

Seduced Into Caring: Virtue and the Life of the Reader

W. Scott Clifton

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Ronald Moore, Chair

Michael Blake

Carole Lee

Andrea Woody

Program Authorized to Offer Degree:

Philosophy

©Copyright 2013
W. Scott Clifton

University of Washington

Abstract

Seduced Into Caring: Virtue and the Life of the Reader

Walter Scott Clifton

Chair of the Supervisory Committee:
Ronald Moore
Philosophy

Can reading novels and watching films make you a morally better person? In my dissertation I argue that living a life in which the experience of fictional narrative artworks occupies a significant role—what I call the life of the reader—can contribute to the cultivation of virtue. I begin by presenting what I take to be a serious moral problem: on the one hand, being able to access the mental lives of others is key to determining and properly weighting the interests of those around us when reasoning morally, while, on the other hand, psychological research has shown that humans are naturally very poor at attaining such access. In the course of presenting the psychological evidence for our evincing poor “empathic accuracy,” I lay out four areas of improvement that researchers have suggested might help us become better at this task. I then turn to showing how living the life of the reader results in improvements in each of these areas, thereby showing that living this kind of life will allow us to make considerable headway toward solving the moral problem described in the early chapters.

Table of Contents

Introduction	5
Chapter 1 Perception and its Role in Morality	14
Chapter 2 The Kind of Motive Necessary for Axiological and Moral Perception	47
Chapter 3 Empathic Accuracy and What It Means to Morality	72
Chapter 4 Art and Morality: A Framework	100
Chapter 5 Particularist Theories	123
Chapter 6 Nussbaum's Hybrid Particularist Theory: Learning Jamesian Perception	155
Chapter 7 Aesthetic Maturity and Narrative Artworks	180
Chapter 8 Defamiliarization, the Aesthetic Trajectory, and Decentering	201
Chapter 9 Heuristic Mindreading and Metaknowledge: An Opportunity to Step Back and Reflect	218
Chapter 10 Moved to Mindread: The Role of Critical Equilibrium in Deliberative Mindreading	247
Chapter 11 Seduced into Caring: Virtue and the Life of the Reader	279

Introduction

An answer sometimes given is that exercise of the symbolizing faculties beyond immediate need has the more remote practical purpose of developing our abilities and techniques to cope with future contingencies. Aesthetic experience becomes a gymnasium workout, pictures and symphonies the barbells and punching bags we use in strengthening our intellectual muscles. [...] Art, long derided as the idle amusement of the guiltily leisure class, is acclaimed as a universal servant of mankind. This is a comforting view for those who must reconcile aesthetic inclinations with a conviction that all value reduces to practical utility.

—Nelson Goodman, *Languages of Art*¹

To understand art, one must be able to disconnect from one's interests (as Kant and his successors put it) to see through others' eyes. And that disinterested sympathetic vision is the very core of moral consciousness. Many of the advertised virtues of aesthetic attention translate cleanly into capacities of moral attention. Artworks and other aesthetic objects teach us to frame things we see so as to make coherent assessment of their composition and interrelation possible. They teach us to admire and protect what is precious because it is rare, or hard to do, or uniquely reflective of historical experience. They teach us to appreciate what is new because it extends our reach, and they teach us to protect what is old because it affirms our grasp. Every prominent aesthetic capacity is on its way to being a moral capacity. On its way, but necessarily never there. [...] [E]thics stands to aesthetics as baseball stands to golf. Baseball shouldn't be confused with golf; there are lots of differences. But there are also lots of serious and subtle connections between them. Most importantly, if one learns how to do the one, one is already part of the way toward learning the other.

—Ronald Moore, “The Aesthetic and the Moral”²

¹ Indianapolis: Bobbs-Merrill, 1968, pp. 256-257.

Did reading *Uncle Tom's Cabin* convince people in the nineteenth century to see black slaves no longer as chattel and instead as persons? Does rooting for the Ku Klux Klan toward the end of the film *Birth of a Nation* cause viewers to have sympathy for the modern white supremacy movement? What effect does viewing Balthus paintings of pubescent and pre-pubescent girls in various stages of undress have on audience members? Do readers of *Lolita*, with its erudite, cunning, and amusing narrator Humbert Humbert, see active pedophilia as less egregious as a result of reading the novel? Is *Catcher in the Rye* a catalyst for teenage rebellion? Do satirical artworks have the power to show a populace the immorality of certain opinions? Can merely reading *The Adventures of Alice in Wonderland* cause someone to sexualize young girls, by virtue of Lewis Carroll's purported inclinations? Is there something wrong about admiring Gauguin paintings when one knows that he left his wife and children to fend for themselves in order to go paint in Tahiti? Can a person aesthetically admire propaganda works, like *Triumph of the Will*, without having to take up a pro-stance toward the organization being propagandized?

Questions like these get to the heart of the debate about whether there are substantive connections between our experience of artworks and our moral lives. They are versions of the broader questions: Is there a causal relation between aesthetic experience and moral behavior? Can engaging with some artworks make a person morally better or worse? And if so, does this causal relation emerge from the respective *natures* of aesthetic and moral experience? Historically, societies have certainly thought that these causal relations exist, though this claim generally manifests itself in its negative form: aesthetic experience has a *corrupting* influence, especially on youth, so access to it must be tightly controlled. This way of viewing art has led to

² *Journal of Aesthetic Education* Vol. 29, No. 2, 1995, pp. 26-27.

book burnings, the removal of films from school libraries, the radical revision of reading lists in high school literature classes, and parental advisory labels on rock and rap albums in stores.

What has been less common in the popular media is the defense of aesthetic experience for its *positive* effects, despite the growing popularity of theories in academia of how such experience can and does improve the moral lives of audience members.

In this dissertation I will argue that art has great power—so great, in fact, that those who worry about the corrupting influence of art are not without basis. Some aesthetic experience is so pervasive that our lives are indeed transformed as a result, and the transformation can be for the worse. Nevertheless, we should value aesthetic experience in part for its very transformative power. In the first part of the dissertation I will present a serious problem for us as moral agents: in order to be of good character, we must be motivated by compassion, understood as empathetic concern, so that we can identify and properly weight the interests of others, but we face tremendous psychological challenges when we try to do so. Empirical research suggests that we are naturally poor at accurately accessing the mental states of those around us, thus making the identification of what is good for others extremely difficult. At the end of this part of the dissertation, I lay out four broad areas that can be improved in order to raise our levels of accuracy.

In the second part of the dissertation I take up the suggestion that aesthetic experience might represent a solution to the problem, first, by providing a conceptual and historical framework surrounding theories whether art and ethics are significantly related; second, by presenting several of the more prominent theories purporting to show how engaging with art can have positive moral effects on audiences. I conclude this part of the dissertation by showing that almost all of these theories are *particularistic* in nature—that is, they focus narrowly on

individual artworks and experiences—and because of their particularist focus, they fail to solve the moral problem discussed in the first part. There is an additional reason, too, why these particularist theories are lacking. They show at most that art can result in effects that could be either morally beneficial or corrupting—what are called “pre-ethical” effects—while the claims found in these accounts are that the effects are necessarily beneficial—what are called positive “ethical effects.” For example, one theorist argues that a particular (kind of) artwork can improve the capacity *to plan* in audience members.³ While this effect certainly could improve the moral lives of its audience, as people will be more able to realize their goals, if these goals themselves are evil in nature, then the artworks would have morally pernicious effects. We have then a local reason to reject these accounts—*viz.*, that they don’t solve our moral problem—but also a global reason—*viz.*, they aren’t able to show what they purport to show.

In the third part of the dissertation I present my argument that art can solve the moral problem discussed in the first part, but that in order for us to see how, we must focus, not on individual works and experiences, but more on a kind of life in which aesthetic experience of the right sort occupies a central role—the life of the aesthetic adult. More specifically, the kind of life under investigation is what I call the life of the reader, since the relevant kinds of artworks will be *fictional, narrative artworks*. I take the notion of aesthetic adulthood from Ronald Moore’s book *Natural Beauty*.⁴ There he argues that there are significant connections between being a *moral adult* and an *aesthetic adult*. “To be a moral adult,” he writes, “is to make a moral decision to live one’s life one way rather than another after having sorted through the normative cards spread on the table.” One’s moral development is a matter of making certain substantial decisions—maybe not *one* substantial decision, but, presumably, a small set of significant

³ See the discussion in chapter 5 of Gregory Currie’s formational account for the details of this argument.

⁴ Peterborough, Ontario: Broadview Press, 2008.

moments determining one's way of living. "To be an aesthetic adult," Moore says, "is to be similarly committed to a course of valuation (which may entail appreciation, preservation, protection, encouragement, among other involvements) regarding both artifactual and natural objects. It is to have worked one's way through the various levels of objective and subjective valuation that one's aesthetic culture presents to the point where one determines that these things count enough to have established a stake in one's life."⁵ I expand on Moore's conception of aesthetic adulthood, arguing that *remaining* an adult is often as important and as difficult as *becoming* an adult. In the third part of the dissertation I attempt to show that living the life of the reader can both improve our ability to empathize and increase our capacity to show concern for others.

The central thesis of this dissertation is the following: there is a *substantive, isomorphic* relation between a person's becoming and remaining an aesthetic adult, thus living the life of the reader, and a person's possessing the right kind of motive necessary for being a virtuous agent. A useful way to think about this isomorphic relation is as training, like that found in sports. Athletes train *prior* to game day, because when things matter, when the stakes are high, the preparatory work necessary for honing the necessary skills and developing the muscles needed for performance should have already been done. For this training to be effective, the activities in the gym must be *similar enough* to those in the game itself, since it wouldn't do the athlete any good to develop skills and muscles that aren't utilized in the actual game, and the training must be *sufficiently challenging*, in order for there to be progress aimed at enhancing game day performance.

In the first epigraph above Nelson Goodman mentions the suggestion that aesthetic experience might be thought of as a "gymnasium workout" in which we use some of our

⁵ Ibid, p. 232.

“intellectual muscles,” which, after becoming stronger, can then be used in other contexts.⁶

Goodman criticizes this view because it implies that “all value [including aesthetic value] reduces to practical utility,” but this conclusion follows only if we hold, as Goodman seems to, *value-monism*—the position that aesthetic works possess only one kind of value. Aesthetic experience can have multiple kinds of value, however, and these different kinds don’t necessarily conflict with one another. We can attribute aesthetic, moral, and cognitive value to artworks, based on their potential effects on our aesthetic, ethical, and intellectual lives without thereby saying that any one of them exhausts their value.

In the second epigraph Ronald Moore provides a second sports analogy resembling the one I have been discussing. He describes Marcia Eaton’s conception of the relation between aesthetics and ethics as akin to the relation between similar sports, such as baseball and golf—they are different in many respects, he says, but similar in some significant ways, too, such that “if one learns how to do the one, one is already part of the way toward learning the other.” These isomorphisms between baseball and golf make it possible that improving in one of the sports can result in improvement in the other.

I find both the gymnasium workout and baseball-golf metaphors helpful to clarify the relation between our aesthetic and moral lives. Some of the processes underlying moral reasoning are the same that underlie making sense of appropriately challenging artworks. As a person gets better at one, she is naturally on her way to getting better at the other, echoing Moore’s description of baseball and golf. They are similar in some respects, and different in others, just as gymnasium workouts are similar to, but different from, performing in the big

⁶ See chapter 10 for discussion of a similar suggestion made by Lisa Zunshine that reading mystery novels in particular does for our mindreading skills what bodybuilding does for athletes. I’ll argue that her conclusion that, just as bodybuilding doesn’t better equip athletes to do daily tasks, pushing our mindreading capabilities to their limits in aesthetic contexts doesn’t result in our being better able to relate to others in real life, is wrong.

game. The gym is a *virtual* space, exhibiting some features found in real game situations and some not. Those features not shared by both the gym and game contexts have to do with the costs of errors and the freedom to be innovative. The boxer spars with her partner, the tennis player hits balls projected by a machine, the balance beam gymnast practices on an outline of the beam on the floor, the football team walks through the various plays they plan to try out on Saturday. The training context is often just a simulation of the game, and as such, allows for the appropriate steps to be taken for genuine improvement and progress.

The same kind of opportunity exists for moral agents when they adopt the life of the reader. The reader confronts challenges in the aesthetic context, some of which require *mindreading*—the act of accessing the mental states of others in order to explain and predict their actions, and, more specific to the discussion that follows, in order to determine and properly weight their interests. Since the margin of error is wide in the aesthetic context and the costs of error are low, the reader can afford to try out different tactics, adopt novel strategies, and take the time to reflect on mindreading failures, isolating which factors were responsible for the failures and thus determining whether she is to blame. She can even try it again, noting what caused the initial failure and taking steps to prevent these things from happening again. In real life failed mindreading, and poor moral decisions based on such failures, often must pass either undetected or unexamined, since life goes on. In aesthetic contexts life doesn't have to go on, because the reader can pause, rewind, or fast-forward, all in the service of getting it right. This is not to suggest that the reader says to herself, "I must get better at mindreading, so that I get better at moral reasoning. I will then pay close attention to my mindreading in this novel or film and then reflect on my failures to determine why I failed." The reader will just try to get things right, and when she doesn't, she will often try again, paying attention to other details. Some of this process

of learning how to get it right crosses over into the real life situations of the aesthetic adult, resulting not only in a better mindreader, but a better moral reasoner, a considerate person, a moral adult. The reader is led into the aesthetic experience by the promise of the pleasure of engaging with the artwork and tempted back to aesthetic experiences by past pleasures, but, after the aggregation of subtle and slight changes resulting from the repeated experiences occurring over time, she ends up increasing her capacity to care for those around her. Therefore, in a sense, the reader is *seduced* by aesthetic experience into caring—not just for fictional characters, not just for her closest confidantes, but also for those around her who stand to be affected by her actions.⁷

⁷ A note on the methodology of the dissertation: it consists partly of philosophical analysis and argumentation and partly of the application of psychological research. It's important not to build the account found here strictly on philosophical argumentation, because the issue is about whether purported causal connections between art and morality really exist. We can't adequately deal with this issue without taking into account facts about our psychology and how we relate to the world. In a recent newspaper editorial Gregory Currie ("Does Great Literature Make Us Better?" *New York Times*, June 1, 2013) claims that those working on this issue fail to recognize that it's essential to base our investigations on empirical work as well as philosophy, saying that "we need to show that exposure to literature itself makes some sort of positive difference to the people we end up being," which "will take a lot of careful and insightful psychological research." Unfortunately, says Currie, "[n]ot nearly enough research has been conducted; nor [...] is the relevant psychological evidence just around the corner." Worse, "advocates of the view that literature educates and civilizes don't overrate the evidence — they don't even think that evidence comes into it," which ends up with most people "soldier[ing] on with a positive view of the improving effects of literature, supported by nothing more than an airy bed of sentiment." I think that Currie is right that we need to draw on psychological research in investigations like this, but he is wrong that the current research is underdeveloped. For this reason I will deal with some of the more compelling studies that shed light on the issues discussed, but it should be recognized that they are there to lend support to the more substantive *philosophical* claims I make and defend throughout.

Part I

A Problem for Morality

Chapter 1

Perception and its Role in Morality

Traditionally discussions in ethics have focused on what makes particular acts *right*, but the recent surge in interest in virtue ethics has tended to shift the focus from acts to persons, and onto what constitutes *good character*. This shift has resulted in criticisms of both utilitarianism and Kantian deontology, as these theories are usually construed to be act-centered, and thought not to be able to answer questions about what makes a person good, rather than what makes an action right. Rosalind Hursthouse¹ raises doubts, however, about distinguishing theories in this way, arguing that it's overly simplistic. Moreover, some theorists have argued that utilitarianism and Kantianism can be described in ways that does accommodate the shift in focus.² Therefore, while the project that I intend to undertake will centrally involve asking whether there is any characteristic essential to being a good person, I make no assumptions about whether virtue ethics is the only kind of theory that can make sense of the issue.

To begin this project, let's consider the world in which we now live. While it may be edifying to investigate what makes actions right *simpliciter*, as discussions have typically run, seeking moral absolutes and constructing universal statements about rightness and goodness, proceeding without a recognition of the way the world currently is might turn out to be fiddling while Rome burns. As we turn our attention to the way the world is and take into account both the demands that moral reasoning rightfully makes upon us and the tensions they create with our

¹ *On Virtue Ethics*, NY: Oxford University Press, 1999, chapter 1.

² See Robert Merrihew Adams, "Motive Utilitarianism" *Journal of Philosophy* 73, pp. 467-481, for an example of how utilitarianism is able to expand beyond such a focus on acts and David Velleman, "Love as a Moral Emotion," *Ethics* Vol. 109 No. 2, pp. 338-374, for how Kantianism is able to do so as well.

psychological tendencies, we will see a significant problem for morality present itself in chapters 1, 2, and 3—a problem to which I will provide a solution in subsequent chapters. I am not prepared to argue that it's the *only* solution, but it's a suitable and satisfactory one. In order to see what the problem is, we have to begin with what can be called *the human predicament*.

1. The Human Predicament

The relevant notion of the human predicament appears in G. J. Warnock's *The Object of Morality*,³ who describes it as being in a Hobbesian world where there is a "certain 'natural' tendency for things to go very badly"⁴ and our having a strong interest in these things not going badly. The use of "badly" here isn't moral, but, Warnock argues, the "'natural' tendency" for things to go badly is not natural either, as it results from our own limitations—these limitations themselves being partially moral. He claims that if human beings were "less vulnerable, less aggressive, less egotistical, less irrational, more intelligent, more self-sufficient, and more favoured by material circumstances,"⁵ things might go much less badly.

Warnock thinks there are two general limitations in human beings today that combine to produce the human predicament: limited (instrumental) rationality and limited sympathies. The first limitation is demonstrated in the many cases of people designating individual ends, and then failing to pursue the means to achieve those ends. Humans, he says, "are not just naturally disposed always to do what it would be best that they should do, even if they see, or are perfectly

³ London: Methuen, 1971.

⁴ Ibid, p. 23.

⁵ Ibid, p. 23.

in a position to see, what that is.”⁶ This first limitation then is a form of what the ancient Greeks referred to as *akrasia*. The notion of limited sympathies, left vague by Warnock, is marked by an orientation toward the self and what is dear to the self. A man possessed of limited sympathies “who does not like being hungry,” Warnock says, “and who is naturally inclined to take such steps as he can to satisfy his hunger, may very well care less, even not at all, about the hunger of others, and may not care at all whether anything is done to satisfy them.”⁷ And if there are some people about whose hunger he does care, like close friends and family, it will only be because they are so dear to *him*.

These two limited capacities combine to produce the human predicament—i.e., things going very badly for us—and, since we have a strong interest in breaking free of this predicament, we have a similar interest in addressing and mitigating these limitations. Though both are serious, Warnock thinks that our limited capacity for sympathy is more pressing, and constitutes the more crucial topic for moral philosophy. This is because people can be perfectly rational and still be faced with the human predicament. He writes that “a man may be wholly rational, clear-headed, sane, and still, if he is not to act destructively towards others, it is essential that he should not simply see, but care, what becomes of them.”⁸ Caring, rather than rationality, seems to be the basic *moral* notion here.

We might ask though why Warnock would think that a deficit in caring is a more serious problem than flawed reasoning. Shouldn't we worry about the possibility that I care for someone while being utterly unable to figure out what to do in order to address that person's suffering? Isn't this possibility as worrisome as a situation in which I know how to help, but don't care enough to do so? The reason Warnock thinks the deficit in caring is more important lies in the

⁶ Ibid, p. 21.

⁷ Ibid, p. 21.

⁸ Ibid, p. 24.

fact that Warnock is trying to elucidate the task of *moral philosophy*. It's not within the purview of this area of inquiry to improve a person's capacity to reason, but one of the purposes of moral philosophy is "to expand our sympathies, or, better, to reduce the liability to damage inherent in their natural tendency to be narrowly restricted."⁹

Thus Warnock makes the two following claims:

- (1) the deficit in caring found in humans is more serious than the deficit in rationality
- (2) it's the object of moral philosophy (a) to expand our sympathies and/or (b) to mitigate the damage done by limited sympathies

It's not clear why Warnock includes the hedge found in (2b), unless it's because he thinks that the chances of success of (2a) are much lower than those of (2b). Nonetheless, if we could effect an adequately wide expansion of human sympathies, damage done by limited sympathies would *a fortiori* be mitigated, since the limitations would have been reduced. Therefore, I am prepared to adopt Warnock's description of the human predicament, but with the following amendments of his claims, in order to be clear that we are headed in a different direction in the ensuing discussion:

- (3) the deficit in caring is supremely important (no other deficit is greater) when we are trying to solve the human predicament
- (4) it's the object of moral philosophy (a) to expand our sympathies, if possible; (b) to mitigate the damage done by our limited sympathies, if (a) is not possible

⁹ Ibid, p. 26.

In this chapter I will offer a justification of (3), while in later chapters I will show that (4) may be true, as far as it goes, but that there is a more effective means than moral philosophy of expanding our sympathies: art.

2. Moral Perception

In *Gravity and Grace* Simone Weil says the following: “A great many people do not feel with their whole soul that there is all the difference in the world between the destruction of a town and their own irremediable exile from that town.”¹⁰ The reason for these people’s thinking in this way, she says, is that their “sense of reality” is “insufficient,” that they are victims of an “attachment” to things of the world so deeply felt that they can’t perceive anything without perceiving it as “something perceived by me.” When they see a chair, it’s not a chair they see, but a chair in relation to themselves. Similarly, when they encounter another person, they see the person only in relation to themselves. In many cases this is a distortion, an inaccurate perception, that leads to an overall perspective on the world grounded in an insufficient sense of reality. We can take her point here and develop it in order to see the importance of addressing the deficit in caring highlighted by Warnock, and subsequently bring out what can be done to expand our sympathies.

Weil describes a significant connection between attending to features in the world and making plausible judgments about what should morally be done. This naturally prompts us to think about this issue within the context of work done on *moral perception*. Theories holding that

¹⁰ NY: Putnam, 1952, p. 59.

perception has a role to play appear in different forms.¹¹ Some theories hold that *all* of the moral work is done by perception, claiming that virtuous agents not only see concrete particulars in a moral light, but also what actions to perform or that the actions being performed are right or wrong. Consider, for instance, the example provided by Gilbert Harman: “If you round a corner and see a group of hoodlums pour gasoline on a cat and ignite it, you do not need to *conclude* that what they are doing is wrong; you do not need figure anything out; you can *see* that it is wrong.”¹² Let’s call this *strong moral perception*. Other theories ascribe a *preparatory* role to perception, claiming that it has a crucial part to play, while not doing all the moral work, and thus isn’t involved at the moment of decision. Lawrence Blum, for instance, says that moral perception “constructs what an agent is faced with as ‘a (moral) situation’ in the first place” and that the “idea of moral judgment as bridging general rule and particular situation depends on a prior individuating of ‘the situation.’” Moral perception, Blum contends, “provid[es] a setting in which moral judgment carries out its task.”¹³ Let’s call this kind *weak moral perception*. Both of these variants should be kept distinct from the weakest kind of theory holding that perception has a role to play in morality—the view that says that perception is necessary for good moral reasoning, but in no way different from how it works in other kinds, such as instrumental reasoning. This view—what we could call *minimalism*—doesn’t posit *moral* perception at all, since it’s merely (non-moral) perception being used for moral purposes.

¹¹ For some recent discussions of moral perception, see Robert Audi, "Moral Perception and Moral Knowledge," *Proceedings of the Aristotelian Society Supplementary Volume LXXXIV*, (2010), pp. 79-97; *Moral Perception*, (Princeton, NJ: Princeton University Press, 2013); Andrew Cullison, "Moral Perception," *European Journal of Philosophy*, Vol. 18, No. 2, (2009), pp. 1589-175; David Copp, "Four Epistemological Challenges to Ethical Naturalism," *Canadian Journal of Philosophy*, Vol. 30 (Supplement), (2000), pp. 31-74; Daniel Jacobson, "Seeing by Feeling: Virtues, Skills, and Moral Perception," *Ethical Theory and Moral Practice*, Vol. 8, (2005), pp. 387-409; Justin McBrayer, "Moral Perception and the Causal Objection," *Ratio*, Vol. 23, No. 3, (2010), pp. 291-307; Sarah McGrath, "Moral Knowledge by Perception," *Philosophical Perspectives*, Vol. 18, (2004), pp. 209-228; Michael Watkins & Kelly Dean Jolley, "Pollyanna Realism: Moral Perception and Moral Properties," *Australasian Journal of Philosophy*, Vol. 80, No. 1, (2002), pp. 75-85.

¹² Gilbert Harman, *The Nature of Morality: An Introduction to Ethics*, (NY: Oxford University Press, 1977), p. 4.

¹³ Lawrence Blum, *Moral Perception and Particularity*, (NY: Cambridge University Press, 1994), p. 21.

Which, if any, of these roles does perception play in morality? Is it stronger than minimalism? As strong as strong moral perception? Or is weak moral perception necessary for moral reasoning? In order to answer these questions we should look closely at what factors are at play.

First, we have to remind ourselves that all observation is theory-laden.¹⁴ Perception itself is not free of knowledge; therefore, it's probably the case that moral perception is not free of moral knowledge. Two individuals, presented with the very same situation, might "see" very different things, depending on what moral knowledge each of them has. Consider the following example:

John and Joan are riding on a subway train, seated. There are no empty seats and some people are standing; yet the subway car is not packed so tightly as to be uncomfortable for everyone. One of the passengers standing is a woman in her thirties holding two relatively full shopping bags. John is not particularly paying attention to the woman, but he is cognizant of her. Joan, by contrast, is distinctly aware that the woman is uncomfortable.¹⁵

Both John and Joan see the woman standing and holding the two bags, but only Joan "sees"—i.e., is aware—that the woman is uncomfortable. The two see the same woman, but Joan sees more of her, about her, and of her situation. She sees not only a woman holding bags and traveling on the subway—which is the entirety of what John sees—but an uncomfortable woman struggling to keep her footing and maintain a grip on her bags. The difference between the two, Blum claims, is that "Joan *saliently perceives* [...] the standing woman's good (i.e., her comfort)

¹⁴ See N. R. Hanson, *Patterns of Discovery: An Inquiry into the Conceptual Foundations of Science*, Cambridge: Cambridge University Press, 1958.

¹⁵ Blum, *Moral Perception and Particularity*, p. 32.

as a stake in a way that John does not.”¹⁶ What does it mean to “saliently perceive” the woman’s good? It is to perceive “a morally relevant value in [a] situation.”¹⁷ One sees more than the concrete particulars, but *moral value embodied in those particulars*. Joan perceives the morally relevant value in the subway scenario, while John doesn’t.

The morally relevant value cannot just be read off of the particulars, but must be assessed through a special kind of awareness, itself emerging from knowledge. One must be fairly familiar with the phenomenon of uncomfortableness, its negative nature, and the possibility of its alleviation as a result of the decision of others to alleviate it, in order to see the subway scenario as one with moral content. Once one sees this situation as a morally significant one, he must make a choice either to continue to sit, not giving the woman with the bags the chance to lighten her load or to rise and give her the seat. Unfortunately, John fails to see the situation as a moral one, because he fails to recognize that the woman has a claim on him—not a moral claim to *the seat*, but to his *consideration*. John doesn’t even reach the stage at which he makes the choice, since he fails to see that there is a choice to make. Standing and giving her the seat might not be an obligation—doing so may be a supererogatory act. But John has no way to make this determination because he fails to see the woman as a moral claimant at all, and this, according to Blum, is due to John’s “situational self-absorption or attentional laziness.”¹⁸ Thus John is morally blameworthy. Since both John and Joan see the woman, but only Joan sees her *as* uncomfortable, and this is the only difference between the two, we should conclude that John *should have* seen her uncomfortableness. He should be prepared to recognize the suffering in those around him, especially when he stands to alleviate the suffering. When he fails to recognize such suffering, he evinces a moral failing.

¹⁶ Ibid, p. 32, my emphasis.

¹⁷ Ibid, p. 32.

¹⁸ Ibid, p. 33.

Consider a second example:

Theresa is the administrator of a department. One of her subordinates, Julio, has been stricken with a deteriorating condition in his leg, causing him frequent pain. He approaches Theresa to help work out a plan by which the company, and in particular their department, can accommodate his disability. Theresa is unable to appreciate Julio's disability and the impact it is having on his work. Although in principle Theresa accepts the company's legal obligation to accommodate Julio's disability, in fact she continually offers him less than he needs and is entitled to.¹⁹

Here Theresa sees Julio as a disabled person, so she could perhaps be said to have a fuller sense of who Julio is than John did with the woman on the subway. Nevertheless she fails to recognize Julio's challenges presented by his disability and the pain caused by these challenges as well as the fact that she is in a position to address these challenges and alleviate his pain. She doesn't "grasp fully what this means for him, and to take in or acknowledge that pain. The level of Julio's pain and its impact on his mental state is insufficiently salient for Theresa."²⁰ His pain doesn't represent for Theresa a moral claim. All she sees is the *legal* claim his disability gives Julio, and this is as far as she allows it to go.

It's possible to interpret each of these scenarios in such a way that John and Theresa are not unaware of the moral claims on them by people suffering near them, but have decided long ago not to be concerned with the suffering of others. For example, consider the following background information that doesn't appear in Blum's Theresa example, but would change our moral evaluation of her or at least our reasons for the evaluation:

¹⁹ Ibid, p. 34.

²⁰ Ibid, p. 34.

Not long after Theresa had entered middle management, before becoming administrator, she had tried to address the needs of Kristina, one of her subordinates who had excessive fatigue issues. She allowed Kristina to split her day into two halves, so that she could go home in the middle of the day to rest and return later in the day, completing her shift after the end of the regular workday. After Theresa had allowed this, however, many other employees began to complain that they had legitimate health issues that required addressing. Pretty soon, she was overrun by employee requests for special accommodations, and her superiors told her to get things under control or she would be demoted. After that period of her career, Theresa decided that in order to guard against charges of favoritism and excessive accommodation requests, she would only grant those requests required by law and ignore all others.

In this amended case Theresa may still be morally blameworthy for her treatment of Julio, but for a completely different reason. She recognizes both (a) the morally salient details of the situation (Julio's challenges and subsequent pain) and (b) the fact that they are morally salient. She is morally at fault because she recognizes that he is in pain and that he has a moral claim on her which isn't mitigated by her earlier experiences. In the original example she is morally at fault because she recognizes (a) but not (b). Thus the relevant knowledge Theresa lacks in the original example is that the details—morally salient details—of the situation before her are in fact morally salient.

By considering this amended version, we can now see the original examples of John and Theresa as ones in which the missing ingredient is a recognition that the morally salient details *are morally salient*. In other words these examples illustrate failures not in *seeing*, but in *seeing-as*, with the implication that John and Theresa are missing some piece of relevant *knowledge*—more specifically, *moral* knowledge.

We haven't found yet where perception belongs, but we can now see that it's not pure perception, free of moral consideration, so the presumption at this point is that we should adopt a

position stronger than minimalism. To illustrate this point, let's look at the following schematic of the moral reasoning process provided by Blum:

Step 1: Accurately recognize the features of the situation.

Step 2: Recognize the features of the already characterized situation as morally significant. (Moment of moral perception)

Step 3: Move from the moral characterization of the situation (step 2) to raising the question of whether one should take some action in the situation.

Step 4: Judge whether or not one should in fact take action.

Step 5: Select a rule or principle that one takes to be applicable to the situation.

Step 6: Determine the act that best instantiates the principle one has selected.

Step 7: Figure out how to perform the act specified in step 6.²¹

It's clear from how these steps are organized that Blum holds a weak moral perception theory. Notice that, according to this schematic, a strong moral perception theory would hold that the moment of moral perception would happen in step 6. One would *see* what act one should perform. But if one just sees what action to perform in step 6, it's not clear that step 5 would be theoretically or temporally prior to step 6. Rather, it would seem that if rules are used at all, they are applied *contemporaneously* with the determination found in step 6. *Prima facie*, at least, it seems that strong moral perception theories must be coupled with a form of *rule-particularism*.²²

Moral particularism comes in several distinct forms, but the one we are most interested in here is the kind that says that moral reasoning is a matter of attending to the one's particular

²¹ Ibid, chapter 3.

²² I take this term from Roger Crisp, "Particularizing Particularism," in *Moral Particularism*, ed. Brad Hooker & Margaret Olivia Little, NY: Oxford University Press, 2000, pp. 23-47.

situation and making a reasoned judgment, without first consulting moral rules and principles and subsuming the situation under the relevant rule. On many characterizations of moral reasoning, the agent confronts a concrete situation with moral rules already in hand and, upon understanding what kind of situation is before the agent, the agent decides what action to perform by applying the rule to the situation. Under this description, possessing moral principles is theoretically prior to any observations of the concrete situation and moral judgment is a matter of *inference*. Consider an alternative to Blum's characterization of moral reasoning—Aristotle's practical syllogism:

Major Premise: A relevant rule or principle

Minor Premise: A set of concrete particulars subsumed under the rule

Conclusion: Prescription for action

According to this scheme, one already has the relevant rule in hand upon confronting the situation. After one has processed the information found in the concrete situation, one makes an inference to the conclusion. Rule-particularism, however, holds that perception is theoretically prior to the calling up of moral principles. Moral principles play only a subsidiary role to perception, since every concrete situation is going to have features that make a straightforward application of general principles difficult, if not impossible. How do I know which is the relevant principle until I have some description of the situation? Charles Larmore argues that the Aristotelian practical syllogism isn't in fact a representation of moral judgment—it's a product of it:

The major and minor premises of the moral syllogism (whose conclusion is the action to be done) express, respectively, the duty to be performed and the way it can be fulfilled in the given circumstances—the good and the possible [...] But this means that the exercise of moral judgment precedes the inference from these premises; it consists in arriving at them in the first place. Thus, we cannot take the practical syllogism itself as a model of how moral judgment works.²³

On strong moral perception theories, where one just sees the action to perform or that the actions one is observing are right or wrong, it doesn't seem to be a matter of applying a moral rule to the observed particulars. Indeed, it doesn't seem to be an inferential process at all. Instead, it's a matter of perceiving the particulars *and* applying moral rules at the same time, thus resulting in a simultaneous perceiving and judging, which leaves no room for inference. The concrete particulars meet the relevant rule and, upon such a meeting, moral judgment just occurs.

There is little reason to think that *any* strong moral perception theory must be tied to a rule-particularist theory, but it's hard to see how such an uncoupled theory would work. We'll return to rule-particularism in chapter 6, where we will look at Martha Nussbaum's theory of how reading novels can make someone morally better. Here I am going to bracket the question of whether perception's role in moral reasoning belongs in step 6 on Blum's schematic, since defending such a view will also require defending rule-particularism, a task that would pull us too far afield at this point. Since we have already seen evidence that minimalism is too weak, this leaves as a viable candidate weak moral perception. In the next section we will look at a particular theory utilizing weak moral perception: the moral philosophy of Iris Murdoch.

²³ *Patterns of Moral Discovery*, NY: Cambridge University Press, 1987, p. 16.

3. Murdochian Moral Perception

We've been investigating the role of perception in morality because of a recognition, based on a point made by Simone Weil, that having knowledge of the world in front of us is essential to making good moral decisions. Put another way, a person of good moral character will have to be able to see the world as it really is—and this notion of seeing is only partly metaphorical. Drawing on Weil's work, Iris Murdoch constructs an account holding that seeing the world as it really is demands a kind of love on the part of the moral agent. If she's right, then this will lend credence to earlier claims about the deficit in caring—being a good person will require caring for others in some sense. Once we have Murdoch's account in front of us, we will have to press further and determine in *what* sense exactly caring is necessary for moral reasoning.

Aside from her novels Iris Murdoch is probably best known for calling attention to the importance in moral philosophy of the private, internal mental lives of moral agents, in contrast to the prevailing theories that focus entirely on agents' public, external, behavior.²⁴ Most relevant to our discussion here, she stresses the role of motive and the necessity of the grounding of one's motive in knowledge for good moral character. Even if my behavior is indistinguishable from that of someone acting morally, my behavior could be reflective of ignorance and thus a flawed character. This point is illustrated in her now famous example of M who initially judges her

²⁴ There has been a recent surge in interest in the moral philosophy of Murdoch. For some of the most recent work, see Justin Broackes, *Iris Murdoch, Philosopher: A Collection of Essays* (Oxford: Oxford University Press, 2012); Maria Antonaccio, *Picturing the Human* (Oxford: Oxford University Press, 2000); *A Philosophy to Live By: Engaging Iris Murdoch* (Oxford: Oxford University Press, 2012); Heather Widdows, *The Moral Vision of Iris Murdoch* (NY: Ashgate, 2005); Megan Lavery, *Iris Murdoch's Ethics: A Consideration of Her Romantic Vision* (NY: Continuum Studies in British Philosophy, 2007); Marije Altorf, *Iris Murdoch and the Art of Imagining* (NY: Continuum Studies in British Philosophy, 2008); and Margaret Holland, "Touching the Weights: Moral Perception and Attention," *International Philosophical Quarterly*, Vol. 38, No. 3, (1998), pp. 299-312.

daughter-in-law D negatively, based on prejudices, which, she decides later on, actually prevented her from *really seeing* D:

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.²⁵

Later on, after M's son and D have moved away, thus forestalling any new observations of D on M's part, M realizes her judgments might have been corrupted by her jealousy and resentment. She then returns to the task of evaluating D as a person, tries to shed her prejudices and really see D for what she is (or was). She then discovers D "to be not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful."²⁶ In other words, in spite of M's behaving "beautifully to the girl throughout," her original treatment of D was grounded in error, due to self-absorption, and was therefore reflective of bad moral character.²⁷ After deciding she may have been in error, M attempts "not just to see D accurately but to see her *justly* or *lovingly*."²⁸ This seeing someone with "justice" and "love" Murdoch calls "attention." "I can only choose within the world which I can *see*," Murdoch writes, "in the moral sense of 'see' which implies that clear vision is a result of moral imagination and moral effort."²⁹

²⁵ Iris Murdoch, *The Sovereignty of Good* (NY: Ark Paperbacks, 1970), p. 17.

²⁶ *Ibid*, pp. 17-18.

²⁷ *Ibid*, p. 17.

²⁸ *Ibid*, p. 23.

²⁹ *Ibid*, p. 37.

We can see here that, as she describes this example, Murdoch makes two claims similar to those considered above: that *knowledge* is essential to moral reasoning and that *perception* is especially apt in acquiring that knowledge. Let's look at each of these in turn.

First, we can see that what makes M at fault in the example is that she lacks certain knowledge. On a strictly act-centered view, M would be morally blameless, since her actual behavior is indiscernible from what it would have been had she genuinely admired D. Now that M is no longer around D, however, she begins to question her earlier evaluations of D, saying to herself at some point, "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." In looking again, she realizes how unjust her judgments of D had been, how tainted her perception had been by prejudice, and how utterly unable she was to really see her daughter-in-law. Since D is no longer present, it's not new observations that have changed M's opinion of her. The revisions come later, after M's behavior has already occurred. If we stipulate that D really was a refined and sophisticated person, and that M was dead wrong in her earlier judgments, then we might ask why M was *morally* at fault. She made incorrect judgments, yet these judgments didn't affect her behavior. The difference between the actual case in which she makes *inaccurate* and *negative* judgments and a counterfactual case in which she makes *accurate* and *positive* judgments doesn't lie in her particular behavior *B*, which is the same in both cases. In the counterfactual case, however, M displays *B* *because* of her judgments, while in the former, she does so *despite* her judgments. It will benefit us to see what weight "despite" carries here.

Consider a third case: M displays *B* despite her negative judgments, but, unlike in the actual case, the negative judgments are accurate. Is M morally at fault here? Certainly not. We're more likely to praise her for being able to cast aside her negative opinion for the sake of civility.

Therefore, in the two counterfactual cases, in which M displays *B* and makes accurate judgments, we morally praise her, while in the actual case we blame her. Why? Behavior isn't the crucial factor, since it's the same in all cases. It must be motive. It cannot be motive alone, however, since in the actual case and second counterfactual case, M displays *B* despite her negative judgments, thus performed from the same motive. It must be motive *grounded in accurate judgments*—i.e., grounded in *knowledge*. What kind of knowledge? Murdoch describes it as “—“not [...] impersonal quasi-scientific knowledge of the ordinary world, whatever that may be, but with a refined and honest perception of what is really the case.”³⁰ This “refined and honest perception” she calls “attention.” It's here that Murdoch moves to the second claim from above—that the knowledge essential to motive derives, at least in part, from perception.

Let's return to Weil's example from before, where she says, “A great many people do not feel with their whole soul that there is all the difference in the world between the destruction of a town and their own irremediable exile from that town.” Murdoch expands on Weil's notion of gaining a sufficient sense of reality. Of course, what's disturbing about a person who sees an equivalence between a destruction of a town and her being exiled from it is, at one level, her failing to prioritize over her own interests the suffering and death experienced by the town's residents. But why is seeing the destruction of the town in this way wrong? Is it merely because it fails to take due account of the residents' suffering or is it the self-absorption that subsequently causes the failure to take such account? Murdoch thinks that it's the latter. She writes that the problem to be solved within moral philosophy is to “suggest methods of dealing with the fact that so much of human conduct is moved by mechanical energy of an egocentric kind.” Echoing Warnock, she says that in “the moral life the enemy is the fat relentless ego.”³¹ The ego allows

³⁰ Ibid, p. 38.

³¹ Ibid, p. 52.

many of us to engage in “personal fantasy,” described as “the tissue of self-aggrandizing and consoling wishes and dreams which prevents one from seeing what is there outside one.”³² If the primary moral failing in Weil’s example were the inability to prioritize the residents’ suffering, then at least in those cases in which an agent who’s engaged in personal fantasy *is* able to prioritize the suffering of others, she would not be at fault. If it turned out, for example, that she accidentally focused on the townspeople’s suffering, as a result of observing the destruction of the town through a self-absorbed lens, there should be no moral problem, but it seems like there would indeed still be a moral problem, even when the suffering of others is (accidentally) recognized as important. Thus what makes the persons described by Weil of a bad character isn’t primarily the failure to take due account of the suffering, but, at a broader level, the failure to see the townspeople free of the prejudices activated by the hold of the ego.

It’s true that part of what makes it bad to allow personal fantasy to affect one’s observations is that it will so frequently result in instances like seeing the town’s destruction in this way, but what’s essential isn’t the prioritizing of the suffering of others: what’s essential is *really seeing* the suffering for what it is and not merely for how it affects us. Being able to construct an accurate description of the world is important insofar as a central component of moral reasoning is the possession of *knowledge*, as we saw above. It’s tempting nonetheless to read Murdoch as making a two-step argument, first showing (1) that allowing the ego to affect our perception almost always results in the kind of egregiousness found in Weil’s example, and then showing (2) that, since it’s personal fantasy that produces this result, we should work to free ourselves of it. This would place the primary importance of the view in seeing the suffering of others in the correct light. If someone could achieve this without working on the purification of perception, so much the better for them. Murdoch says some things that encourage this reading,

³² Ibid, p. 59.

saying, for instance, that the “more the separateness and differentness of other people is realized, and the fact seen that another man has needs and wishes as demanding as one’s own, the harder it becomes to treat a person as a thing.”³³ This might suggest that what is most important is not to treat others as things. If this is the case, then *whatever* has the transformative potential to motivate us to treat others as ends in themselves would be as beneficial as producing in ourselves “a refined and honest perception of what is really the case.”

This is clearly not what Murdoch ultimately argues, however, for she also says,

All just vision, even in the strictest problems of the intellect, and *a fortiori* when suffering or wickedness have to be perceived, is a moral matter. The same virtues, in the end the same virtue (love), are required throughout, and fantasy (self) can prevent us from seeing a blade of grass just as it can prevent us from seeing another person.³⁴

By equating seeing another person, and that person’s suffering when suffering is present, with seeing a seemingly inconsequential blade of grass, Murdoch shows that what she holds most valuable is seeing the *world* as it really is. When one sees the world as it really is, it follows that one will see suffering when it is present. Yet this is not the most morally significant piece of development. The seeing is itself a moral achievement, because “the realism (ability to perceive reality) required for goodness is a kind of intellectual ability to perceive what is true, which is automatically at the same time a suppression of self.”³⁵ Thus what makes attention and what she calls *love* a moral achievement isn’t found in the specific cases in which an agent is able to see

³³ Ibid, p. 66.

³⁴ Ibid, p. 70.

³⁵ Ibid, p. 66.

the suffering of others, but in all cases in which the agent sees the world aright as a result of the “suppression of self.”

It’s clear then that the role Murdoch ascribes to perception is more than the minimalist view described earlier that accurate perception is important for moral reasoning, but no more than for prudential reasoning, as it’s the same perception used for moral purposes. We might be better reasoners when we have accurate information resulting from a reliable perceptual system, but, the minimalist could argue, this perception really isn’t *moral* perception at all. Murdoch envisions the ability to perceive reality, however, as falling under certain norms, one of which is breaking free of the hold of the ego. This breaking free requires a struggle, however, and is often painful. One of the undeniable facts about us, she says, is that we are naturally selfish. One must *resolve* to become unselfish and *strive* to achieve what she calls “unselfing.” This is the only way we will accomplish a “refined and honest perception of what is really the case.” Thus what makes Murdochian moral perception *moral* is (a) the recognition that one’s own interests are not more valuable than those of others simply because they are one’s own and (b) the ceaseless attempts to purify one’s capacity for attention.

Because Murdoch refers to “the ability so to direct attention [...] outward [and] away from the self,” and to the attitude of the person who sees the world as it really is, as “love,” some people are tempted to think that Murdoch’s account is built around the kind of love we show for our closest friends and confidantes. Getting clear about what she means by love is important for us to understand how we might best use her moral perception to address the deficit in caring

highlighted by Warnock. In order to see how her notion of love is *not* the standard conception of love, we should consider a criticism raised against Murdoch's account by David Velleman.³⁶

Velleman's worry—what might be called the problem of partiality—is that, if love is a moral emotion, as it seems to be for Murdoch, then we have a tension between love and morality, because, on the one hand, the “moral point of view is impartial and favors no particular individual,” while, on the other, “favoring someone in particular seems like the very essence of love.”³⁷ If Murdoch's notion of love includes “favoring someone in particular,” then her account would appear to suffer from this tension. Velleman mentions one strategy for resolving this tension:

One way to bring [love and morality] into convergence would be to reject the Kantian conception of morality as impartial. Lawrence Blum endorses the view, which he attributes to Iris Murdoch, that ‘the moral task is not to generate action based on universal and impartial principles but to attend and respond to particular persons.’ The way to effect a convergence of spirit between love and morality, according to Blum, is to allow greater partiality in our conception of morality.³⁸

Velleman finds this strategy headed in the wrong direction and opts instead to “rethink the partiality of love.” He goes on to provide an account intended to show that “really looking” as it appears in Murdoch's work, and *Achtung*, understood as respect or reverence, as it appears in Kant's work, enjoy a “deep conceptual connection.”³⁹ Love and respect for the rational being of

³⁶ See my “Murdochian Moral Perception,” *Journal of Value Inquiry*, forthcoming, where I expand the present argument in order to address, in addition to Velleman's, a criticism of Murdoch raised in Charles Starkey, “On the Category of Moral Perception,” *Social Theory and Practice*, Vol. 32, No. 1, (2006), pp. 75-96.

³⁷ Velleman, “Love as a Moral Emotion,” p. 338.

³⁸ *Ibid*, pp. 341-342.

³⁹ *Ibid*, p. 343.

other persons is purported to be much more closely aligned—love, he argues, is more impartial than it seems, but Kantian respect too is more suited than it seems “to matters of the heart.”⁴⁰

Velleman’s account is in response to his reading of Murdoch as providing a fairly standard conception of love, which includes partiality. By offering up a revised conception of love, containing a greater degree of impartiality, that is consistent with morality’s demand for impartiality, he believes that he has eliminated the tension. He does concede at one point that “Murdoch’s ethic of attending to the particular is not necessarily at odds with the ethics of impartiality,” noting that Murdoch describes attention as “impersonal” and “an exercise of detachment,” and that in her work “there is no suggestion that particularity entails partiality,” but his worry is that her promise of impersonality, which he construes as impartiality, is not made good by her account. He thus goes on to draw connections between love and Kantian respect in an attempt to show that the two are much more closely aligned.

Velleman’s project, while valuable in its own right, is nevertheless motivated by a flawed understanding of Murdoch’s love and what she is trying to do. In particular, he doesn’t take seriously enough her claims about love’s—indeed, the “highest love”—being impersonal. There are two reasons this is problematic for his criticism of Murdoch: first, he implicitly synonymizes “impartial” and “impersonal,” which gives the false impression that Murdoch tries and fails to give an account committed to the requirement that morality be impartial; and second, he fails to recognize that a love described as impersonal is likely not going to be what we standardly think of as love.

When Velleman concedes that Murdoch describes the highest love as impersonal, he—perhaps out of charity—imputes to her a commitment to the requirement that morality be impartial. But impersonality is significantly different from impartiality. John R. Wright

⁴⁰ Ibid, p. 344

characterizes Murdoch as holding a form of moral realism, but, a realism that “signifies a kind of serious engagement with the individuals we encounter.” This serious engagement requires the freeing of the individual of her own blinding interests, but also makes on the agent demands that are incompatible with impartiality, since such engagement may necessitate a development “toward idiosyncrasy as we encounter individuals who challenge us to see them as embodying” the qualities emerging from the just and loving gaze.⁴¹ We often then have to be partial—understood as non-objective—in order to see individuals the way they are. Thus Murdoch’s claim that the highest love is impersonal is not equivalent to saying that it is impartial. Clearly, Murdoch is setting out to show something different from what Velleman thinks she is.

Velleman’s mischaracterization of Murdoch’s project leads to the second reason why giving so little weight to Murdoch’s claim about impersonality is problematic: it fails to recognize that Murdochian love is not just a variation of what we might call the standard conception of love. An impersonal love will be at odds with such a conception. Based on her description, we have a *prima facie* reason for not reading her as offering a standard conception. Generally love is construed as (a) an emotion that is (b) other-directed.⁴² But for Murdoch, love is constitutively bound up with the act of unselfing, the breaking free of personal fantasy, and engaging in attention aimed at the world. The highest love is impersonal, because, when one’s heart is full of love, it’s empty of the irrational concerns for the self. Thus even if attention is aimed *outward*, love is actually directed at *the self*. More specifically, it’s an *attitude* toward the self—namely, that the self doesn’t deserve the amount of concern we naturally feel toward it. Even if we approach the world differently as a result of developing love, strictly speaking,

⁴¹ John R. Wright, “Transcendence Without Reality,” *Philosophy*, Vol. 80, No. 3, (2005), p. 378.

⁴² A worry might arise here that my characterization of the standard conception of love doesn’t allow for self-love; thus (b) would not be correct. I will assume nonetheless that love should be taken to be other-directed with the condition that, if self-love is genuinely possible, then the claim could be weakened to say that the *paradigmatic examples* are of love as an emotion directed outward, away from the self

Murdoch's love is the recognition that the ego's hold on us must be loosened and eventually removed, and the further realization that, in spite of the ego's natural grip upon us, our interests should not necessarily be ranked highest in our hierarchy of values. Thus what distinguishes genuine love from similar but distinct states is this attitude toward the self.

There is a flip side to this attitude, of course, which might restore some of the hominess of the standard conception to Murdoch's conception of love. My recognizing that my interests should not automatically be ranked highest in my hierarchy of values allows me to prioritize the interests of others, when doing so is called for. Thus I will respond to the suffering of others more appropriately, when I can ascribe the right priority to their suffering. In some sense then love is directed outward—but only as a result of the unselfing that must come first, according to Murdoch.

If this is the right way of understanding Murdoch's notion of love, there's no special problem of partiality here. While Murdoch is not trying to provide a reconceptualization of love that includes partiality, as I argued earlier, the worry that Velleman has that Murdoch's love unjustifiably licenses the favoring of a particular individual over another isn't well-placed. Murdochian love is constituted by the ability to de-emphasize one's own interests, so that one is able to see clearly the interests of others—presumably every person with whom one comes in contact, unless there are other reasons which are unrelated to the ego's hold on the agent. We will, of course, treat some people differently, evident in Bernard Williams's example of the man who saves his wife rather than a stranger, when he can save only one of the drowning people. This isn't a problem for Murdoch, however, since seeing others clearly will allow us also to see clearly the demands put upon us by certain relationships of which we are a part. This might be a problem for an account that argues that standard love is a moral emotion, but this is not the kind

of account Murdoch is offering. Since Murdoch's love prepares the moral agent to really see others, no matter who these others are or what relationship they bear to us, there is no problem of partiality—or no special problem for the account, in spite of what Velleman claims.

As I said above, I think that Murdoch's moral perception is best characterized as weak moral perception. There are places where she may be read as offering a strong moral perception theory, but the reason for this is that she believes so much work is done by seeing the world aright that we *almost* have nothing left to do when the time to choose arrives. Consider the following: "It is what lies behind and in between actions and prompts them that is important, and it is this area which should be purified. By the time the moment of choice has arrived the quality of attention has probably determined the nature of the act."⁴³ Notice that the moment of choice still comes *after* attention has done its work, which suggests that it's still preparatory, even if there is really very little for the other faculties of the agent to do other than complete the job already begun by attention. "Action is an occasion for grace," she says, "or for its opposite," meaning that, even if moral perception has determined the nature of the act, the choice still has to be made to respond to the information provided by the perception, in which case moral judgment and moral choice are, for Murdoch, not just moral perception.⁴⁴ She's thus offering a weak conception of moral perception.

We might stop here and begin trying to determine what the right kind of motive would be in order to accommodate Murdoch's plausible claims, since she never tells us exactly how love and attention are to be realized in our motivations. Before doing so, however, I would like to offer up a different role that perception plays in morality, one that has heretofore gone unnoticed by moral perception theorists, but that is essential to Murdochian moral perception.

⁴³ Murdoch, *Sovereignty of Good*, pp. 66-67.

⁴⁴ *Ibid*, p. 71.

4. Building Up Structures of Value: Axiological Perception

We've seen that Murdochian moral perception relies essentially on what Murdoch calls love—i.e., the recognition that one's own interests should not be prioritized over the interests of others for the sole reason that they are one's own interests. Displaying love allows a moral agent to properly place her own and others' interests in the hierarchy of importance, such that she wouldn't lament the destruction of a town merely because she won't get to visit it ever again. Thus moral perception allows a proper weighting of the interests of all those included in the situation before the agent.

But perception has a role to play prior to this moment of weighting interests. In order to place interests in their proper hierarchy, the agent has to know what the interests of those around her are. Moral perception therefore presupposes the ability to determine these interests and this too is a matter for perception—though not, strictly speaking, moral perception. Recall the first two steps of Blum's schematic:

Step 1: Accurately recognize the features of the situation.

Step 2: Recognize the features of the already characterized situation as morally significant. (Moment of moral perception)

We can see that in step 1 the perception involved is not moral. Accurate perception of the situation is merely the gathering of sensory data, cognized correctly. Correct cognizance of the sensory data presupposes the conceptualization of the raw information, such that a chair is perceived as a chair and not just a rectangular patch of brown, with jagged corners and planar

surface, and so on with other objects. If this were the full extent of perception's part in morality, minimalism would be correct. But we've seen Blum and Murdoch argue that perception extends beyond step 1, both arguing explicitly that moral perception comes early on in the process, at step 2. From this characterization it follows that non-moral perception happens in step 1 and moral perception happens in step 2, and if there is any role remaining for perception, it happens later in the process than step 2 and would most likely be strong moral perception. I will argue, however, that missing between steps 1 and 2 is an intermediate step—one essential to the later steps involving a kind of perception that, for reasons I will give shortly, can be called *axiological perception*.

As it stands step 1 is uncontroversial, since there must be a point at which the agent comes across an “inchoate” situation—i.e., “not even yet a distinct ‘situation’”—requiring some cognitive and perceptual processing in order just to recognize it *as* a situation. At this point the perception is non-moral; if there is any value-attribution here, it involves instrumental value, aimed at organismic survival. But Blum further describes the first step as involving the taking of the situation “to be [...] of a certain character.”⁴⁵ This “certain character” gets taken up later in the process in step 2 when a determination is made about which features are morally salient, but taking the situation to be of a certain character extends beyond instrumental value-attribution, yet involves a kind of value-attribution that doesn't have to do with moral value. What is going on then between steps 1 and 2?

Murdoch introduces a metaphor that can help us get clear on the intermediate step, writing that the “work of attention [...] imperceptibly [...] builds up structures of value round about us.”⁴⁶ She says this in order to show that “moral life [...] is something that goes on

⁴⁵ Blum, *Moral Perception and Particularity*, p. 58.

⁴⁶ Murdoch, *Sovereignty of Good*, p. 37.

continually, not something that is switched off in between the occurrence of explicit moral choices,⁴⁷ but the metaphor of building a structure of value allows us to think of moral judgments as the capstone constructed atop the foundational elements and to recognize how hugely important these foundational elements are to what emerges at the top. It's at the base of the structure that non-moral value is found—the instrumental value pertinent to the survival of the organism, but also *what is good* for the constituents of the situation. In other words, at or near the foundation is where the interests of others and the agent herself are determined, and the agent must get this right, lest the whole structure run to ruin when we move toward the top. Determining these interests requires perception as well—what we can call *axiological* perception.⁴⁸

If we understand strong moral perception as requiring *seeing-that*—i.e., seeing (in the semi-metaphorical sense moral perception theories intend⁴⁹) that an act being performed is right or wrong or that a certain act should be performed—and weak moral perception as requiring only *seeing-as*—i.e., seeing the morally significant features of a situation *as* morally significant—we might characterize axiological perception as *seeing-in*—i.e., seeing in a certain situation certain features relevant to making an evaluative, yet not moral, judgment. In this sense seeing-in goes beyond the making of an inchoate situation something cognizable. Thus two people might both be said to properly perceive a situation, while one fails to see in the situation value that the other does see.

⁴⁷ Ibid, p. 37.

⁴⁸ I hope it's clear in the text that my calling this kind of perception axiological doesn't imply that the perception occurring in step 1 is non-axiological, since I have explicitly recognized that instrumental value is determined at this stage.

⁴⁹ Though see Robert Audi, *Moral Perception*, for an account that takes moral perception in a literal sense—that is, literally perceiving moral properties. See also Justin McBrayer, "Moral Perception and the Causal Objection," for a discussion of what is called the causal objection to moral perception, which says that moral perception presupposes direct causal contact with moral properties, a claim some find dubious.

Recall Blum's John and Joan example. Arguably both recognize that it would be good for the woman not to have to stand, just as they both also recognize that it's good for themselves to stay seated. It's in the woman's interest to attain a seat and it's in their interests to keep their seats. Each of them has performed an adequate piece of axiological perception, even if Joan is able to perform weak moral perception while John isn't. But consider a third seated person, Janine, who is so self-absorbed that she cannot even recognize that there is a person in front of her who has an interest in sitting. It's completely beyond her ability to see the woman with the cumbersome bags and to realize that sitting would ease the woman's burden. In fact, she fails even to recognize that the woman is burdened. While John is morally culpable because he never acknowledges that the woman has a claim on his consideration, Janine cannot even recognize the woman as an agent with interests at all. Or, if she claims to recognize the woman's interests, it's not a genuine recognition, because her fixation on her own interests obscures, and perhaps even prevents, her vision of the woman as an autonomous agent with beliefs and desires of her own. John has one step to make in order to overcome his moral failings, while Janine has two—seeing what is good for those around her and then seeing how she might be obligated to contribute to the realization of what is good for those around her.

Why might Janine be in this position? She's isn't deaf or blind. Presumably she sees the packed subway just as John and Joan do; she sees the woman laden with bags and she must understand that the reason the woman isn't seated because there are no open seats. So what explains why John and Joan do see that the woman has an interest in gaining a seat, while Janine doesn't? I believe that what distinguishes John and Joan from Janine is that John and Joan perceive the situation from under a different *gestalt* from how Janine perceives it. That is, while all three individuals can be said to perceive the situation, Janine perceives it with certain

elements foregrounded and others backgrounded, and one of the elements backgrounded is the woman holding the bags. Perhaps in Janine's perception of the situation, what is in the foreground is the speed of the subway, since she is late for an appointment; or the smell of those around her, since she is feeling slightly nauseated by the prevailing odor; or the relative comfort she is feeling while seated, since she rarely is able to acquire a seat on the subway. With each of these, what is foregrounded is her own interests and how those interests are being satisfied or not. What is in the background is the interests of others. If they come into the picture at all, they are there as rival interests, those that stand to thwart Janine's own desires and preferences. And since Janine cannot recognize the interests of others strictly *as* interests belonging to these people, she is completely unable to see what is good for them. She fails to see in the situation value that John and Joan see, which is a failure of *seeing-in*, and thus a failure of axiological perception.

An aesthetic analogy might help us see this more clearly. Think about how two people could view the same painting, successfully seeing all of the constituents in the tableau, but one might see-in the painting more than the other. For example, if I am looking at a portrait of Winston Churchill, but I am unaware that the subject is Churchill, I might be said to be seeing Churchill, but not Churchill in the painting. Moreover, I certainly won't see in the painting the former prime minister of Great Britain, since doing so would require that I see Churchill. A friend of mine, aware that the subject is Churchill and that he was prime minister of Great Britain, would see in the painting the prime minister. In one sense we are both perceiving the same painting, but in a second sense, my friend is perceiving more than I am.

The seeing-in that I am describing is similar to Richard Wollheim's notion of seeing-in.⁵⁰ For Wollheim, however, seeing-in is constitutive of what he calls representational seeing—that is, the kind of seeing that we utilize when viewing visual representations, like paintings and photographs. However, insofar as I am suggesting that seeing-in happens in straightforward perception, my notion is significantly different from Wollheim's. There is one point of contact between Wollheim's seeing-in and my own, evident in his characterizing seeing-in as deriving from “a special perceptual capacity” to have “perceptual experiences of things that are not present to the senses: that is to say, both of things that are absent and also of things that are non-existent.”⁵¹ One sees more than what is there to be seen—that is, as one literally sees what is there, one non-literally sees what is not there to be literally seen—or as Wollheim puts it, “a state of affairs can be seen in a particular.”⁵²

In order to bring this point home, consider a discussion by Noël Carroll of the Bruegel painting *Landscape with the Fall of Icarus*.⁵³ Carroll is raising an objection to the formalist definition of art, the most famous of which belongs to Clive Bell, that says that what makes a work a work of art is its possession of *significant form*. On such a theory we identify works of art as art, not by attending to any content, but solely to the work's formal qualities—shape, texture, color in the case of visual art, rhyme, meter, assonance/alliteration in the case of literary art. Carroll worries that on the formalist definition, we would have to attend only to the formal qualities of the painting in order to recognize the work as art, but that in this painting, we cannot perceive the formal qualities without attending to the content, since the actual fall of Icarus cannot be seen in the painting without the viewer's detecting an errant leg in the water off to the

⁵⁰ “Seeing-As, Seeing-In, and Pictorial Representation,” in *Art and its Objects*, NY: Cambridge University Press, 1980, pp. 205-226.

⁵¹ *Ibid*, p. 217.

⁵² *Ibid*, p. 223.

⁵³ See Figure 1.1.

right and in the background. In the foreground of the painting are peasants performing their daily work, which suggests to Carroll that the “painting is expressive of insouciance,” since it’s an observation of the “irony of the way in which everyday life goes on quietly and imperviously, while legendary events occur unnoticed.”⁵⁴ But one must notice, Carroll says, that the “narrative center” is not the actual center of the painting in order to see how the background and foreground are the reverse of what they should be. What the Bruegel painting illustrates is how a viewer of the scene—the implied painter if you will—can perceive the scene completely and yet lack something essential to seeing in that scene something vital to seeing the scene in its entirety, because of a failure to see what is “not present to the senses” as Wollheim says.

Of course, in the Bruegel painting, inference plays a key role, too, since seeing the legs jutting out of the water isn’t seeing the fall of Icarus. One must still infer from the legs and the title of the painting to the fact that Icarus had just fallen from the sky. Axiological perception is not to be understood as involving inference, so the analogy fails at this point. But what we can take from this example is that it makes a difference to what one sees when one foregrounds certain elements and backgrounds others. This isn’t the same phenomenon as the aspect shift, made famous by Wittgenstein’s duck-rabbit, since it’s not a complete gestalt switch. If Janine were to properly foreground the interests of those around her on the subway, she wouldn’t see the situation as completely different in kind—but she would see it as possessing a “certain character” which was missing before the switch. Originally Janine sees x , y , and z in the situation before, with x and y in the foreground and z in the background, but if she eventually comes to see the situation as John and Joan see it, she might foreground y and z , seeing x in the background, which then opens her up to see a completely different element a , that was obscured because of the difference in what occupied her center of attention. And that a , in my example, is the

⁵⁴ *Philosophy of Art: A Contemporary Introduction*, NY: Routledge, 1999, p. 123.

interests of those around her. Once she makes the gestalt switch, she can recognize what is in the woman's interests, at which point she can either go on to see the woman as having a claim on her consideration (Joan) or not (John). But until she is able to see the woman has interests and what they are, she fails to see-in the situation the good of those around her, and she is thus completely unable to move on to step 2.

If I am right that perception has this additional role to play, then it's crucial that moral agents be able to perform both axiological and moral perception. Even on a view that holds that moral judgment isn't just moral perception—that is, a view not advocating strong moral perception—getting it right in steps 1, 2 and the intermediate step I have been describing would still be essential to those later steps. The structure of value that we build must be solidly constructed at its base. If there are problems at these levels, we're going to see problems when a moral decision is to be made. And when these problems present themselves, we'll see the deficit in caring discussed earlier. In other words, we are just now getting into view the explanation for the deficit both Warnock and Murdoch talk about.

Before we are able to complete the explanation, noting why the problem arises, we need to see what kind of motive would be necessary for an agent to successfully perform both axiological and moral perception. What kind of motive should an agent evince that would enable her both to know what the interests of others are and what weight to give those interests? In the next chapter we'll look at several candidates and see that one stands out as having the most promise.

Chapter 2

The Kind of Motive Necessary for Axiological and Moral Perception

In the last chapter I argued that a key component to moral reasoning is actualizing both axiological and moral perception so that the agent has knowledge of what the interests of others around the agent are and how to properly weight those interests. The next question we should try to answer is: what motive best accomplishes this? The right motive will require the right kind of attitude toward others, since we've seen that an attitude of pure self-interest tends to blind one to the needs of others. Therefore, the appropriate motive for a virtuous moral agent will have at its core an element enabling her to access those interests. We have to be careful on two fronts here, however: first, we must take steps not to overreach, since it's easy to confuse supererogatory motivations and actions for those incumbent on the virtuous agent; second, even if we lapse into talking about motives as if they're pure—e.g., acting from kindness or spite or generosity—they are usually complex and uneasily described. This latter point should compel us then to seek out not the right motive, but the right *kind* of motive. We'll look at three of the more compelling candidates: sympathy, altruism, and compassion.

1. Sympathy

The first candidate I want to consider is *sympathy*. The question to answer is whether acting from the motive of sympathy, as a general rule, would allow an agent to access and

properly weight the interests of others when the agent is making a moral decision. Right at the beginning, however, we have to recognize that this term, along with others to follow, are defined in a number of different, and sometimes mutually inconsistent, ways. At each step, we'll need to be clear just how we are defining the terms we use. In the case of sympathy, it's generally described as both *negative*—it's a response to an occurrence interfering with a particular pursuit of certain ends or causing a state of suffering—and *other-directed*—one feels sympathy for *others* or on their behalf because it's *their* pursuit of ends that has been thwarted or *their* suffering that is occurring. Peter Goldie defines “sympathy” as “emotional experience, involving recognizing another’s difficulties, and having feelings of distress about them, as well as being motivated to alleviate those difficulties in some way.”¹ Elliot Sober and David Sloan Wilson define it this way: : “*S* sympathizes with *O* precisely when *S* believes that something bad has happened to *O* and this causes *S* to feel bad for *O*.”²

On definitions such as Goldie’s and Sober & Wilson’s sympathy would be a moral motive when it results in *concern* for others. One prominent account of sympathy as a moral motive is that of Stephen Darwall, who treats “sympathy” as synonymous with “sympathetic concern,” defined as “a feeling or emotion that (a) responds to some apparent threat or obstacle to an individual's good or well-being, (b) has that individual himself as object, and (c) involves concern for him, and thus for his well-being, for his sake.”³ This characterization of sympathy treats (a) and (b) as instrumental to the primarily moral component found in (c). “Seeing the child on the verge of falling,” he writes, “one is concerned for his safety, not just for its (his

¹ “How We Think of Others’ Emotions,” *Mind & Language* Vol. 14 No. 4, 1999, p. 396.

² *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: Harvard University Press, 1998, p. 235.

³ “Empathy, Sympathy, Care,” *Philosophical Studies* Vol. 89 No. 2/3, p. 261.

safety's) sake, but for *his* sake. One is concerned for *him*. Sympathy for the child is a way of caring for (and about) him."⁴

If sympathy is the state we're looking for, it's going to be because of this notion of concern. In order to be able to determine what is good for those around one and when what is good for others should affect our own decisions, some concern will likely be necessary. Therefore, we should look more closely at Darwall's account to see if it fits the bill. He argues that the kind of concern I should have is for the *actual* welfare, and not simply the *perceived* welfare, of the other. This has two consequences. First, concern isn't person-relative. When I have concern for someone's actual welfare, this welfare presents itself as of "neutral disvalue." This means that, while I do feel bad for the person whose welfare has been diminished, I further judge that the world itself is worse off for having this diminishment. Thus everyone's welfare matters equally, which implies that our own welfare matters as well.⁵ Second, since we feel concern for others based on what is *actually* in their best interests, it's possible that we don't feel concern for someone whose occurrent desires have been thwarted, if such thwarting isn't in fact bad for her. A person's desires may be for things *not* in her best interest, so we have concern for the fulfillment of "what we or she would rationally desire *for her*, that is, *for her sake*."⁶ Note here, however, that this second consequence does not render superfluous the agent's being able to access the desires of the other. He will still need to be able to know what she desires, even if her desires are not consistent with what is really in her best interests.

⁴ Ibid, p. 261.

⁵ Note that Darwall's account of sympathetic concern ends up where an ethical egoist account might end up—with a recognition that everyone's welfare matters. But an ethical egoist would reason, first from the value of his own welfare, and, after recognizing how his own welfare could be augmented by the augmenting of the welfare of others, then to a recognition of the value of the welfare of others. Darwall's ordering is exactly the opposite: by recognizing the value of the welfare of others, the agent comes to recognize the value of his own.

⁶ Darwall, "Empathy, Sympathy, Care," p. 277.

Darwall's notion of sympathetic concern is appealing, involving some of the elements discussed in Chapter 1. There is a gap, however, between sympathetic concern and the motive we are seeking—that gap has to do with knowledge. In contrast to sympathy Darwall describes a mental state called “proto-sympathetic empathy.” Like “sympathy,” “empathy” has been defined in a number of divergent ways. Some definitions treat empathy as a sharing of emotion, some as a sharing of perspective, and others as some combination of both. Some theorists, like Darwall, respond to the divergent definitions by positing *kinds* of empathy.⁷ The kind of empathy marked by perspective-taking Darwall calls “projective empathy.” Because “projective empathy is a projection into the other's standpoint,” proto-sympathetic empathy is not genuine sympathy, since “[a]ttention is focused, not on the other, but on her situation as we imagine she sees it, or as we think she should see it.”⁸ At this point it's purely imagining what *a* person would experience in that situation, and not what *that* person is experiencing in her situation. Also, at this stage, it's not concern either. This state is *proto-sympathetic* empathy “because it brings the other's relation into view in [a] way that *can* engage sympathy on his behalf.”⁹ Of course, as “projective empathy” lacks concern, nothing follows about how the results of such imagining get implemented. “Someone in the grip of resentment, envy, or the desire for revenge may take delight in the vivid appreciation of another's plight he gets from imagining what another's situation must be like for her.”¹⁰ Thus projective empathy on its own is not sufficient for good moral reasoning, as it can be used for evil purposes as well. Proto-sympathetic empathy can lead to sympathy nevertheless. For example, a person may feel *empathic distress*,¹¹ and consequently

⁷ See my “Closing In on Empathy: A New Model” for a fuller discussion of defining “empathy.”

⁸ Darwall, “Empathy, Sympathy, Care,” p. 270.

⁹ Ibid, p. 271, my emphasis.

¹⁰ Ibid, p. 272.

¹¹ This term appears in Martin Hoffman, *Empathy and Moral Development: Implications for Caring and Justice*, Cambridge: Cambridge University Press, 2000.

take steps to address the needs of the other person in order to diminish the empathic distress. But as Daniel Batson¹² points out, what makes empathic distress painful is the “aversive arousal” resulting from seeing the person in need. It’s just as effective, and often easier, to remedy empathic distress, not by addressing the need, but by escaping the situation in which one is confronted with the need. Thus without concern from the beginning, nothing guarantees that the agent will be motivated to do anything for those in need. Proto-sympathetic empathy is insufficient for morality because it lacks concern, but Darwall doesn’t just add concern to proto-sympathetic empathy in order to get sympathetic concern. This would be empathetic concern. Therefore, while projective empathy is a possible route to sympathetic concern, it isn’t a necessary route as Darwall sees it.

This feature of Darwall’s account, I suggest, presents him with a problem. If sympathy is an emotional response to a threat to another person on that person’s behalf, and involves concern for the *actual* best interests of the other, there needs to be a reliable mechanism for determining what is in the actual best interests of the other. This is what we want in the motive we are seeking. Nowhere in his description, however, is there a mechanism that can *accurately* determine what is in the actual best interests of the other. He rules out empathy, because it’s based on *imagining* how the other person sees her situation or gives us an idea of how she *should* see it. This presupposes something about the nature of empathy, which is that there is no guarantee that it gives us an insight into how the other person *actually* sees the situation, yet sympathy doesn’t seem to be able to achieve this purpose either. Sympathetic concern alone cannot do the job. In order to determine the interests of others, we generally need to have information about their occurrent mental states, but sympathetic concern cannot give us this information. Moreover, it’s not clear why empathy, *combined with concern*, couldn’t give a

¹² *Altruism in Humans*, NY: Oxford University Press, 2011.

fairly accurate picture of how the other is seeing her situation. Therefore, if we add to Darwall's account projective empathy, we would have at least a mechanism for providing the requisite knowledge, whereas in his own account, none is evident. Once we do this, however, it's no longer clear what role sympathy plays. Unless Darwall would respond by saying that *empathetic concern* is insufficient for moral reasoning while *sympathetic concern* is—and reasons for making such a claim are not apparent to me—then the role played by sympathy in his account could be played by empathy and sympathy would be otiose. I propose, therefore, that we reject sympathy as the motive necessary for actualizing axiological and moral perception, and pursue the possibility of empathetic concern instead.

2. Altruism

Darwall's account shows us that sympathy, understood in the traditional way, fails to provide the right kind of access necessary for determining and weighting the interests of others, but it also shows us that if we combine, not sympathy, but empathy and concern, we might be on the right track. Batson's account makes central what he calls "empathic concern." In this section we'll consider Batson's account, but we will see that ultimately this account values actions from *altruism*, which will turn out to be too strong a requirement.

Batson characterizes empathic concern as "other-oriented emotion elicited by and congruent with the perceived welfare of someone in need."¹³ This definition blends the epistemic value of empathy with the moral value of concern, resulting in what Batson argues is the moral

¹³ Ibid, p. 11.

value of empathic concern. Jesse Prinz has criticized this feature of the account, arguing that it claims to “show the value of empathy,” when it “may really show [only] the value of concern.”¹⁴ But as we’ve seen, the difference of *kinds* of concern can be relevant, as sympathetic concern differs from empathic concern, and is unsatisfactory because it lacks epistemic justification. It seems to matter then that Batson investigates not just concern, but concern backed by a purportedly reliable epistemic mechanism.

Empathy isn’t always treated as an epistemic mechanism, however. As I mentioned before, “empathy” too possesses a large number of divergent descriptions, to which some, like Darwall, respond by positing kinds of empathy. Batson discusses several different conceptions of empathy in modern use:

1. Knowing another person’s internal state, including his or her thoughts and feelings
2. Adopting the posture or matching the neural response of an observed other
3. Coming to feel as another person feels
4. Intuiting or projecting oneself into another’s situation
5. Imagining how another is thinking and feeling
6. Imagining how one would think and feel in the other’s place
7. Feeling distress at witnessing another person’s suffering¹⁵

On conception (1) empathy is straightforwardly an epistemic state, while on conception (3) it’s straightforwardly emotional. The co-existence of these two conceptions has prompted some theorists to differentiate broadly between two different kinds of empathy called cognitive and

¹⁴ “Is Empathy Necessary for Morality?” in *Empathy: Philosophical and Psychological Perspectives*, ed. Amy Coplan and Peter Goldie, NY: Oxford University Press, 2011, p. 212.

¹⁵ Batson, *Altruism in Humans*, pp. 12-19.

affective empathy respectively, found in Hoffman's description of *cognitive* empathy as "the cognitive awareness of another person's internal states, that is, his thoughts, feelings, perceptions, and intentions;" and *affective* empathy as "the vicarious affective response to another person."¹⁶ As we move further down the list, however, we get away from the conceptions that stress one or the other of the cognitive or affective elements and more toward conceptions blending the two, such that empathy is conceived of as both an imaginative projection and an emotional response. By the time we reach conception (7) we also have the notion of vicarious distress, which is close to the notion of concern, while still leaving room for the escape option found in aversive arousal discussed earlier. It's only with Batson's description of empathic concern that we get all three: an imaginative projection into the perspective of another, congruent emotional states, and concern. It's thus concern that results from the empathic sharing, and not *just* concern, that matters to moral reasoning, *pace* Prinz.

Like Darwall's account, Batson's account of empathic concern initially looks promising. When we see the role Batson attributes to empathic concern, however, it's clear that he expects too much of moral agents, at least in the context within which we are working. For Batson empathic concern on its own is only instrumentally valuable, fitting into the "empathy-altruism hypothesis"—that is, the claim that empathy naturally results in altruistic behavior. Empathy is morally valuable only insofar as it results in altruistic helping behavior, itself the inherently valuable element. But is altruistic behavior the moral ideal? I will argue that it's not.

Darwall's sympathetic concern, while it might often result in behavior intended to address the needs of others, is supposed to be morally valuable on its own, as a motive. Batson's empathic concern, however, is valued because of its naturally resulting in helping behavior. One reason for this is that Batson isn't constructing a normative account, but a descriptive one,

¹⁶ *Empathy and Moral Development*, p. 29.

opposed to other descriptive accounts concluding that humans are naturally egoistic—the view called *psychological egoism*. These descriptive accounts relate to normative accounts by virtue of the ought-can implication thesis. If ought implies can and if people are naturally unable to act out of the interests of others (psychological egoism), then people cannot be morally obligated to act out of the interests of others. The way to *ethical egoism*—the view that people *should* act only out of their own interests—is thus opened up by the truth of psychological egoism. Rejecting psychological egoism doesn't require accepting psychological altruism, however, if the latter is construed as the view that people *always* act out of the interests of others, but one doesn't have to be this kind of psychological altruist to accept *ethical altruism*, the view that people *should* act out of the interests of others. Once one accepts that people *can* act out of the interests of others, the way is thus opened to ethical altruism. Of course, the falsity of psychological egoism also makes possible views less extreme than ethical altruism, views that conclude that in *some* instances people should act out of the interests of others. Batson's descriptive account is intended to refute psychological egoism, but it's not aimed at establishing psychological altruism. His burden is to show only that altruistic behavior results from capacities that we already have, one of which is the capacity to feel empathic concern. The empathy-altruism hypothesis is still a strong claim, however, as we can see in the following statement: "The hypothesis states that feeling other-oriented emotion elicited by and congruent with the perceived welfare of another person in need (i.e., empathic concern) produces a motivational state with the ultimate goal of increasing that person's welfare by having the empathy-inducing need removed (i.e., altruistic motivation)."¹⁷ Batson claims that empathic concern is *one*, but not the *only*, means to altruistic motivation.

¹⁷ Batson, *Altruism in Humans*, p. 29.

Since it's altruism, and not empathic concern, that has intrinsic value for Batson, we need to be more precise about what altruism is. Batson defines "altruism" as "a motivational state with the ultimate goal of increasing another's welfare," in contrast to egoism, which is "a motivational state with the ultimate goal of increasing one's own welfare."¹⁸ This conception of altruism is mirrored in Sober and Wilson's characterization of "psychological altruism" as "car[ing] about the welfare of others as an end in itself." That is, "[a]ltruists have irreducible other-directed ends."¹⁹ Sober and Wilson's characterization of altruism is contrasted with two other states: (1) psychological egoism, which they define as the view "that the only ultimate goals an individual has are *self-directed*; people desire their own well-being, and nothing else, as an end in itself;"²⁰ and (2) evolutionary altruism, which is the claim that throughout evolution, organisms displayed behavior that "increases the fitness of others and decreases the fitness of the actor."²¹ They are careful to distinguish psychological and evolutionary altruism, since the latter claims only that actors displayed such behavior, "regardless of how, or even whether, they think or feel about the action." On these two characterizations of altruism, altruistically motivated behavior occurs when someone acts strictly out of the consciously recognized interests of others, with the welfare of others as his ultimate end. Batson and Sober and Wilson are not suggesting that because people are able to behave altruistically, they *ought* to do so; thus they are not arguing for ethical altruism. Defending ethical altruism would be very difficult. Even if psychological altruism is distinguished from evolutionary altruism primarily by the fact that in the former no personal cost is *necessary* to such behavior, while it is in the latter case (at least extended to the species

¹⁸ Ibid, p. 20.

¹⁹ *Unto Others*, p. 228.

²⁰ Ibid, p. 224.

²¹ Ibid, p. 17.

overall), in *many* cases, acting altruistically will require acting *against* the interests of the agent. In some of these cases, incurring such costs will be *unjust*.

Consider the following example:²² A farmer has worked all winter long preparing his fields, all spring planting crops, all summer tending and watering the plants, and ultimately stands to reap the rewards of his labor. During this same time, his neighbor, another farmer, ignores the preparatory steps and fails to plant his crops. Then, when the time comes for reaping, he has nothing, and is consequently hungry and in need. While the industrious farmer may be moved to share his harvest with the lazy farmer, he isn't morally obligated to do so. We may praise the industrious farmer for sharing, but we wouldn't blame him for not doing so.²³ This shows that altruism, rather than being morally obligatory, is supererogatory—altruistic behavior is morally good, but not morally required. Thus ethical altruism is false. But if exhibiting altruistic behavior is not morally obligatory, then, while altruistic motivation would be good to have, it's not the motive that virtuous moral agents generally *ought* to have. And if empathic concern is valuable only as a means to altruistic behavior, as Batson has it, then it's something that would be good to have, but isn't morally required. Thus empathic concern, as a means to altruism, isn't the motive we are after. We want a motive that generally holds for the agent. On Batson's account, having empathic concern is, *ceteris paribus*, better than not having it; but not

²² This example appears in John Arthur, "Famine Relief and the Moral Code," in *Ethics: History, Theory, and Contemporary Issues*, ed. Steven M. Cahn and Peter Markie, NY: Oxford University Press, 2009, pp. 829-840.

²³ Of course, there are situations in which we might accuse the industrious farmer for being heartless or cruel. Suppose that the lazy farmer and his family are visible to the industrious farmer, and day after day, the lazy farmer's family gets thinner and thinner, weaker and weaker, and eventually dies of starvation. At this point, we might blame the industrious farmer for not sharing, but this is probably because of the hard-heartedness needed for someone to watch a person in need slowly die before his very eyes when he could have prevented it. Even here, however, the blame for the situation is still placed on the lazy farmer since he put himself and his family in it to begin with.

having it isn't a moral failing. Therefore, empathic concern as a motive cannot be what we are after either.²⁴

3. Compassion

The reason Batson's account cannot give us what we need is that it treats altruism as an inherently valuable motive, and, while sometimes altruism is the kind of motive an agent should possess, it's not the kind of motive that would actualize axiological and moral perception, since we are after a kind of motive that an agent should *generally* possess. But we looked closely at Batson's account because it seemed to capture part of what Darwall's account was missing: the prioritizing of the agent's ability to access the interests of the other person. This is why a combination of empathy and concern seemed more promising. In this section I will consider the motive of *compassion*. The notion of concern is not identical to, but is closely related to, caring. *Prima facie* empathy combined with concern and/or caring gets us a state like compassion, but we will see that some conceptual work will have to be done in order to give a satisfactory account.

As with other terms discussed before, "compassion" is defined in a number of different ways. For example, Martha Nussbaum describes the compassionate person as acquiring "motivations to help the person for whom she has compassion," because she will have concern that "is motivated or supported by the thought that one might oneself be, one day, in that

²⁴ Michael Slote uses Batson's empathic concern in related ways. Aside from his advocating along with Batson the claim that empathy often results in helping behavior, he argues in *The Ethics of Care and Empathy* (NY: Routledge, 2007) that empathy can help in moral evaluation and in *Moral Sentimentalism* (NY: Oxford University Press, 2010) that empathy can help us make meta-ethical determinations of what constitutes moral approval and disapproval. I don't address these accounts here because they don't directly relate to motivation in the way Batson's account does.

person's position."²⁵ I find this conception of compassion unduly egoistic, however, since compassion is often a state people find themselves in when it's practically impossible that they will ever be in a situation similar to the target's. Moreover, compassion as a motive is often praised precisely because it's selfless, as occurs when we generally approve of those working with populations of the suffering and underprivileged by describing them as "compassionate." Therefore, whatever else compassion is, at this point it seems to me that it would imply a *lack of* egoism, which runs counter to Nussbaum's conception.

This opposition of compassion to egoism naturally prompts us to consider what may be the most famous account prioritizing compassion as a motive: that of Arthur Schopenhauer. While the metaphysical picture that Schopenhauer presents in his *The World as Will and Representation* is almost universally rejected, I will argue that the account of compassion that appears there and in his less controversial *On the Basis of Morals*²⁶ can, when amended, give us what we need.

Schopenhauer sets out to give a descriptive moral theory, but part of our reformulation of his account will require normativizing it. Nevertheless, we can evaluate some of his descriptive claims first, in order to see whether they hold water. He contrasts compassion (*mitleid*) with two other motives, the "anti-moral" motives egoism and malice. Egoism is marked by the seeking of the benefit of the self, while malice is the seeking of the harm to others, even at the cost of causing harm to self, and compassion is the seeking of the benefit of the other. He claims that all actions are motivated by one of these and calls egoism and malice anti-moral because they don't have positive moral worth, while compassion does. Egoism has no positive moral worth because it has no moral worth at all, while malice has only negative moral worth. Morally good actions

²⁵ *Upheavals of Thought*, NY: Cambridge University Press, 2003, p. 342.

²⁶ In *The Two Fundamental Problems of Ethics*, trans. David E. Cartwright and Edward E. Erdmann, NY: Oxford University Press, 2010, pp. 121-272.

are those motivated by compassion, morally evil actions motivated by malice, and morally *worthless*—not morally bad, but of no moral worth at all—actions motivated by egoism. These latter actions are the most common, according to Schopenhauer, echoing claims made by Warnock and Murdoch. He describes egoism further:

Egoism by its nature is boundless: the human wills unconditionally to preserve his existence; wills it unconditionally free of pains, among which are included all want and privation; wills the greatest possible amount of well-being, and wills every pleasure of which he is capable, even seeks wherever possible to develop new capacities for pleasure. Everything which opposes the striving of his egoism provokes his animosity, anger, hate: he will seek to annihilate it as his enemy. He wills to enjoy everything possible, have everything; since, however, this is impossible, at least to be master of everything: ‘everything for me, and nothing for others’ is his motto. Egoism is colossal: it towers over the world.²⁷

Since egoistic actions lack moral worth, the fact that most actions fall into this category doesn’t imply that the world is full of evil actions. On the contrary, most actions for Schopenhauer are simply morally worthless.

At this point we might be worried about his descriptive claim that all actions can be said to derive from the motives of either compassion, egoism, or malice. As I noted at the beginning of the chapter, motives are rarely pure, so it’s doubtful that every motive can be described in one of these three ways. One might try to read Schopenhauer as claiming only that all *morally evaluable* actions derive from one of these motives, but this would conflict with his contention that egoistic actions are morally worthless, rather than of negative moral worth; thus this reading fails. It’s hard to see how we can accept this descriptive claim, but we might not find that terribly

²⁷ Ibid, p. 202.

troubling, since we aren't looking to Schopenhauer for a descriptive account, but rather a framework for a normative account. Unfortunately, his claim that egoistic actions are not of negative moral worth conflicts with what I argued in chapter 1, which is that egoism is a moral vice—i.e., of negative moral worth. But even independent of my argument, his claim about egoism is false on its face, as he admits that egoism produces the following traits: greed, gluttony, lust, self-interest, avarice, covetousness, injustice, hard-heartedness, pride, and arrogance. At least some of these are vices, which makes Schopenhauer's insistence that egoism is of no moral worth all the more mysterious. For this reason, I suggest we excise this feature of his account. Once we've done that, we can return to the claim about every action's deriving from either compassion or egoism or malice and narrow it to imply only that every morally evaluable action comes from one of these motives in order to give it more plausibility. So amended the new account would say that it is of great moral importance whether a person is motivated primarily by compassion, malice, or egoism, with the implication that a malicious person is morally worse than an egoistic person, even if an egoistic person does not possess a good moral character.

After making this first amendment to Schopenhauer's account, we can move on to see what he thinks constitutes acting from compassion, egoism, or malice, and what justifies acting from compassion, rather than from the other two. In arguing for the negative moral worth of malice, he asks what is the most "devilish" kind of action, suggesting that, while actions from envy run a strong second, arguably the worst kind of action results from *Schadenfreude*—the pleasure in seeing and/or participating in the suffering of another.²⁸ "There is no more unfailling sign," he says, "of a thoroughly bad heart and profound moral baseness than an inclination to

²⁸ He pokes fun at Kant's suggestion that lying to the crazed killer is as morally bad as malicious actions motivated by *Schadenfreude*. It is an absurd picture, indeed: putting the liar who wants to save the potential victim from murder on a moral par with the person who not only tells the killer where the potential victim is, but pulls up a chair to delight in the ensuing bloodshed. But perhaps the comparison is unduly stark.

pure, heartfelt *Schadenfreude*.”²⁹ When we compare actions performed from *Schadenfreude* to egoistic actions, we see at least the one difference that the latter “*can* lead to crimes and misdeeds of all sorts, but the misfortune and pain it causes another is merely a means and not an end,” while the former take “another’s suffering and pains [as] an end in itself the attainment of which is a delight.”³⁰ For Schopenhauer egoistic actions have no moral worth, positive or negative, because they are *self*-directed. Malicious actions have moral worth because they are *other*-directed, but their worth is *negative* because they are directed at the augmenting and delighting in the suffering of others. Compassionate actions too have moral worth because they are other-directed, and positive moral worth because they lament and seek to ameliorate the suffering of others. The reason for this taxonomy is that Schopenhauer believes there is a central moral principle almost all people hold: *Harm no one; on the contrary, try to help people as much as possible*.³¹ This principle reflects two virtues: *harming no one* corresponds to the virtue of justice; *trying to help others as much as possible* to the virtue of loving kindness.³² Even on our adaptation of his account, a person acting purely from egoism would be unable to embody either virtue, insofar as harming is intentional and trying to help people is selfless.³³ Schopenhauer believes that intuitively we all hold the central moral principle to ground moral behavior, but that for most of us this intuition is likely not grounded by any fundamental principle. He believes that it does have a metaphysical justification, however, and it’s found in his own system constructed in *The World as Will and Representation*, which describes the world of concrete individual particulars as being illusory, in actuality only a manifestation of one ultimately real underlying

²⁹ Schopenhauer, *On the Basis of Morals*, p. 205.

³⁰ *Ibid*, p. 206, my emphasis.

³¹ *Ibid*, p. 216.

³² *Ibid*, p. 216.

³³ It’s probably possible for someone to try to help people for egoistic reasons, but I am not going to explore that possibility here. If it turned out to be the case, we could amend the central moral principle so that the virtue of loving kindness corresponded to selflessly trying to help people or something similar.

substance: Will. We cannot go into a discussion of Will and its relation to the world of representation, but the general point is that individual persons are manifestations of the one substance Will, so that another person's suffering is ultimately everyone's suffering. There is no metaphysical distinction between me and the other; truly seeing the other is seeing "I once more." Thus this person's suffering is my own suffering, which logically implies that I shouldn't add to it (virtue of justice) and I should do what I can to ameliorate it (virtue of loving kindness).

We will have to reject the metaphysical basis that Schopenhauer posits, since, among other reasons, it requires accepting his dualism between noumenon and phenomenon, adopted and adapted from Kant's system as found in the first critique. But the basic point fits in with what we found appealing in both Darwall's and Batson's accounts. While the other's suffering is supposed to be my suffering, what counts is the amelioration of suffering *simpliciter*, whoever it is.³⁴ This mirrors the feature of Darwall's sympathetic concern that suffering is of "neutral disvalue," meaning that suffering is not person-relative. It also guards against the phenomenon of aversive arousal, as discussed by Batson, a purely egoistic strategy for removing the empathic distress. Schopenhauer's claim isn't that because the other's suffering causes me to suffer, I should address the other's suffering. It's because I recognize the other's suffering as having no distinction from my own suffering that I have as much reason to address it as I have to address

³⁴ Does this imply that *any being's* suffering should be ameliorated, such as the suffering of animals? I won't explore that issue here, but merely note that Schopenhauer is recognized to be one of the first western philosophers to give extended consideration to the question of what duties humans have to animals. Here's a choice quote:

[T]he above-mentioned capacity of animals to be more satisfied than we by mere existence is abused by the egotistic and heartless man, and is often exploited to such an extent that he allows them absolutely nothing but bare existence. For example, the bird that is organized to roam through half the world is confined to a cubic foot of space where it slowly pines to death and cries [...] and the highly intelligent dog, man's truest and most faithful friend, put on a chain by him! Never do I see such a dog without feeling of the deepest sympathy for him and of profound indignation against his master. I think with satisfaction of a case, reported some years ago in *The Times*, where a lord kept a large dog on a chain. One day as he was walking through the yard, he took it into his head to go and pat the dog, whereupon the animal tore his arm open from top to bottom, and quite right too! What he meant by this was: You are not my master, but my devil, who makes a hell of my brief existence! May this happen to all who chain up dogs. (*Parerga and Paralipomena*, Vol. 2, trans. E. F. J. Payne, NY: Oxford University Press, p. 297)

my own. Thus the general claim that suffering, whosever it is, is *bad* and should be squarely in the focus of our moral consideration is all that is necessary, and moves us away from the deeper metaphysical basis that Schopenhauer posits.

Our first amendment of Schopenhauer's account—i.e., the replacing of the descriptive claim that we take moral virtue to require acting from compassion with the normative claim that moral agents *should* act from compassion, rather than from egoism and malice—*normativizes* his non-normative account. Our second amendment—i.e., removing the metaphysical basis as justification and replacing it with the normative claims emerging from Darwall and Batson—*naturalizes* his account. There is one last amendment to make, having to do with the details of his very conception of compassion. After this last amendment is made, we will have what I take to be the kind of motive that can actualize axiological and moral perception, though I will leave the question unanswered as to whether the resulting motive is consistent enough with the spirit of Schopenhauer's account to be thought of as truly Schopenhauerian in nature.

The term Schopenhauer uses, that is almost universally translated into English as “compassion,” is *mitleid*, which literally means “suffering with.” He describes *mitleid* as “the quite immediate *participation*, independent of considerations of any sort, primarily in the *suffering* of another,” with the implication that “the prevention or removal of this suffering is that in which ultimately all satisfaction and all well-being and happiness consist.”³⁵ But given the general point considered earlier that what matters is the prevention and amelioration of suffering *simpliciter*, the “immediate participation” in the suffering of another will be self-undermining. This can best be illustrated by two objections to this account raised by Friedrich Nietzsche.

First, if only actions performed from *mitleid* have positive moral worth, then acting from *mitleid* is a moral ideal—the more of it, the better. But Nietzsche asks us to suppose that acting

³⁵ Schopenhauer, *On the Basis of Morals*, p. 213, emphasis in original.

from *mitleid* “was dominant even for a single day.” Would we find the world on this day morally ideal? It’s more likely that such a widespread experience of *mitleid* would be a nightmare, and “mankind would immediately perish of it.” (*Daybreak* 134)³⁶ On this one day, if Schopenhauer is correct, the world would be full of morally virtuous agents, and yet the world would contain *more* suffering in it than on any other day. Therefore, even if the end result is the amelioration of the suffering that exists before we experience *mitleid*, the stages through which we pass would actually increase suffering, since at this point we and the target of the *mitleid* would both be suffering. For a period the amount of suffering would be multiplied many times over and we would actually make the world worse before we made it better, a consequence best avoided if at all possible.

Nietzsche’s first objection then is that we add to the general amount of suffering when we experience *mitleid*, because upon such experience we suffer on the behalf of others, whereas beforehand, we didn’t. His second objection is that acting from *mitleid* can add to the general suffering by virtue of the target’s suffering *more* as a result of our experience of *mitleid*. If, for example, the sufferer feels *pitied* as a result of my expressing my state of *mitleid*, then she now suffers more than before my expression. Especially in those cases, which will undoubtedly exist, in which I can do nothing to ameliorate her suffering, my demonstration of *mitleid* might be perceived as condescension, casting the sufferer in an inferior role, and thus adding an extra layer of suffering. (*Daybreak* 135)³⁷ Instead of suffering with my friend, I should perhaps respond as Zarathustra prescribes: “be a resting place for his suffering, but a hard bed as it were, a field cot: thus will you profit him best.” (*Thus Spoke Zarathustra* “On the Pitying”)³⁸

³⁶ *Daybreak: Thoughts on the Prejudices of Morality*, trans. R. J. Hollingdale, NY: Cambridge University Press, 1997.

³⁷ *Ibid.*

³⁸ *Thus Spoke Zarathustra: A Book for None and All*, trans. Walter Kaufmann, NY: Penguin Books, 1978.

Compassion doesn't have to be tied so closely to suffering, even if it should be a response to it. A compassionate person doesn't have to *feel* the suffering of others in order to be able to know that it exists and to what extent it exists. Clearly, Schopenhauer didn't hold this to be the case, arguing instead that the only way to know about such suffering is to feel it oneself. But we can see then that *mitleid* serves an epistemic role for Schopenhauer, instrumentally valuable for its giving us knowledge about the suffering of others. A slightly different conception of compassion, one that moves us away from the requirement that one *participate* in the suffering of others, but that includes (a) a reliable mechanism for knowing when, and to what extent, suffering occurs and (b) the appropriate response to such suffering, would be an improvement on Schopenhauer's conception. This brings us back to the notion of empathetic concern. Empathy, understood in a certain way I will discuss next, satisfies (a) while concern satisfies (b). This is markedly different from Batson's empathic concern, since the value of empathy here is not in its connection to altruism, but through its aptness to give us knowledge of the mental and emotional lives of others. This is not to say that we don't often decide to help others as a result of empathizing. I agree with Murdoch when she writes that the "more the separateness and differentness of other people is realized and the fact seen that another man has needs and wishes as demanding as one's own, the harder it becomes to treat a person as a thing."³⁹ Moreover, there is empirical evidence, discussed by Hoffman,⁴⁰ suggesting that when people feel vicarious distress through empathizing with someone having a bad experience, they are more likely to help. What I am arguing here is that what makes empathy morally beneficial lies primarily in its being an epistemic resource, irrespective of whatever kind of behavior results.

³⁹ Iris Murdoch, *The Sovereignty of Good* (NY: Ark Paperbacks, 1970), p. 66.

⁴⁰ *Empathy and Moral Development*, p. 29.

The epistemic value of empathy runs deeper than some accounts would suggest. For example, Nussbaum thinks that its value is found in helping us “recognize humanity in others.” I believe we get more than that when we empathize—or when we empathize well.⁴¹ In order to see why, let’s recall something alluded to earlier. Some accounts posit different kinds of empathy, as we saw with Darwall, leaving room for cognitive empathy, which Hoffman describes as “the cognitive awareness of another person’s internal states, that is, his thoughts, feelings, perceptions, and intentions.” Calling this genuine empathy has an unwelcome implication for some theorists, however, which is that *cold* empathy would then be possible. Most of us have the intuition that, whatever empathy is, it’s an affective state of some sort, which would make the term “cold empathy” oxymoronic. Moreover, since another kind of empathy called affective empathy is characterized by Hoffman as “the vicarious affective response to another person,”⁴² it would seem that genuine empathy needn’t include any sort of perspective sharing, which again runs counter to many people’s intuitions about empathy. In order to guard against the temptation to speak of kinds of empathy, which is counter-intuitive, some theorists have set out to give a componential definition of “empathy.” In conceiving of empathy as constitutively cognitive, found in its perspectival nature, and constitutively affective, usually accounted for in the vicarious experience of the target’s emotional states, we might be said to preserve people’s intuitions.

I don’t have the space to canvas the various componential definitions, so I will state without argument that of the componential definitions on offer, I find the one given by Amy

⁴¹ Some theorists argue that “empathy” is a success-term, concluding that we empathize only when we get things right, whatever one’s definition would say getting things right amounts to. I have argued against this position in my paper “Is Empathy a Success-State?”

⁴² *Empathy and Moral Development*, p. 29.

Coplan⁴³ to be the most plausible.⁴⁴ Coplan argues that there are three essential features of empathy: affective matching, other-oriented perspective-taking, and self-other differentiation. She describes affective matching as occurring only when “an observer’s affective states are qualitatively identical to a target’s” and coming “about in a particular way, namely through other-oriented perspective-taking.” *Other*-oriented perspective-taking is contrasted with *self*-oriented perspective-taking. I’m engaged in the latter when “I imagine what it’s like *for me* to be in *your* situation,”⁴⁵ while someone engaged in the former, “represents the other’s situation from the other person’s point of view and thus attempts to simulate the target’s individual experience as though she were the target individual.”⁴⁶ It’s important for Coplan that the perspective-taking be other-oriented; otherwise, any actions resulting from the self-oriented perspective-taking are likely to be purely egoistic.

What’s important for our purposes here is that, on Coplan’s account, an imaginative projection into the perspective of the target is an essential component of empathy. If one is an effective empathizer, one will be in a position to gather a wealth of information about the target. When one empathizes well, one will be able to access the desires, beliefs, evaluations, dispositions, and other mental states that can help determine the interests of the target. Indeed, it’s hard to see how one could get the same amount of information without possessing this ability. Empathizing is thus a highly valuable tool for facilitating axiological perception.

⁴³ “Will the Real Empathy Please Stand Up? A Case for a Narrow Conceptualization,” *The Southern Journal of Philosophy* 49, Spindel Supplement, 2011, pp. 40-65; “Understanding Empathy: Its Features and Effects,” in *Empathy: Philosophical and Psychological Perspectives*, ed. Amy Coplan & Peter Goldie. NY: Oxford University Press, 2011, pp. 3-19.

⁴⁴ I should say that I find Coplan’s the most plausible from a range of definitions that implicitly adopt what I call the “shared-emotion model of empathy” in my “Closing In on Empathy: A New Model” where I ultimately argue that we should reject this model. My own conception overlaps with Coplan’s *vis-à-vis* the cognitive component, so I have chosen not to include my account of empathy here, as it’s surely to be more controversial than Coplan’s and obscure the central point about perspective adoption necessary for completing the present discussion.

⁴⁵ Coplan, “Understanding Empathy,” p. 9.

⁴⁶ *Ibid*, p.10.

If compassion is best conceived of as empathetic concern, then acting from the motive of compassion will actualize axiological perception, insofar as the latter is the accessing of the interests of those around one. But more than this, concern will actualize the moral perception described by Murdoch and Blum. In order to feel the concern that is coupled to empathy, as described here, the agent will put herself in a position to determine at any given time whether the interests of others should be prioritized over her own. Note here that concern is not the same thing as altruism or *pure* selflessness. The agent is under no moral obligation to act *out* of the interests of any others, as we saw earlier in the discussion of ethical altruism. What the virtuous moral agent will do, however, is give *consideration* to those around her and to what would be good for them. If it turns out that someone else is in greater need than herself, *ceteris paribus* she will do what she can to address that need, as Joan does in Blum's subway example. If not, then she won't—but she has to make that determination first. And this is what happens in step 2 in Blum's schematic.

The motive we were looking for, I contend, is compassion, conceived of as empathetic concern. We haven't mentioned yet the possibility of a particular *virtue* corresponding to or resulting from the possession of this motive, but I suggest that the virtuous moral agent will display, among other virtues, what we can call *considerateness*—i.e., the *moral* consideration of the interests of others and what relationship those interests have to our own. And possession of this virtue will require the agent to be able to engage in both axiological and moral perception.

Murdoch describes *humility* as “the central of all virtues,” but her description differs from how most people think of it: it's “not a peculiar habit of self-effacement, rather like having an inaudible voice,” she says, but a “selfless respect for reality.”⁴⁷ We must resist the temptation to engage in personal fantasy and really see the world around us for what it is. This requires seeing

⁴⁷ Murdoch, *The Sovereignty of Good*, p. 95.

and relating to *other people* with the same kind of “selfless respect.” When an agent is engaged in moral reasoning, she must have the same concern for others as she has for herself, not giving undue weight to her own interests over the interests of others. In exhibiting humility she takes into account the interests of others, accessed through the act of empathizing, locating them in the same space of concern as her own. Similar to Murdoch’s notion of love, Percy Bysshe Shelley writes in *A Defence of Poetry* the following: “The great secret of morals is love; or a going out of our own nature, and an identification of ourselves with the beautiful which exists in thought, action, or person, not our own. A man, to be greatly good, must imagine intensely and comprehensively; he must put himself in the place of another and of many others; the pains and pleasures of his species must become his own.”⁴⁸ It’s clear from each of these statements that Murdoch’s humble person and Shelley’s “greatly good” man are, at the very least, also the considerate person.

4. The Problem Comes into View

In Chapter 1 I argued that axiological and moral perception are necessary for good moral reasoning because these capacities give the moral agent crucial information—i.e., what the interests of others are and how these interests should be weighted respectively. In the present chapter I have argued that the motive of compassion is effective in actualizing both kinds of perception and, further, that an agent who typically acts from compassion displays the virtue of considerateness. A necessary component of compassion, as I have characterized it, is empathy,

⁴⁸ Book Jungle, 2009, p. 54.

which serves as an epistemic resource. A necessary component of empathy, on Coplan's account that I have adopted, is the imaginative projection into the perspective of another. Therefore, one must be fairly accurate at locking onto the perspective of those around one and reliably so. If this isn't the case, the agent's entire process will be based on error—and this will manifest itself most clearly in poor moral decisions at the end of the process. Unfortunately, as we will see in the next chapter, by and large humans are naturally inaccurate at accessing the mental lives of others. And this will allow us to see a problem—perhaps *the* problem—for morality and help explain why the deficit in caring described earlier exists.

Chapter 3

Empathic Accuracy and What It Means to Morality

At the end of the last chapter I alluded to the emergence of a problem for morality. Here is where we stand at this point: at a minimum, a person of good moral character must be able to engage in axiological and weak moral perception, which requires acting from the motive of compassion, understood as empathetic concern, reflected in the agent's possessing the virtue of considerateness. Up to now, most of the arguments I have made have been primarily philosophical in nature, but now that we have made clear what is required of moral agents, we should ask whether normal human beings are able to satisfy these minimal requirements. When we ask this question, it's going to be necessary for us to extend our resources beyond the philosophical, since determining how we are as reasoners is a question for the empirical sciences. In this chapter I will present evidence coming out of a body of research that suggests that we are poor at what is referred to as *mindreading*, which is a technical term defined as the accessing of the mental lives of others, understood to include beliefs, judgments, intentions, and desires. More specifically, we suffer from a relative lack of *empathic accuracy*, meaning that most of our attempts to access the mental lives of others fail to give us accurate descriptions of what these others are thinking and feeling. Inasmuch as being accurate mindreaders is minimally required of moral agents, based on the arguments found in chapters 1 and 2, this should worry us quite a bit. Either we have an obligation to try to improve our mindreading ability, to the extent that this is possible, or we have many fewer moral obligations than we think, since "ought implies can."

Fortunately, improvement is possible, and after presenting the evidence for our poor empathic accuracy, I will discuss the ways that such improvement can be had. In the second part of the dissertation I will lay out my argument for why I think the most effective way to realize these improvements is through living a life in which a particular kind of aesthetic experience is central. Before giving the solution, however, I need to complete the description of the problem.

1. Bad Mindreading

It should come as a surprise to us that we are poor at mindreading. After all, evolutionarily speaking, if our ancestors commonly attributed the wrong mental states to those around them, the species likely wouldn't have survived. But we might conclude from this, not that the purported evidence of bad mindreading must be wrong because of evolution, but that the margin of error was wide enough in those contexts for such inaccurate attributions not to have had such a profound impact. Stephen Jay Gould¹ famously says that evidence for evolution is not found in optimal adaptations, but in jury-rigged solutions to evolutionary problems—what Herbert Simon calls “satisficing.”² Natural selection has only existing systems to operate on, and since the sole “goal” is species-survival, capacities only have to be *good enough*. The same goes for the human mind. Gary Marcus calls the mind a “kluge,” which is a term derived from engineers to refer to “a clumsy or inelegant—yet surprisingly effective—solution to a problem.”³ The mind is a patchwork of “fixes” emerging in the face of various adapt-or-die situations, and concerning the mind's adaptive responses to these situations, Marcus's point is apt: it has been

¹ *The Panda's Thumb: More Reflections in Natural History*, NY: Norton, 1992.

² “Rational Choice and the Structure of the Environment,” *Psychological Review* Vol. 63 No. 2, pp. 129-138.

³ *Kluge: The Haphazard Construction of the Human Mind*, Boston: Houghton Mifflin, 2008, p. 2.

“surprisingly effective” at enabling the human species to survive. But one shouldn’t dwell too long on the success of the species at surviving, especially when we turn our attention to contexts that reach beyond mere survival. What was good enough in the former context might not be good enough in the newer contexts. This applies to the skill essential to interpersonal understanding: mindreading. Even if our ancestors were good enough at mindreading to survive, even if we now are good enough at it to negotiate the various exchanges occurring between us on a daily basis, this says nothing about whether we are good enough at it in other contexts, such as moral reasoning. And the psychological evidence suggests that we aren’t good enough in this context, as we’ll see shortly.

The pioneer of the body of research suggesting we’re bad mindreaders is William Ickes. He and his colleagues devised a way to measure what they call *empathic accuracy*—that is, how often we accurately attribute mental states to those around us. Prior to their experimental design, empathic accuracy was measured by asking subjects how accurate they believed themselves to be at judging the mental and emotional states of other persons. This procedure was shown to be extremely flawed once other methods were implemented, demonstrating that subjects lacked insight into their own cognitive abilities. Actually, subjects turned out to be *worse* judges of their own abilities than others who made purely external judgments about them.⁴ The new experimental setup, called the “unstructured dyadic interaction paradigm,” was designed in the following way. The experiment begins with subjects unknowingly being videotaped in a “waiting room” while the experimenter leaves, ostensibly to retrieve something needed to begin the experiment. The subjects are then videotaped having a six minute conversation. After the conversation, subjects are informed that the experiment has already begun with the taping and

⁴ *Everyday Mind Reading: Understanding What Other People Think and Feel*, Amherst, NY: Prometheus Books, 2003, pp. 173-174.

are asked for permission to continue. Once permission has been granted, the experiment proceeds in two stages. In the first stage the subjects are played the videotape and asked to “tab” points in the conversation when they remember having some thought or emotion and, further, what thought or emotion they remember having. In the second stage, subjects are played the videotape again and, at the points tabbed by their partners—indicating the occurrence of some thought or emotion—the subjects are asked to say whether they had judged their partner to have had some thought or emotion at that point and what the thought or emotion was. After these two stages are complete, the subjects are dismissed. Independent raters are then employed to determine which answers given in the second stage match up appropriately with answers given in the first. The scale used by the raters range from 0 (different content judged by the subject in the second stage in comparison to the partner’s statement given in the first stage) through 1 (similar but not same) to 2 (the same content). In this way empathic accuracy is determined.⁵

The results, which were replicated by others using the same setup, are striking:

Ickes (2003) showed that accuracy in interpersonal judgments is far from perfect. Using the empathic accuracy methodology, which measures the degree to which the thought-feeling events reported by a target person are accurately inferred by a target’s interaction partner (the observer), the data from many studies show that the typical range of mean empathic accuracy scores is from 15% to 35% (W. Ickes, personal communication, February 2, 2005). These numbers suggest that observers have some ability to accurately infer what another person is thinking or feeling, but that they are often relatively unsuccessful. In most social interactions, people are ‘good enough’ perceivers rather than ‘perfectly accurate’ in understanding others’ subjective experiences (Fiske, 1993).⁶

⁵ Ibid, Ch. 3.

⁶Liesbet Goubert, Kenneth D. Craig, Ann Buysse, (2009) “Perceiving Others in Pain: Experimental and Clinical Evidence on the Role of Empathy,” in *The Social Neuroscience of Empathy*, ed. Jean Decety and William Ickes, Cambridge, MA: MIT Press, 2009, p. 155.

At best, subjects accurately attributed mental and emotional states barely one out of three attempts, and at worst, fewer than two out of ten attempts. “In most social interactions,” it’s said, “people are ‘good enough.’” But given the arguments made in chapters 1 and 2, a moral reasoner needs to be accurate more reliably than this. What this suggests is that the average moral agent has a flawed conception of what the interests of others are, and it’s even possible that the gap between the actual and the perceived interests is wide. Is it any wonder then that moral agents care much less than they should? They seem to lack the ability to know when caring is called for. A problem for morality indeed!

2. What Can Make It Better?

So there is evidence that we are relatively poor at accessing the mental lives of others, which should make us very uneasy. We saw in earlier chapters that being able to determine the interests of others is vital to moral decision-making and that determining the interests of others centrally involves being able to mindread. Now it looks like we are constitutively unable to be good moral reasoners. There may be hope, however, since as we see what explains the poor mindreading skills, we will further see that they can be improved. Unfortunately, as we go through the various capacities and skills underlying mindreading, we may get the impression that the situation is much worse than even the research of Ickes shows us. But we should take heart. In later chapters I will show that we can become better mindreaders and thus better prepared to reason morally.

Why are we bad mindreaders? There is a vast literature on this, but broadly speaking four factors play most directly into our enjoying such low empathic accuracy: (1) our natural egocentrism, which makes it very difficult for us to de-center our own perspective and re-center on the perspectives of others; (2) an underdeveloped capacity to gather metaknowledge about our relative successes and failures at mindreading; (3) relatively little motivation to mindread well; and (4) flaws in the underlying cognitive mechanisms responsible for the ability to mindread. Let's look at each of these.

First, we are naturally egocentric, which means that it's very difficult for us to see a situation from any perspective but our own. This becomes manifest in several different ways, including the following:

- *Egocentrism in language speakers* – It's been shown that when formulating utterances people naturally adopt an egocentric perspective. Boaz Keysar and Dale J. Barr⁷ suggest that this is contrary to what we might expect, since the goal of communicating content is for the listener to understand us. We might think that, in order to accomplish this goal, language speakers would “design utterances bearing in mind the other's thoughts, beliefs, assumptions, and so on.” This is what we would expect if language use conformed to the “principle of optimal design.” Speakers would anchor onto the cognitive states of their conversational partners and try to tailor their utterances to these states. Speakers don't conform to this principle, however. What happens is they “anchor on their egocentric perspective and then attempt to adjust to their partner's perspective” as the need arises.

The need usually arises only in cases of ambiguity, which still represents a considerable

⁷ “Self-Anchoring in Conversation: Why Language Users Do Not Do What They Should,” in *Heuristics and Biases*, ed. Thomas Gilovich, Dale Griffin, & Daniel Kahneman, Cambridge, UK: Cambridge University Press, 2002.

number of instances. Speakers adopt what Keysar and Barr call the “perspective adjustment theory of language use,”⁸ where they anchor on their *own* cognitive states and then make adjustments when necessary. Thus language speakers are naturally egocentric.

- *Egocentrism in language listeners* – Listeners of spoken language also naturally adopt an egocentric perspective. This has been shown in experiments in which subjects hear a speaker refer to an object’s relative location, like “the peg in the bottom row,” in instances when the speaker cannot possibly know that what she is referring to as the bottom row is actually the middle row, since her vision is occluded and she can’t see the additional row. The listener, however, does see the additional row and knows that the speaker doesn’t. Thus the speaker is actually referring to what the subject knows is the middle row, when she says the words “the bottom row.” The subject should automatically infer this, given his knowledge that the speaker is unaware of the extra row. Response times show, however, that listeners automatically attend to the actual bottom row, because from their perspective, “bottom row” refers not to the middle row, but to the actual bottom row.⁹ This phenomenon is also demonstrated in contexts of pronoun use. A subject is prompted to think of a particular person such that his partner would have no way to know that he is thinking of this person. Further, he knows that his partner is unable to know this. In conversation the partner uses a pronoun matching the gender of the person about whom the subject is thinking. Even though there is no way that the partner could be using the pronoun to refer to the person being thought about, the subject

⁸ Ibid, p. 152.

⁹ See Dale J Barr & Boaz Keysar, “Mindreading in an Exotic Case: The Normal Adult Human,” in *Other Minds*, ed. Bertram Malle & Sara D. Hodges, NY: The Guilford Press, 2005.

will still automatically interpret it as referring to this person, rather than someone about whom both the subject and the partner were conversing.¹⁰

- *Egocentric perspective in pain attribution* – Research in pain attribution suggests that individuals process judgments about the pain of others primarily in the same regions of the brain that are activated when the individuals experience pain themselves. Philip Jackson et al have shown “that the anterior cingulate and anterior insula cortices, regions often reported as being part of the pain affective system, are recruited when watching someone else’s pain.”¹¹ It’s as though we have to locate the pain in ourselves before we can attribute the pain to someone else. We feel their pain, perhaps, in some sense, but not immediately as *their* pain. First we have to experience it as our own.

Evidence found in each of these categories suggests that the natural perspective of individuals is egocentric. We shouldn’t conclude from this that people always relate to the world from their own perspective and are unable to do otherwise, but the research does show that much of the automatic processing is egocentric, and that a subject must make adjustments in order to co-opt what is naturally the default.

This co-opting requires what is called “decentering,” which involves putting aside one’s “own perspectives, worldviews, values, and preferences as” one tries to fully enter the world of the other person and experience the world as she does.¹² Decentering occurs as a result of a shift in perspective. Since the natural egocentric perspective is primarily pre-conscious, decentering occurs primarily as a conscious process, as when a subject intends to make egocentrism less

¹⁰ See Keysar & Barr, “Self-Anchoring in Conversation.”

¹¹ Philip L. Jackson, Andrew N. Meltzoff, Jean Decety, “How Do We Perceive the Pain of Others? A Window into the Neural Processes Involved in Empathy,” *NeuroImage* 24, 2005, p. 777.

¹² Jeanne C. Watson and Leslie S. Greenberg, “Empathic Resonance: A Neuroscience Perspective,” in *The Social Neuroscience of Empathy*, ed. Jean Decety and William Ickes, Cambridge, MA: MIT Press, 2009, p. 131.

automatic. This suggests that de- and re-centering is within our ability, so long as we know that it's something that must be consciously pursued.

This brings us to the second area in which improvement should occur in order to effect better empathic accuracy: metaknowledge. When it comes to empathic accuracy, people are rarely aware that they have attributed the wrong mental states to those around them. As suggested above, usually the baseline empathic accuracy is good enough in social interactions, since adjustments can be made quickly and misattributions can be relatively benign. Ickes¹³ thinks that this impoverished metaknowledge—i.e., knowledge about how many of our beliefs about the mental states of others are true—is probably due to a lack of appropriate feedback in ordinary situations along with the natural egocentric perspective just described. Thus improving one's metaknowledge requires the ability to decenter, as well as opportunities for the right kind of feedback related to individual instances of mindreading.

The ability to decenter and the possession of metaknowledge are necessary, but not sufficient, conditions for improving empathic accuracy. One must be motivated to mindread well. Some research shows that increased motivation alone might improve empathic accuracy. For example, Ickes and colleagues set out to test whether it was true that women were better mindreaders than men. It's often said that women—through “women's intuition”—are better at mindreading than men. Several early experiments, intended to test this belief and designed according to the unstructured dyad interaction paradigm, suggested that women were *not* better at mindreading than men; statistically there was no gender differential at all. But at a later date, they ran a different set of experiments with the same purpose and they suddenly found a marked difference between women and men. The method was the same—conforming to the same

¹³ “Empathic Accuracy,” *Journal of Personality* 61, 1993, pp. 587-610.

paradigm—so some other factor must have changed between the earlier and later experiments. After investigating, the researchers found a difference in the forms that subjects were asked to complete, indicating the thoughts and feelings they thought their partners possessed during the conversation (stage 2). Earlier forms displayed a column for target state valence (Do you think the target's state is a. positive; b. neutral; c. negative?) while the later forms displayed a column asking the subject how accurate he or she thought her current inference to be. It was in experiments using this second form that the marked difference resulted. The experimenters determined that the reason women were making more accurate inferences in the later experiments was that they had been reminded by the new column on the form of the societal expectation that as women they are *supposed to be* better mindreaders than men.¹⁴ Once they consciously took up that presupposition, they were more motivated to mindread well and as a result, they became better at it. This suggests that empathic accuracy relies, in part, on motivation: “Although women, on average, do not appear to have more empathic *ability* than men,” Ickes writes, “there is compelling evidence that women will display greater accuracy than men when their empathic *motivation* is engaged by situational cues that remind them that they, as women, are expected to excel at empathy-related tasks.”¹⁵

Kristi Klein and Sara Hodges¹⁶ tried to determine whether motivation *in general* could be a factor in improving empathic accuracy. Even though it was societal expectations and perhaps the centrality of strong empathic skills in the self-concept held by many of the female subjects that were responsible for the stronger motivation in the experiments run by Ickes et al., it was

¹⁴ See Tiffany Graham & William Ickes, “Why Women’s Intuition Isn’t Greater Than Men’s,” in *Empathic Accuracy*, ed. William Ickes, NY: The Guilford Press, 1997; and William Ickes, Paul R. Gesn, & Tiffany Graham, “Gender Differences in Empathic Accuracy: Differential Motivation or Differential Ability?” *Personal Relationships* 7, 2000.

¹⁵ *Everyday Mind Reading*, p. 135.

¹⁶ Kristi J. K. Klein & Sara D. Hodges, “Gender Differences, Motivation, and Empathic Accuracy: When It Pays to Understand,” *Personality and Social Psychology Bulletin* 27, 2001.

hypothesized that motivation was the crucial factor for the improvement in accuracy. Thus if experimenters could increase the motivation of both male and female subjects to be empathically accurate, then the gender difference should disappear and there should be a marked difference between control subjects and those subjects whose motivations were increased. Klein and Hodges attempted to increase motivation using monetary rewards for accurate empathic inferences. Their findings supported their hypothesis: greater motivation in general improves empathic accuracy. They conclude: “In sum, motivation seems to be a key component in the process of empathizing with another person. The gender difference that was found in our studies seems to be adequately explained by differential motivation between men and women under different circumstances. We believe that this is an encouraging finding, suggesting that greater empathic accuracy can be achieved by virtually anyone who is given the proper motivation.”¹⁷

As motivation plays such a crucial role, we might be tempted to conclude that ability is not that important in this context. After all, if it’s just a matter of trying harder, then much of what is necessary for improvement would already be in our possession. Unfortunately, researchers believe that the mechanisms underlying our ability to mindread need to be improved as well. Some of these improvements might come along with improvements in the first three areas, but we have to know what these mechanisms are in order to know what would improve them. This presents us a difficulty, however, since there isn’t widespread agreement about what the mechanisms are. In the next section we’ll look at the debate about this issue.

¹⁷ Ibid, pp. 728-729.

3. Modes and Mechanisms of Mindreading

The debate over what mechanisms underlie mindreading is in part empirical and in part philosophical. Philosophers have entered the fray because the evidence that empirical science can give us—the nature of the evidence as well as the amount—is insufficient to explain how we mindread. One way that has seemed especially promising for constructing an account of how normal subjects mindread is to consider the abnormal cases of mindreading, the most prominent example of which is autism. Several influential collections of papers¹⁸ dealing with autism and mindreading have appeared, though what emerges from these exchanges is an apparent lack of consensus about what distinguishes autistic from normal subjects. There does seem to be agreement, however, about one coarse-grained deficit underlying autism: *mindblindness*. Simon Baron-Cohen¹⁹ uses this term to refer to what is perhaps the most telling deficit in autistic subjects. Just as there is a deficit in people who are blind to things in the physical world, mindblind people are blind to the very existence of mental states in others. Another way of putting this is that autistic people are unable to “metarepresent;”²⁰ that is, they cannot represent other persons representing anything in the world through beliefs, desires, judgments, and intentions. Extremely autistic individuals are unable to access, or recognize as real, the mental states of others. Thus in these individuals, we have a picture of an *inability* to mindread. What is missing in autistic subjects might be what is responsible for mindreading in non-autistic subjects.

¹⁸ See Simon Baron-Cohen, Helen Tager-Flusberg, & Donald J. Cohen, ed., *Understanding Other Minds: Perspectives on Autism*, NY: Oxford University Press, 1993; *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. NY: Oxford University Press, 2000; Janet Astington, Paul L. Harris, David R. Olson, ed., *Developing Theories of Mind*, NY: Cambridge University Press, 1988; Peter Carruthers & Peter K. Smith, ed., *Theories of Theories of Mind*, NY: Cambridge University Press, 1996.

¹⁹ *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press, 1995.

²⁰ This term appears in Alan Leslie, “Pretence and Representation,” *Psychological Review* 94, 1987.

Today researchers consider autism a spectrum disorder, with extreme cases on one end marked by a complete lack of ability to represent others as having mental states, and less extreme cases on the other end, usually referred to as Asperger's syndrome, marked by only partial inabilities. There isn't a consensus yet about whether the ability to mindread itself is located on a spectrum, with the worst ability represented by the cases of extreme autism, and progressively better abilities the further up the continuum we go. Many people do hold that there is a threshold of ability, beneath which persons are considered autistic, or as having Asperger's, and beyond which, persons are considered "normal". On this *absolute threshold model*, there would be a range in which subnormal abilities to mindread cause drastic disabilities in subjects, but there would also be a range beyond a certain threshold point within which people exhibit no deficits. Taking the analogy of visual blindness further, according to this model, autistic individuals might be described as mindblind to varying degrees, all the way up to the highly functioning subjects with Asperger's. Once beyond that point, there would be mind-sighted individuals, who, in spite of the fact that there is a variation of ability from the low end to the high end, experience no lack of functionality, with perhaps additional assistance from other mental faculties. Just as myopic individuals are able to function with the aid of eyeglasses or contact lenses, those who have poorer mind-sight than others would still be able to make their way in the world relatively well.

But some researchers have suggested that that normal range itself can be quite wide, suggesting that instead of the absolute threshold model, we should adopt an *absolute continuum model* where the ability to mindread can be located anywhere on the continuum, with extremely autistic individuals on the near-zero ability end, and other, perhaps only hypothetical, cases on

the perfect ability end.²¹ There would still be a threshold differentiating autism and Asperger's cases from non-autistic cases, while there could still exist a differential between those just beyond the threshold and those further up the continuum, a differential that in theory could be just as wide as that between extremely autistic individuals and those with Asperger's. This model fits nicely with the findings that we discussed in the section on empathic accuracy, since the mean accuracy scores Ickes reports ranged from 15 to 35%, with some subjects mindreading accurately slightly more than one times out of ten and other subjects successful around one times out of three—and these are normal subjects! The range found within the work done on empathic accuracy suggests that we do have a wide range of abilities, supporting the absolute continuum model. On adopting this model, we could entertain the possibility that people have the potential to improve their empathic accuracy by improving their ability to mindread, and hope to see success rates of 50, 60, or even 70%.

In the next section I am going to give a sketch of how the debate has proceeded about the mechanisms of mindreading. In order to make our job easier later, I will break mindreading down into three *modes*: (1) heuristic mindreading; (2) deliberative mindreading; and (3) mindwriting. Most discussions have conflated (1) and (2), often not even recognizing that mindreading might occur in different ways depending on context, while very few theorists have recognized the importance of mindwriting at all. Briefly, heuristic mindreading is the process of gaining epistemic access to other minds in everyday contexts—a process that must by its very nature be fast, automatic, and likely unconscious, while deliberate mindreading is the process of gaining access to other minds in contexts requiring extended reflection, and often involving higher stakes than in everyday contexts. Mindwriting is the process of making one's mental

²¹ This raises an interesting question: would those who have perfect ability to mindread be as negatively affected as autistic individuals who have zero ability to mindread? Would perfect mindreading represent an enviable way of life?

states more easily accessible to other mindreaders. Although very few discussions have broken down the larger issue in this way, I contend that taking account of each of these will throw the overall question of improvement of mindreading into greater relief.

4. Theory vs. Simulation

The debate over the mechanisms of mindreading quickly divided into two opposing camps: the theory-theorists and the simulation-theorists. The theory-theorists—including Shaun Nichols, Stephen Stich, Jerry Fodor, Simon Baron-Cohen, Alan Leslie, Alison Gopnik, Henry Wellman—hold that human beings understand the actions of others principally as a matter of inference from observation and some body of (probably) implicit information to a belief about the mental states of the other persons. It's argued²² that humans move through the world possessing and utilizing a folk psychology—that is, a body of unifying principles very similar to the sets of principles we find in paradigmatic scientific theories. Theory-theory is the view, then, that whenever we are confronted with a situation in which we desire to understand, explain, or predict the action of others, we draw on the relevant principles, based on the situation we face, and then make an inference. The process described by theory-theory is very similar to scientific reasoning, which is why theory-theorists hold that the primary factor responsible for our mindreading is, in fact, actually *a theory*.

None of this need be at the conscious level, as we've seen. Theoretical inferences are “heuristically valuable,” which means that they get us what we need to know, even if they aren't

²² See Paul Churchland, “Eliminative Materialism and the Propositional Attitudes,” *Journal of Philosophy* 78, 1981; David Lewis, “Psychophysical and Theoretical Identifications,” *Australasian Journal of Philosophy* 50, 1972.

the optimal solutions to the problem of interpersonal understanding. Fodor makes this point in the following:

Commonsense psychology works so well it disappears. It's like those mythical Rolls-Royce cars whose engines are sealed when they leave the factory; only it's better because it isn't mythical. Someone I don't know phones me at my office in New York from – as it might be – Arizona. “Would you like to lecture here next Tuesday?” are the words he utters. “Yes, thank you. I'll be at your airport on the 3 p.m. flight” are the words that I reply. That's *all* that happens, but it's more than enough; the rest of the burden of predicting behavior—of bridging the gap between utterances and actions—is routinely taken up by theory. And the theory works so well that several days later (or weeks later, or months later, or years later; you can vary the example to taste) and several thousand miles away, there I am at the airport, and there he is to meet me. Or if I *don't* turn up, it's less likely that the theory has failed than that something went wrong with the airline. It's not possible to say, in quantitative terms, just how successfully commonsense psychology allows us to coordinate our behaviors. But I have the impression that we manage pretty well with one another; often rather better than we cope with less complex machines.²³

Here Fodor is arguing against people like Churchland²⁴ and Stich²⁵ who think that the theoretical terms of folk psychology are not appropriate for inclusion in cognitive science, but all three agree, however, that this is the method we use in our navigation through the world in our everyday life. They are all theory-theorists under the current conceptual framework.²⁶

²³ Jerry Fodor, “The Persistence of the Attitudes,” in *Philosophy of Psychology: Contemporary Readings*, ed. Jose Bermudez, NY: Routledge, 2006, p. 219.

²⁴ “Eliminative Materialism.”

²⁵ *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA: MIT Press, 1983.

²⁶ I ignore some issues that are internal to the theory-theory camp. For example, there is disagreement about whether the represented folk psychological principles are modular, and, if so, whether the contents of the theory-module are hard-wired (see Jerry Fodor, *Modularity of Mind*. Cambridge, MA: MIT Press, 1983 for a description of modularity; and Gabriel Segal, “The Modularity of Theory of Mind,” in *Theories of Theories of Mind*, ed. Peter Carruthers and Peter Smith, NY: Cambridge University Press, 1996 for an argument that mindreading is modular). Also, there is a debate in the developmental literature about whether children develop a body of theory early on, in which case, at any given time, the origin of the set of principles can be found in early childhood. Some theorists

An alternate kind of process has been proposed to explain our interpersonal understanding. The simulation-theorists—including Alvin Goldman, Robert Gordon, Jane Heal, Greg Currie, Paul Harris—argue that understanding the actions of others is not a matter of inference at all, but occurs as a result of a kind of imaginative sharing of perspective. They call this sharing *mental simulation*. The central idea of simulation-theory is that whenever we are faced with a situation in which we are trying to understand, explain, or predict the actions of another person, we place ourselves in his or her situation or perspective and simulate the mental processes likely to occur, resulting in an output state, consisting of a behavioral response and/or resultant mental states. The person simulating (called the source) feeds into the simulator (the source's own cognitive and affective system) the initial conditions, including details about the external situation and occurrent mental states of the simulated person (called the target). The simulation is then run, beginning from the initial conditions, via an imagining of what kinds of mental states would likely result from those conditions. The entire process is run “off-line” which prevents the source from realizing the output behavioral responses and taking up the resultant mental states into the source's own cognitive framework.²⁷ The mental states and behavioral states remain pretend states. After the simulation has ended, the source sees what output states the target is likely to be in at the time in question.²⁸ There are two competing versions of simulation that find appropriate places in our discussion. The first, what I call here

(Gopnik, Wellman) further argue that children are born with a very simple theory. This theory gets overturned, however, when the child learns that it's not working. Throughout the developmental stages, new theories come along to replace the old ones, when the old ones are found to be ineffective (similar to a Kuhnian crisis/replacement sequence—the theory is often referred to as “child as scientist theory”). I don't get into these details, since the general features relevant to our purposes are the same in each of these amended versions and, also, because these versions seem to be outliers, with fewer adherents than the central model.

²⁷ There can be degrees of “off-lineness” evident in the fact that some *affective* states can result from the simulation, depending on the source's intention in running the simulation (on this point, see Ian Ravenscroft, “What Is It Like to Be Someone Else? Simulation and Empathy,” *Ratio (new series)* 2, 1998). It's hard to see how non-affective states could be allowed to run on-line, however, since one would have to forget that one was running a mere simulation in order to take up resultant beliefs, desires, hopes, etc., after the simulation.

²⁸ This phrase “time in question” is intentionally vague in order to account for the fact that simulations are supposed to be used for both explanation and prediction, which would involve present and future output states respectively.

transference simulation, is adopted by Goldman and others, while the second, what I call transformation simulation, is defended primarily by Gordon. Once we see the difference between these two kinds of simulation, we will be in a position to understand why transference simulation, combined with some theoretical content, explains heuristic mindreading, while transformation simulation, also linked up to theoretical content, explains deliberative mindreading. The account I will advocate will be what is called a *hybrid* account, since it posits a combination of both theoretical content and simulation as the mechanisms underlying both heuristic and deliberative mindreading.

5. Transference vs. Transformation Simulation

The main difference between transference and transformation simulation lies in the attribution process. Transference simulation takes the overall simulative process to be analogical in nature, with attribution coming after the simulation is over—and therefore *transferred* to the target—while transformation simulation takes it to be imaginative—that is, one re-centers perspectives prior to the simulation such that at the end of the process, the output states would properly belong to the perspective of the target, and not the source, thus making transference unnecessary. The source would have *transformed*, in some suitably metaphorical sense, into the target prior to running the simulation.

In order to evaluate each kind of simulation, we would have to keep in view two points in the simulation process that things are most likely to go wrong: (a) in the input of the initial conditions and (b) in the simulating itself. Consider the transference simulation process:

- (1) the input of initial conditions, including relevant information about the target (mental states, life history, idiosyncrasies, etc.)
- (2) the simulation
- (3) the output states
- (4) an analogical inference from the source to the target

One of the worries Goldman rightly raises about simulation is that the process might be contaminated by the source's own idiosyncrasies. He writes that a successful simulation will have to have a way "for the attributor [source] to *quarantine* his own idiosyncratic desires and beliefs (etc.) from the simulation routine. If the attributor has desires or beliefs that aren't shared by the target, allowing them to seep into the routine could contaminate it."²⁹ This would mean that within the simulation process, anything particular to the target would be accounted for in (1), but in (2), the simulation should proceed in the same manner for anyone. In other words, once the initial conditions have been fed into the mechanism, the simulation would run in roughly the same way for any target. No idiosyncratic facts about the target would bear at all on the simulation process itself. In the end, the output states would be transferred to the target, based on the judgment that anyone who began in the initial conditions would be likely to reason in the same way as becomes manifest in the simulation.

²⁹ *Simulating Minds*, NY: Oxford University Press, 2006, p. 29.

Goldman thinks this worry about “contamination” is a reason to accept transference simulation as an explanation of mindreading. One would just have to be a reliable simulator in roughly the same manner as any other source acting as a simulator. Gordon³⁰ worries, however, that the analogical inference found in Goldman’s (4) weakens the general case for simulation as a rival to theory-theories. One of the purported strengths of simulation-theory is that the process it depends on is relatively information-poor.³¹ It’s argued that theory-theory cannot account for children’s mindreading abilities, since children wouldn’t have built up enough folk psychological principles or amassed enough concepts in order to complete the analogical inference. One way that theory-theorists respond to this charge is to adopt the modular or the nativist versions of theory-theory (see fn. 26), which allows for hardwired implicit representations of the folk principles. These versions have problems of their own. Simulation-theory, however, isn’t supposed to rely on any conceptual knowledge. One just needs to know what initial conditions to feed into the mechanism and then run the mechanism. If there is need for transference, however, simulation-theory needs to be able to account for matching up source and target mental content, in order to justify the analogical inference. At best, says Gordon, this cancels out the purported strength of simulation-theory, and, at worst, it pushes simulation-theory toward collapse into theory-theory.

The problem, according to Gordon, is that simulation-theorists have assumed that simulation “is, or is part of, an introspection-based methodology.”³² This methodology would exhibit the following sequential order: the source surveys the situation of the target; asks herself

³⁰ “Simulation without Introspection or Inference from Me to You,” In *Mental Simulation: Evaluations and Applications*, ed. Martin Davies and Tony Stone, Cambridge, MA: Blackwell Publishing, 1995; “Ascent Routines for Propositional Attitudes,” *Synthese* 159, 2007.

³¹ See Gregory Currie, “Simulation-Theory, Theory-Theory, and the Evidence from Autism,” in *Theories of Theories of Mind*, ed. Peter Carruthers and Peter Smith. NY: Cambridge University Press, 1996.

³² “Simulation without Introspection or Inference from Me to You,” p. 63.

what mental states she would occupy, if she were in that situation; runs the simulation, proceeding from having pretended to be in those initial mental states; produces the new (off-line) mental and behavioral states; comes out of the simulation; and infers that the output states in which she herself ended up are likely to be the ones which the target is in. The whole process centers on the source's own mental states, which would take as its object certain conceptual content—content that is first and foremost part of her own mental architecture. In order to make the analogical inference, one would have to justify the belief that the target possesses the same conceptual content, which is where simulation loses its appeal.

The problem disappears, Gordon argues, if we reject the assumption that the methodology is introspection-based. Rather than running a simulation based on introspective access to what mental states the source would be in, given the situation of the target—a process that afterward requires the inference from the source's output states to a belief about the target's states—Gordon argues that the connection between the simulation and the target is made prior to the simulation. The source performs an *egocentric shift* before determining the initial conditions. In trying to come up with the initial conditions, the source asks, “What do I think or feel in this situation?” where “I” doesn't refer to the source, but to the target, whose persona the source has now adopted. The source recenters, such that she has now imaginatively projected completely into the target's perspective. The source thus *transforms* into the target, in a metaphorical sense, and then runs the simulation under that perspective.

Gordon likens this process to acting. “Such recentering,” he says, “is the prelude to transforming myself in imagination into [the target] much as actors become the characters they play. Although some actors (‘method’ actors, for example) occasionally step back from the role they are playing and ask, ‘What would I *myself* do, think, and feel in this situation?’ and then

transfer their answer (with or without adjustments) to the character, the typical stance of modern actors is that of being, not actors pretending to be characters in a play, but the characters themselves.”³³ The source antecedently transforms herself into the target, via imagination, and runs the simulation under the perspective of the target, rather than running the simulation under her own perspective, after having imagined herself in the situation of the target, and then transferring the output states from herself to the target. There is no longer any need for analogical inference.

One objection to Gordon’s view might be that we have merely moved the point of inference to an earlier stage—i.e., at the beginning when the source determines that the mental states she imagines herself being in under the perspective of the target are the same kind of states the target would be in. Gordon disagrees, however, saying that when the source performs the egocentric shift, she imagines that *she* is in the target’s situation. When she determines which mental states to feed into the mechanism, she doesn’t look inward and see what states she occupies—she looks outward, at the situation she has imagined herself to be in, and determines which mental states would follow from the way the world is. She performs what Gordon calls an ascent routine. Whenever we try to determine what attitude we hold toward any proposition, Gordon contends, we don’t focus on the attitude purportedly held toward the proposition. We focus on the proposition itself. For instance, suppose someone asks me whether I believe that the University of Washington is in Spokane. I wouldn’t, Gordon claims, introspect and look at my set of beliefs, to determine whether the proposition “The University of Washington is in Spokane” is there. I would, instead, ask the question, “Is the University of Washington in Spokane?” I “step down a semantic level,” says Gordon, and try to “answer the corresponding

³³ Ibid, p. 55.

question about the matter the belief is about.”³⁴ If I answer yes to the question, then I ascend a semantic level and make the statement “I believe that the University of Washington is in Spokane” and if I answer no, then I ascend and make the statement “I don’t believe that the University of Washington is in Spokane.” Similarly, when the source is preparing to perform a simulation, she looks at the world outside her—the world surrounding the target represented in imagination—and performs the ascent routine. In this way she produces the mental states³⁵ that she (as the target) would occupy, given the world represented in imagination. She does not introspect at any point in the process.³⁶

There are other theories of simulation that overlap to some extent with Goldman’s and Gordon’s accounts, but I take the major disagreements to be about whether simulation is run from within a particular perspective or a general conception of the rational person. As I said before, the debate has been about heuristic mindreading—both Goldman and Gordon intend their accounts to explain mindreading in everyday interactions with others. I want to suggest here, however, that *if* transference and transformation simulation have roles to play in mindreading, the role for transference simulation is properly located in the heuristic context, while transformation simulation would properly belong in the deliberative context. I won’t take the time here to defend this claim; instead I will make the assumption that in heuristic contexts, transference simulation would be more appropriate, while in deliberative contexts, transformation simulation would be more appropriate. Again because of space constraints, I will assume without argument that both heuristic and deliberative mindreading require more than

³⁴ “Ascent Routines for Propositional Attitudes,” p. 153.

³⁵ Gordon (“Ascent Routines for Propositional Attitudes,”) has an account of how ascent routines can be performed for other mental states, like desire, hope, fear, regret, wish, etc. I refer the reader to this account, rather than present it here.

³⁶ One might object that introspection still happens, since the source performs the ascent routine, formulates the thoughts about the mental states, then refers to the set of mental states to determine what to feed into the mechanism. This is one step too many, however, since the source would be in the (pretend) mental states already, having performed the ascent routine. This is all she would need to run the simulation.

mere simulation—they also require some theoretical content. Consequently, heuristic mindreading involves both theoretical content and transference simulation, while deliberative mindreading involves both theoretical content and transformation simulation. Therefore, improvements in each of these areas will represent some improvement on the ability to mindread, and thus on our empathic accuracy.³⁷

6. Mindwriting

Finally there is the mode of mindreading that I call *mindwriting*. Strictly speaking, when one mindwrites—that is, when one attempts to make his mental states accessible to others—he isn't mindreading at all. Nevertheless, successful mindreading overall requires successful mindwriting, which we might think of as *writing the text* of the mind. Sources will be more accurate at reading the minds of targets when the targets are better able to write the texts of their minds.³⁸

³⁷ There are a couple of issues I don't address in the text. First, since I am breaking down mindreading into heuristic and deliberative mindreading, it might seem that my arguments later will rely crucially on whether these modes of mindreading are genuine modes. This would necessitate my giving a fuller defense of my claim about what constitutes heuristic and deliberative mindreading. But my arguments later will show that engaging in a specific way with specific kinds of artworks improves on each mechanism: theoretical inference, transference simulation, and transformation simulation. If one or more of these mechanisms aren't involved in mindreading, the improvements I describe can still apply to the others. The only claim that would have to be true is that at least *one* of these mechanisms underlies our capacity to mindread. I will defer to the most recent research on mindreading as support for this claim. Second, while I don't note it in the text, by claiming that transference simulation is proper to the heuristic context and transformation simulation is proper to the deliberative context, I have shifted the debate from a purely descriptive account to a normative account, which is not reflective of the way the debate has occurred in the literature. I believe that, regardless of whether it is the case, transference simulation combined with theoretical inference *should* predominate in heuristic contexts and transformation simulation combined with theoretical inference *should* predominate in deliberative contexts.

³⁸ At least for one kind of disorder—borderline personality disorder (BPD)—the inability to mindwrite has been judged to be a significant factor. BPD is a disorder marked by “a pattern of intense, unstable personal relationships, identity disturbance (a marked and pervasively fragile sense of self), negative affect, impulsivity, [...] an elevated sensitivity to environmental circumstances [...] a continual and intense fear of rejection or abandonment and

Adam Morton³⁹ has written on this issue of the importance of being understood. He argues that it is “of great benefit to know what others will do, and that we often do have this knowledge,” but that the power of folk psychology does not come “primarily from the ability to make bare predictions of future behavior.” What underlies this ability, he says, is the creation of conditions in which persons’ actions are predictable. The creation of these conditions is as much reliant on actions of the target as on the source. As the target, you have the responsibility of “giving the person an understanding of your motives and course of action.” While mindreading is nature’s answer to the problem of the opacity of the mind of the other, this opacity can be ameliorated to some extent by the target’s making his own mental states more accessible. What is important, says Morton, is “mutual intelligibility.” The process is bidirectional. The source tries to make sense of the actions of the target by gaining access to the target’s mind, but the target too tries to make his actions intelligible by giving cues facilitating such access. Morton uses as an example the phenomenon of driving in a place where you haven’t driven before:

[Y]ou are confronted with an environment of lunatic car drivers. They seem completely unpredictable.

Perhaps some of their choices are made completely at random; others are driven by motives that you cannot fathom. (Even if you could fathom them they might not tell you what the people were going to do.) But you do know a few things. You know that they do not want to collide with you, if they can help it. [...] And you know that they will take precautions to avoid a collision. So what you can do is this. You head out slowly

[avoidance of] real or imagined abandonment by others” (Judith M. Flury, William Ickes, William Schweinle, “The Borderline Empathy Effect: Do High BPD Individuals Have Greater Empathic Ability? Or Are They Just More Difficult to ‘read’?” *Journal of Research in Personality* 42, 2008, pp. 312-313). Researchers had originally suspected that the primary cause of this disorder was a heightened sense of empathy, but it’s since been suggested that the primary reason for the family of symptoms is that subjects’ “own reported thoughts and feelings are atypical and quite difficult to ‘read’ in comparison to those reported by their non-BPD interaction partners” (William Ickes, “Empathic Accuracy: Its Links to Clinical, Cognitive, Developmental, Social, and Physiological Psychology,” in *The Social Neuroscience of Empathy*, ed. Jean Decety & William Ickes, Cambridge, MA: MIT Press, 2009, p. 61). Thus the inability to mindwrite can be quite devastating, both to those who cannot mindwrite and those who cannot mindread as a result of the failure of mindwriting.

³⁹ *The Importance of Being Understood: Folk Psychology as Ethics*, NY: Routledge, 2003.

into the traffic, making your intentions as clear as possible and acting in a collision-minimizing way. Other drivers will then be able to avoid you, and they will, as long as they can anticipate your actions. What your survival requires is not that you predict their actions but that they be able to predict yours.⁴⁰

This is, of course, an exaggerated situation, but it shows what role target behavior has in the overall process of mindreading. In the extreme case, the source has no inherent ability to predict the actions of the target, but can, through the aid of carefully provided cues by the target, adjust her behavior to coordinate with the behavior of the target. In less extreme cases, the source has mindreading ability, the target has mindwriting ability, and the two abilities mesh to form one fairly successful enterprise. But only fairly successful, given the relatively weak requirements for species survival. As we've seen, a very low level of empathic accuracy is required for species-survival, while a higher level is necessary for moral reasoning.

We can now see that an improvement in mindreading and empathic accuracy requires an improvement of at least the following four capacities:

1. decentering
2. metaknowledge regarding our empathic accuracy
3. motivation to mindread
4. the ability to mindread (heuristic mindreading, deliberative mindreading, and mindwriting)

Inasmuch as being a virtuous moral agent requires that we be relatively accurate mindreaders, solving the moral problem described in this part of the dissertation will minimally require that we improve in these areas. In part II we will consider some of the issues surrounding the

⁴⁰ Ibid, p. 27.

possibility that art provides us a solution. We will have to wait until part III, however, to see what I believe is an adequate solution to the problem.

Part II

Art as the Solution?

Chapter 4

Art and Morality: A Framework

We ended chapter 3 with the description of a significant problem for morality as well as a specific way that this problem might be solved. I will argue that art represents such a solution, but this claim comes with provisos. In order to see why these provisos are necessary, we must explore some issues surrounding the question of whether there is any legitimate relationship between art and morality. After this exploration, we will need to look at several of the most prominent accounts claiming that there *is* a legitimate relationship in order to see whether any of them presents a solution to our problem. I will argue that none of them is adequate for our purposes and each for roughly the same reason: they are all primarily particularistic in nature—i.e., they focus on individual artworks and experiences, which leaves out the potential of significant moral effects that happen over time and as a result of a kind of habituation. My own account, which will come in Part III, will focus on the latter feature as it pertains to what I will call the life of the reader.

1. Historical Dimensions

Today there are several issues that get discussed under the heading “art and ethics.” These include (1) whether there is any relation between art and ethics at all; (2) what that relation

is; (3) whether our experience of art contributes positively or negatively to ethics; (4) whether ethical concerns should have any bearing on aesthetic judgments of works of art; and (5) whether our experience of art can have any effect on us as moral agents. Issue (5) is the one we're interested in here, since solving the moral problem presented in Part I will require that moral agents *change* in a significant way. But before we get to the solution, we need to see how this issue has been dealt with historically.

The question of what moral effects art produces in audience members has its roots in ancient Greece, with Plato and Aristotle adopting almost diametrically opposed views about whether art improves or corrupts audience members.¹ While the issue cropped up periodically on the margins throughout history, it again gained prominence in the late nineteenth century, after the notion of disinterestedness had become popular in the work of Kant and Schopenhauer, and many of the issues that are at play today can be traced back to this period. Oscar Wilde famously said that art exists for its own sake. Clive Bell, founder of the Formalist movement in visual art, argued that the only element constitutive of art is "significant form." It was with the introduction of these similar views that the modern claim about connections between art and ethics was put into question. Asking whether, or what kind of, ethical judgments could be made about artworks was deemed irrelevant, or worse, logically incoherent.

Formalism emerged as a response to what had become a common way of looking at artworks: as mere didactic objects, as vehicles for delivering messages, usually moral in content. In the medieval era, prior to the Renaissance, much of visual art was religious art, serving to inspire religious devotion and to sermonize about good and evil. Art moved away from this theme, but not completely. We can call this view the *kernel-husk theory*, which Robert Scholes

¹ See Plato, *The Republic* and Aristotle, *The Poetics* for these views.

and Robert Kellogg² define as holding that an artistic work contains an “ethical chestnut” and that, if we remove the external husk of the work, we find the universal truth—the kernel—waiting for us at the center. They suggest that this way of approaching works of art reached its climax in the medieval era and is now out of fashion. They write that this “dichotomy between ‘kernel’ and ‘husk’” is a ‘critical concept that must now be learned; [it] no longer comes to us naturally.’³ Why does it not come naturally to us now?⁴ Because of the Formalist movement led by Wilde and Bell.

Wilde makes his famous art-for-art’s-sake claim in reaction to the kernel-husk theory. Whereas someone might think that the truth contained within a didactic work (which would be all valuable works on a kernel-husk theory) is limited to what lies inside the husk—that is, the kernel—Wilde claims that this is absurd. Truth in art is not a matter of correspondence; it’s “the unity of a thing with itself.”⁵ Wilde’s concern with the kernel-husk theory is that a work’s purported value would lie in its message, and not in its formal consistency, where he believes it belongs. The nature of the artwork, its being *art*, is ignored when one places such value on the ethical truth sitting in the center of the husk. Works of art are not to be shucked.

A related worry he has is that approaching artworks as vehicles for content treats them as *mere* vehicles—objects that have, at most, instrumental value. A work of art, in order to be given its due *as art*, should be handled as an object with intrinsic value. Wilde writes,

² *The Nature of Narrative*, NY: Oxford University Press, 2006.

³ *Ibid*, p. 150.

⁴ We might wonder whether narrative theorists are the best judges of what critical concepts come naturally. It might be that at a certain point it seems to be the case that the kernel-husk perspective is a learned concept, simply because narrative theory is so opposed to this way of approaching artistic works. We often implicitly accept the dictum that “a spoonful of sugar helps the medicine go down” in ethical lessons we learn as children. But even if it is a critical concept that is now learned, it’s undeniable that this way of approaching art is, and has been, prominent in art history.

⁵ Quoted in Frank Palmer, *Literature and Moral Understanding*, Oxford: Clarendon Press, 1992, p. 129.

A work of art is useless as a flower is useless. A flower blossoms for its own joy. We gain a moment of joy by looking at it. That is all that there is to be said about our relation to flowers. Of course a man may sell the flower and so make it useful to him, but this has nothing to do with the *flower*. It is not part of its *essence*. It is accidental. It is a misuse.⁶

He claims that making any instrumental use of a work of art is a *misuse*, like trying to make a flower useful. He is suggesting that it's *inappropriate* to treat the work as though it could be valuable for anything other than aesthetic pleasure.⁷ Aesthetic value exhausts the value of art. *Moral* value, then, is out of bounds.

Bell⁸ makes a similar argument. His view derives from his definition of “art,” which makes the possession of “significant form” the only necessary and sufficient condition for something being an artwork. What constitutes significant form is not clear in Bell’s writings, but for our purposes we can say that form itself is distinguished from content and is the only thing that counts toward a work’s status as art. Bell holds further that *any* work of art is a means to a good state of mind—i.e., “the state of aesthetic contemplation”—and, insofar as anything that is a means to a good state of mind is good, all works of art are good. This is the *only* kind of ethical statement we can make about a work of art. He criticizes Tolstoy for saying that art’s justification lies in its power of promoting good action. Moreover, he argues that it’s inexcusable to allow moral considerations to come into the discussion when comparing the merits of individual works of art, saying that whenever we make ethical judgments about works of art, we judge these works “not as works of art, but as members of some other class, or as independent and unclassified parts of the universe,” doing something akin to taking into “account the area of

⁶ Quoted in Palmer, *Literature and Moral Understanding*, p. 129.

⁷ But isn’t valuing the work for the aesthetic pleasure it gives still treating it as having instrumental value, with aesthetic pleasure perhaps the intrinsically good end? Wilde doesn’t address this worry.

⁸ *Art*, NY: Stokes, 1914.

the canvasses, the thickness of the frames, and the potential value of each as fuel or shelter against the rigours of our climate.”⁹ In other words, when we make moral judgments about individual works of art, we consider the work not as art, but as just some object in the physical universe. We cease to treat the work as art once we include it in any kind of ethical deliberation.

Most theorists today think of these Formalist reactions to the kernel-husk theory as *overreactions*. The kernel-husk theory places all of the value of the artwork in its message, while Wilde and Bell place the value entirely in its internal formal relations. Formalists hold that going beyond the formal relations is a logical error—the commission of a category mistake. In doing so we stop treating the work *as* art, much as we no longer treat a sculpture as art when we use it as a doorstep. Whenever we make ethical claims about a work of art, the object’s being art is now regarded as a contingent property. The object could just as well be non-art and the same inferences would still be possible. Arguing that Formalism is a swing too far in the other direction, Jerrold Levinson (1997) relates an anecdote about Jack Warner, an early mover and shaker in Hollywood: “When asked what the message was in one of his recent offerings, he replied that he just made movies, and that if he had wanted to send a message, he would have used Western Union.” Levinson claims that Warner is under the spell of the “formalists and art-for-art’s-sakers,” and that the notion of messages in art is not as absurd as Warner makes out. Levinson writes, “it seems hard to deny that artworks [...] very often do have messages, and far from inexpressible ones. In other words, many works are reasonably taken as *saying* something, in an extended sense, that is, as implicitly advancing some proposition, endorsing some perspective, or affirming some value.”¹⁰ Of course, this is not an endorsement of the kernel-husk theory, which claims that the message is all that is valuable in a work of art. Levinson argues that

⁹ Ibid, pp. 84-85.

¹⁰ “Messages in Art,” in *Art and Its Messages*, ed. Stephen Davies, University Park, PA: Pennsylvania State University Press, 1997, p. 70.

there can be more to value about works than their messages, but that *part* of their value could be in what they assert. This is in direct opposition to the claims made by Wilde and Bell.

Wilde and Bell defend views that Noël Carroll¹¹ calls *radical autonomism*. Carroll provides a framework from within which we can think about the issue of art and ethics, consisting of four views: (a) radical autonomism, the view that it is *logically* incoherent to make ethical assessments of artworks, as we've seen reflected in Formalism; (b) moderate autonomism, the view that even if it's not logically incoherent to assess artworks in ethical terms, moral defects and virtues do not contribute to the aesthetic worth of the work; (c) radical moralism, the view that it is always the case that moral virtues contribute positively and moral defects contribute negatively to a work's worth; and (d) moderate moralism, the view that it is sometimes the case that moral virtues contribute positively and moral defects contribute negatively to a work's worth.

As we are interested in issue (5) above—whether our experience of art can have any effect on us as moral agents—radical autonomism is the only one of the four views we need consider here. The other three views Carroll describes—moderate autonomism, radical moralism, and moderate moralism—address the issue of whether *moral and aesthetic value* are related in a substantive way. They are views related to what Gaut calls the “intrinsic issue,”—“whether the ethical value of artworks conditions their aesthetic value.”¹² We aren't interested here in the intrinsic issue,¹³

¹¹ “Moderate Moralism,” *British Journal of Aesthetics* Vol. 36, No. 3, 1996.

¹² *Art, Emotion, and Ethics*, NY: Oxford University Press, 2007, p. 9.

¹³ For views on the intrinsic issue, see Noël Carroll, “Moderate Moralism;” “Moderate Moralism Versus Moderate Autonomism,” *British Journal of Aesthetics*, vol. 38, no. 4 (1998), pp. 419-424; Berys Gaut, “The Ethical Criticism of Art,” in *Aesthetics and Ethics*, ed. Jerrold Levinson. (NY: Cambridge University Press, 1998), pp. 182-203; Robert Stecker, “The Interaction of Ethical and Aesthetic Value,” *British Journal of Aesthetics* Vol. 45, No. 2 (2005), pp. 138–50; James C. Anderson and Jeffrey T. Dean, “Moderate Autonomism,” *British Journal of Aesthetics*, vol. 38, no. 2 (1998), pp. 150-166; Daniel Jacobson, “In Praise of Immoral Art,” *Philosophical Topics* Vol. 25 (1997), pp. 155-189; Matthew Kieran, “Forbidden Knowledge: The Challenge of Immoralism,” in *Art and Morality*, ed. Jose Bermudez and Sebastian Gardner. (NY: Routledge, 2006), pp. 56-73; and Alessandro Giovannelli,

which is contrasted with the “causal issue,”—“whether exposure to works of art tend to affect us morally—morally to improve or morally to corrupt us.”¹⁴ We are interested only in the causal issue, but inasmuch as asking whether “works of art tend to affect us morally” as a result of aesthetic features of the work requires that we aren’t committing a category error, we must argue against the radical autonomist.

Now that I have sketched the historical dimension of the discussion about the relation of art and ethics, I need to provide some conceptual distinctions as well. As we are interested in the causal issue, we have to take into account the nature of the alleged effects on audience members, as well as the range of artworks and experiences under consideration. For this reason, we have to make three distinctions that will help us locate the accounts on offer as well as the one I will defend in Part III: (a) ethical vs. pre-ethical effects; (b) informational vs. formational effects; (c) particularism vs. non-particularism.

2. Ethical and Pre-Ethical Effects

In stating the causal issue Gaut rightly includes the possibility of art’s having both positive and negative effects. Broadly speaking, a positive effect would inherently make the moral agent *better*, while a negative effect would make the agent *worse*. But there is a third possibility: art might make someone *better prepared* to be a good moral agent, while not necessarily making the agent morally better. This wouldn’t, strictly speaking, be an ethical effect at all, since no moral change necessarily occurs as a result of being better prepared. For example,

“The Ethical Criticism of Art: A New Mapping of the Territory,” *Philosophia*, vol. 35, (2007), pp. 117-127. Also see my “Non-Branching Moderate Moralism,” *Philosophia*, forthcoming.

¹⁴ *Ibid*, p. 6.

someone might learn the lesson from some incident that thinking twice before acting is beneficial. This person, who was once a hothead, is now no longer so rash and presumably performs more carefully thought-out actions. If we ask, however, whether this change is *morally* beneficial, we will need more context. Does Mr. Previous Hothead now handle traffic situations better, so that he no longer puts others in danger in order to punish those that have cut him off? If so, the effect would be positive. But perhaps he uses what he has learned to exact revenge more effectively on co-workers that he doesn't like, now that he is able to think things out more fully and to concoct plans that will maximize their suffering. If this is the case, then the effect is negative. And if he handles traffic altercations more reasonably, but also exacts revenge more effectively, both as a result of what he has learned, then it's not clear whether the effects are positive, negative, both, or neither.

What the previous example illustrates is that learning to think things through and not to act rashly can be used for good or bad purposes, but taken alone doesn't constitute either a moral asset or liability. If this new skill had been acquired as a result of aesthetic experience, it would be what Jèmeljan Hakemulder calls a *pre-ethical* effect of art, since it's an "enhancement of abilities which are likely to help us in making ethical inquiries."¹⁵ Pre-ethical effects are distinguished from ethical effects, which are those that are directly connected to moral reasoning and behavior and implicitly get classified as positive or negative. When making moral decisions we *make use* of the pre-ethical abilities, which include "being able to understand the conflicting demands, being able to determine our own norms and values, and predicting the consequences of either option of [our] dilemma."¹⁶ These abilities—which don't necessarily exhaust the abilities relevant to moral decision-making—are perhaps essential to making decisions, but aren't

¹⁵ *The Moral Laboratory : Experiments Examining the Effects of Reading Literature on Social Perception and Moral Self-Concept*, Philadelphia: J. Benjamins Publishing, 2000, p. 4.

¹⁶ *Ibid*, p. 4.

sufficient for making good decisions. While they cannot be considered ethical effects, they still increase “the likelihood of such effects.”

It’s often implicit, but a condition generally held of any adequate account dealing with the causal issue is that the claims focus on ethical, and not merely pre-ethical, effects. An account that demonstrates the existence of an ethical effect *ceteris paribus* will be stronger than an account demonstrating the existence of a merely pre-ethical effect of aesthetic experience.

3. Informational Effects

A second distinction also has to do with the nature of the effects of aesthetic experience: informational vs. formational effects. Informational effects are those that occur when an audience member learns new information as a result of some aesthetic experience, while formational effects occur when an audience member learns a new skill or cultivates an existing skill. Some controversy surrounds claims made about informational effects, so we will spend some time looking at this controversy and how it might be resolved.

Consider the following example: I have read Dickens’s *A Tale of Two Cities* and after having done so, I claim to have learned some propositions about the French Revolution. In one sense this appears rather straightforward. Perhaps I learned that the Revolution occurred in 1789 and it would be strange for someone to question whether I have learned this new information, given that I didn’t know it before reading the novel and did know it after reading the novel because the true statement is asserted that the French Revolution began in 1789. But something odd also seems to be going on here, because the novel is full of false statements. How did I

determine that the statement about the Revolution was true while the others were false? In short, how can any information learned from an artwork be *justified*, when some of the statements are false and some of them are true and there are few explicit rules for determining which is which?

One possible means of sorting the false statements from the true is picking up on the *flavor* of claims to truth. For example, the use of precision—e.g., specific dates, full names, textual references—gives the reader the impression that even if statements about some of the characters are not to be taken as true, the more precise statements should be so taken. But the possibility of false precision should worry us here. Consider the rich and detailed descriptions that appear in the short stories of Jorge Borges—descriptions that have the ring of truth, but cannot possibly be true, perhaps because they contradict laws of nature or even of logic. Apparent precision cannot therefore be a reliable guide.

Perhaps then precision backed by statements made in other contexts supporting the claims made in the novel would do the trick. The statement made in the novel that the Revolution happened in 1789 gives me reason to think that it's true, so I go look up the dates in a history book and I thus have my confirmation. If the confirmation is necessary, however, then it looks like I learned the information only upon checking the history book and not by reading the novel. The cognitive value of reading *A Tale of Two Cities* that initially appeared to be found in my learning information about the Revolution would now be doubtful, since I still had to confirm it by consulting a history book.

One way to preserve the cognitive value of the novel is to claim that it *presents* the information in a different way from how it's presented in the textbook.¹⁷ This line of argument gives artworks a distinctive functional role, different from the role occupied by textbooks, and

¹⁷ See Colin McGinn, *Ethics, Evil, and Fiction*, NY: Oxford University Press, 1997, for an example of this argument.

these artworks would have cognitive value not so much in *what* is learned, but in *how* it's learned. But then the claim that I can learn something unique from my experiences of art is no longer true and threatens the view that there is distinct cognitive value found in art.

The view (a) that artworks can present us propositions—truths—that are not attainable from other sources, or give us easier access to such truths, and (b) that these truths are non-trivial is called *cognitivism*. These two components—what we might call the uniqueness and non-triviality components—of cognitivism have come under fire from non-cognitivists. We should look at how cognitivism can survive these attacks, since I believe cognitivism is a live option that should be pursued. Non-cognitive attacks generally focus on each component separately. The two most serious attacks upon cognitivism have been made by Jerome Stolnitz and T. J. Diffey, which I look at in turn.

4. Non-Cognitivism

Stolnitz¹⁸ attacks both the uniqueness and non-triviality components. In opposition to the uniqueness component, he makes what could be called the argument from dissimilarity. He argues that art isn't structurally similar enough to our paradigmatic truth-producing/discovering fields of inquiry, like science and history, to support the claim that art provides us valuable truths not as easily found in other areas of inquiry. Moreover, even when we consider a more controversial example of a truth-producing endeavor like religion, we see that religion has some structural similarities to science and history that the institution of art doesn't. Thus the case for art's being able to provide unique truths looks hopeless.

¹⁸ "On the Cognitive Triviality of Art," in *Philosophy of Literature*, ed. Eileen John & Dominic Iver Lopes, Malden, MA: Blackwell, 2004, pp. 317-323.

He turns his attention to the non-triviality component and provides what could be called the argument from absence of examples, based on some of our most revered artistic works—works such as *Pride and Prejudice*, *Oedipus Tyrannus*, *Bleak House*, *Uncle Tom's Cabin*, *Crime and Punishment*. Presumably if there are non-trivial truths to be found in art, these would be the exemplars of where to find them. Unfortunately for the cognitivist, Stolnitz is able to come up only with platitudinous statements. Thus, Stolnitz concludes, the propositions we glean from art can only be trivial, or if not trivial, as easily gained from other sources: “[Art] is replete with epigrammatic observations hardly distinguishable from folk sayings [and] all [of its truths] are proper to some extra-artistic sphere of the great world. So considered, there are no artistic truths. Not one.”¹⁹ Cognitivism thus fails.

I believe that the cognitivist can answer Stolnitz. Stolnitz's argument from dissimilarity doesn't really get him anywhere until he can convince us that uniqueness is necessary for having cognitive value or that the cognitive value that we are able to gain from our experience of art relies crucially on truths learned from other sources (a part of Diffey's non-cognitive attack we will consider shortly). A cognitivist could simply reject the notion that uniqueness is a necessary condition for cognitivism and I see no reason why we should feel compelled to maintain that non-triviality of what is learned is part of what constitutes the cognitive value of art.

The argument from absence of examples fails too if we can give a convincing reason different from Stolnitz's for why we can't provide non-trivial statements derived from artworks. One way is to explain Stolnitz's failure not, as he has it, by supposing that there are no non-trivial truths to produce, but by showing that it's extremely difficult to (re)produce them in a way that satisfies the constraints he has assumed.

¹⁹ Ibid, p. 321.

To get at this difficulty consider Peter Lamarque's²⁰ discussion of how it might be possible for moral learning to occur as a result of experiencing tragedy. While Lamarque is speaking specifically of moral learning, the point can be generalized to learning as a whole. Moreover, his discussion of the issues concerning tragedy and morality can easily be applied to all art. Lamarque suggests that there are two ways in which tragedy might be said to have moral content: by teaching or expressing a moral lesson and by presenting a moral vision. It's his discussion of moral lessons that I want to consider here, for he says that it is always easy enough to draw moral lessons from great works of literature, but the real question is "what status should be given to them, or [...] where they should be located within an adequate theory of literary appreciation."²¹

One challenge to producing an example of an artistic truth—moral or not— that might lead someone like Stolnitz to think that it cannot be done is that there is a very fine line between being too specific and too general. Lamarque makes this point in the following:

Perhaps the problem can be summarized like this: either the moral lesson is too close to the work, tied too specifically to the characters and incidents in the work, in which case it cannot function as an independent generalizable moral principle, or the moral lesson is too detached, too loosely connected to the specifics of the work to be perceived as part of the literary content or meaning that the work expresses. The tension here is precisely between the *derivability* of the moral principle and its *independence* as a general moral truth.²²

²⁰ "Tragedy and Moral Value," in *Art and Its Messages*, ed. Stephen Davies, University Park, PA: Pennsylvania State University, 1997, pp. 59-69.

²¹ *Ibid.*, p. 62.

²² *Ibid.*, p. 62.

Lamarque quotes Orwell to illustrate his point. Orwell says that one particularly shallow lesson to be learned from *King Lear* is “Don’t relinquish power, don’t give away your lands.” But Orwell believes a deeper lesson can be learned: “Give away your lands if you want to, but don’t expect to gain happiness by doing so. Probably you won’t gain happiness. If you live for others, you must live *for others*, and not as a roundabout way of getting an advantage for yourself.”²³ Orwell’s contention with the shallow interpretation—that of the Fool—is that it’s too watered down, too simplistic, and doesn’t give the Shakespearean work its due. It reduces a very complex and sophisticated statement—or set of statements, if it’s too complex to be summed up in one statement—to something that *seems* platitudinous in isolation. But even Orwell’s statement seems to suffer from the same flaw—only perhaps not to the same degree. Can *King Lear*’s moral lesson really be summed up in Orwell’s three sentences?

The problem is that we have to abstract away from the specifics of the work in order to produce something that could be applied to new situations with perhaps structural similarities to those situations depicted in the work. But as soon as we do so, much of the original content is lost, and the lesson suddenly seems trite. To re-gain substance, we bring it back to the work itself, within the literary framework. But now we have two choices: we can again risk being too tied to the situations of the work—overspecificity—or we can try to achieve the right level of specificity while capturing the valuable moral content that can be carried away from the work. According to Lamarque, however, we are now in danger of making thematic statements about the work itself. That is, we are no longer formulating moral statements, but statements about the work that contain some moral content. This isn’t something that we can take away from the work, since thematic statements won’t have anything useful to apply to the world at large. For

²³ Quoted in Lamarque, “Tragedy and Moral Value,” p. 62.

this reason Lamarque finds this approach flawed. He doesn't believe that trying to derive moral lessons from works is a fruitful strategy.

I think that this explains, in part, Stolnitz's difficulty in finding contentful and unique statements derived from art. What Lamarque has shown isn't that there are *no* moral lessons to be learned from art, but he has merely made the correct point that it's a high hurdle. It's thus a pragmatic point Lamarque is making here—we won't get very far trying to reproduce genuine statements derived from art. It's not a logical point, however, since he hasn't shown that it's impossible to produce examples of moral learning, just that it's not useful.

Someone might think that there is a logical point to make in this context, however. To see how this would go, consider the following argument. Let's look again at the non-triviality and uniqueness requirements. It might seem that if we are able to articulate the lesson, then it has to fail either one or the other constraint. If the lesson is truly unique, then we as audience members should be hard pressed to state the lesson in words. But if it's a lesson that we have learned, then it would seem that we would *have* to be able to state it in words. Otherwise, what reason would we have for the claim that we had learned anything? If we can state the lesson, then it would have to be reducible to a small set of statements—since this is the most audience members are perhaps able to do—in which case the lesson seems like a platitude, thus failing the non-triviality constraint. We have then a similar problem to Lamarque's of there being a tension between derivability and independence. The tension appears to be unavoidable because of the non-triviality and uniqueness constraints. I believe that it is this tension, as well as the one discussed by Lamarque, that leads Stolnitz to think that we won't *be able to* provide any examples.

Tensions are not logical obstacles, however. *All* that has been shown is that moral lessons learned from art are just *very difficult* to put into words. This shouldn't surprise us, though, if the

lessons are both substantive and uniquely derivable from art. In fact, *not* being able to articulate the truths turns out to be evidence of the work's genius, and not of its triviality. Unless Stolnitz is prepared to claim that he has the same level of artistic genius as Austen, Sophocles, Dickens, or Dostoevsky, his failure to produce non-trivial restatements of the truths found in art is to be expected.

So the cognitivist can defend herself against Stolnitz. A different attack on cognitivism is waged by Diffey.²⁴ This attack, focusing exclusively on the uniqueness claim, cannot be parried in the same way as Stolnitz's. As I remarked earlier, a cognitivist shouldn't take the uniqueness claim as part of her view if that just means that the truths learned from a work of art must be accessible from *no* other sources. There is a way of construing the uniqueness claim that a cognitivist might be swayed to adopt. If it's not possible to justify a truth you have learned from a work of art without consulting another, non-artistic source, then it might be thought that the cognitive value of the work, insofar as cognitive value consists only of propositional knowledge, is minimal or non-existent. Diffey makes just such a claim.

To get the full effect of Diffey's non-cognitivist attack, we have to consider the well-known theory of fiction proposed by Monroe Beardsley,²⁵ because Diffey intends to widen Beardsley's view to include the "mediums of art generally." Since Diffey's is a broader theory than Beardsley's, getting the core of Beardsley's view correct is imperative. Beardsley argues that the distinguishing feature of fictional writing is that none of the language appearing in fictional works is what J. L. Austin calls "illocutionary acts,"²⁶ acts performed by virtue of and in

²⁴ "What Can We Learn from Art?" in *Art and Its Messages*, ed. Stephen Davies. University Park, PA: Pennsylvania State University, 1997, pp. 26-33.

²⁵ *Aesthetics: Problems in the Philosophy of Criticism*, NY: Harcourt Brace, 1958.

²⁶ *How to Do Things with Words*, Cambridge, MA: Harvard University Press, 1962.

addition to the act of utterance.²⁷ For an utterance to be an illocutionary act, what Beardsley calls “constitutive conditions” must hold for the particular act.²⁸ These constitutive conditions are conventional linguistic rules already in place at the occurrence of the utterance. For example, in order for the statement “By the power vested in me by the prevailing government, I hereby pronounce you husband and wife” to perform the illocutionary act of wedding two individuals, the utterer must have the authority he claims to have. If an actor onstage says this to two other actors who are playing parts in a play, no wedding has actually taken place, because the constitutive conditions don’t hold. The utterer doesn’t have the appropriate authority. Similarly, warnings require the existence of certain constitutive conditions. I can’t be said to warn someone of some state or consequence if the state or consequence would obviously be desired by the warnee. If I say “I warn you not to open that door or you’ll be presented with a birthday cake with thirty candles” and you obviously want to be presented with this kind of cake, then it’s not that I have given you an empty warning—I haven’t given you a warning at all.

Beardsley argues that in fictional works, none of the language performs any illocutionary acts. Utterances in works of fiction are at most imitations of illocutionary acts. But, as Beardsley points out, imitative illocutionary acts are not illocutionary acts at all. We imitate illocutionary acts all the time in everyday conversation. For example, I might tell my friend about my wife’s promise to me this year that I don’t have to buy her anything for Valentine’s Day. I might do so by imitating my wife’s statement to me: “I promise you that I won’t expect any gift on Valentine’s Day this year.” I have said the words “I promise you that ...” to my friend, but I have not thereby made this promise to my friend.²⁹ I have imitated my wife’s promising me,

²⁷ “Utterance” is understood to include written statements as well as spoken ones.

²⁸ Austin describes a similar set of “felicity” conditions.

²⁹ The distinction here is, of course, between use and mention. I mention the word “promise” without using it.

which itself is not a promise at all. Beardsley claims that in fictional works, every utterance is an imitative illocutionary act and not a genuine illocutionary act.

If this is true, however, then, since *assertions* are illocutionary acts, no genuine assertions are made in fictional works. Thus even if a statement in a fictional work says “The Titanic set sail for New York on Wednesday April 10 1912” it’s not being asserted that the actual ship called the Titanic set sail on April 10 1912, but that the fictional ship called the Titanic set sail on April 10 1912 within the fictional world. There is no text-world reference on Beardsley’s view. Any reference is strictly confined within the fictional world.

Diffey’s argument is a variation on this theme that there is no text-world reference in fictional artworks. He begins by characterizing the standard stance—what he calls the aesthetic stance (and arguably what Nicholas Wolterstorff calls “the fictive stance”³⁰)—as one “taking the work to be holding up states of affairs for inspection, scrutiny, or, to use the traditional term, contemplation.”³¹ That is, the literary work, being presented merely for our reflection, depends on no text-world relation of reference. If any reference relation is born, it’s by virtue of some action on the *reader’s* part. Thus literary works on their own don’t refer to the real world. Any referential relation between the text and the world has to be provided by the reader, who would then have to have three pieces of knowledge: (a) what is being described by the literary work, (b) how the real world is, and (c) that the world is the way the literary work portrays it. But (c) is part of what we are supposed to learn from the artwork. Therefore, either we purportedly learn (c) from the artwork while having no justification for believing it (which would not be learning) or we have justification for (c), but such justification comes from non-artistic sources, such as life-experience or non-fictional sources (in which case we would not have learned the

³⁰ *Works and Worlds of Art*, NY: Oxford University Press, 1980.

³¹ Diffey, “What Can We Learn from Art?” p. 30.

proposition from the artwork at all). In either case there is no learning of propositions from artworks.

The problem here is even more serious than in Stolnitz's attack. While Stolnitz argued that any truth we learn from artworks can just as easily be gained from "extra-artistic" sources, Diffey's claim is that we won't be able to know *whether* we have learned any truths from an artwork without consulting "extra-artistic" sources. This is why I think Diffey's attack succeeds against the original statement of cognitivism, while Stolnitz's doesn't.

One objection that could be raised against Diffey's account is that it depends on Beardsley's theory of fiction, a theory which on its face is absurd. Of course there is text-world reference in artworks. When we read about Napoleon's marching on Waterloo in *Charterhouse of Parma*, the natural assumption is that the character named Napoleon is supposed to be *Napoleon*. That is, "Napoleon," the name used in the book, refers to Napoleon, the man who died on St. Helena. Thus there are some *parts* of an artwork that refer to the world, even if the whole artwork doesn't. But Diffey can respond to this objection by denying that he accepts the Beardsley account in full. Yes, indeed, "Napoleon" can refer to Napoleon. There may be assertions in the work. But how does the reader of *Charterhouse of Parma* conclude that the name refers to Napoleon, unless she is drawing from knowledge she already has? Even if we enlarge the scope, and ask whether an entire sentence refers to the actual world—i.e., whether the sentence is used to assert a proposition about the actual world—we still have to determine whether the words used in the sentence are intended to refer to the actual world. Again, doing this would require knowing enough about the actual world to know that the terms do refer. And if we take every purported truth learned from works of art as propositional, then we have to know whether the constituents of the proposition exist in the actual world. We have to know

whether the proposition is true, based on prior knowledge, before we can be said to have learned it. Therefore, any sentence we come across that refers to the real world, any proposition we assent to as applying to the actual world, we must have already known to be true before encountering it in an artwork.

There are more promising ways that a cognitivist can respond to Diffey's attack, each of which will be considered in the next chapter. One way is to argue that the information we learn from artworks might not be uniquely accessed in art, and may not even be recognized as true of our world without prior knowledge of how the world is, the cognitive value of artworks lie in their potential for delivering the content in a way distinct from other media. This line of argument is taken up by Richard Eldridge³² and Gregory Currie.³³ A second way would be, again to concede the point that we don't learn new propositions, but that the value is found in the acquiring of a deeper understanding of what the proposition means, what it implies, and how it affects our place in the world. Noël Carroll³⁴ and Martha Nussbaum³⁵ take this approach. A third way to respond to the non-cognitivist attack just considered would be to criticize the justification of the claim that we don't learn new propositions from artworks. Both Stolnitz and Diffey seem to think that the primary and most effective way to demonstrate that artistic truths are gained is by providing a list of them—i.e., being able to *articulate* such truths, but surely there are other ways to confirm that someone has learned a valuable truth from art than listing them. This kind of cognitivist response is adopted by Catherine Wilson³⁶ and John Gibson.³⁷ Finally, one could

³² *On Moral Personhood: Philosophy, Literature, Criticism, and Self-Understanding*, Chicago, IL: University of Chicago Press, 1989.

³³ "Moral Psychology of Fiction," *Australasian Journal of Philosophy* 73, 1995.

³⁴ "Art, Narrative, and Moral Understanding," in *Aesthetics and Ethics*, ed. Jerrold Levinson, NY: Cambridge University Press, 1998.

³⁵ *Love's Knowledge*, NY: Oxford University Press, 1990.

³⁶ "Literature and Knowledge," in *Philosophy of Literature*, ed. Eileen John & Dominic Iver Lopes, Malden, MA: Blackwell, 2004, pp. 324-328.

³⁷ *Fiction and the Weave of Life*, NY: Oxford University Press, 2007.

abandon the notion of propositional information as the only kind of relevant information, and focus more on what could be called phenomenal information. Thus the cognitive value of artwork would reside, in part, in learning *what it's like* to live life in certain ways and have particular experiences starkly different from one's own. This approach is adopted by Dorothy Walsh.³⁸

5. Formational Effects

Informational effects are contrasted with formational effects, which have to do with skill acquisition and/or cultivation. The lead-up here is much more straightforward, since there's no stress on uniqueness. It's not controversial to say that we can learn *how* to do certain things as a result of engaging with art. Indeed, *learning how to engage with art* is something that occurs as a result of engaging with art. It's very difficult to teach people how to properly approach art without just telling them to go out and do it. At a even more local level, there might be certain art movements—like Cubism in visual art, surrealism in literature, atonalism in music—which require audience members to learn how to engage with individual works falling under the respective labels and this can even be true at an individual level. I might have to acclimate myself with a particular method of story-telling, say, contained in individual literary works. I might have read James Joyce's *Ulysses*, and, after some degree of struggle, have gotten used to his stream of consciousness method. This ability, however, doesn't necessarily translate to other works, even others by Joyce, like *Finnegans Wake*. Later I read William Faulkner's *The Sound and the Fury*, and realize that there is a stream of consciousness theme here too, but that a bigger

³⁸ *Literature and Knowledge*, Middletown, Conn.: Wesleyan University Press, p. 1969.

challenge is the unusual time scheme. Thus I have to learn how to read this book. I might then experience the same *kind* of challenge of learning how to read Vladimir Nabokov's *Ada*, Anthony Burgess's *A Clockwork Orange*, Thomas Pynchon's *Gravity's Rainbow*. With each instance I have to learn how to engage with the individual work, but it could also be said that after awhile, I get good at learning how to learn how to engage with individual works. In other words, after having read all of these books, when I encounter a challenging work like *The Naked Lunch*, which is difficult to read for reasons completely different from the earlier reasons, I know (a) how to explore (b) how to read this book as a result of encountering and wrestling with the other works.

Of course, the kinds of skill-learning that can occur as a result of engaging with art is not limited to art-specific skills. We could, I am sure, come up with many examples of skills learned or improved on as a result of engaging with art. The kind of formational effects we are interested in are those that represent pre-ethical or ethical effects—i.e., skills the creation of or improvement on which is pertinent to moral reasoning and behavior. The primary formational account I will look at is that proposed by Currie.³⁹

6. Particularism and Non-Particularism

The third distinction I want to discuss has to do not with the effects themselves, as we've seen in the previous two, but with the focus of the theories dealing with the causal issue. My contention is that most theories that have been offered heretofore have focused primarily on

³⁹ "Realism of Character and the Value of Fiction," in *Aesthetics and Ethics*, ed. Jerrold Levinson, NY: Cambridge University Press, 1998.

individual works of art, and the power of these works to produce certain ethical effects, or individual experiences, such that if everything is happening correctly, a particular experience of a particular artwork has certain ethical effects. These theories I call particularistic, since they focus on individual elements, works, or experiences. I don't deny that individual works can have ethical effects, but I worry that we no longer see the beautiful complexity of artistic experience when we isolate works and experiences in this way. In the next chapter I will argue that most of the prominent particularistic theories suffer weaknesses because they focus too narrowly on individual experiences. More specifically, since the problem we are trying to solve requires a significant change in character, because particularism is focused on *moments*, rather than on *extended periods of time*, it is ill-equipped to give us a solution to the problem.

Particularism is contrasted with what I will call non-particularism. When we call an account non-particularist, this will mean that it addresses the power of regular and sustained engagement with art and an habitual familiarity with the experience to produce ethical effects. The relevant context here is just life as lived. Rather than plucking an artwork or an experience of an artwork out of the museum or off the shelf or away from the auditorium, and exploring its various moral facets, non-particularist theories address the question of whether making art a part of one's life, as life is lived, has ethical effects. It looks at aesthetic experience embedded within everyday experience and considers whether any ethical effects emerge as a result.

I will argue that in order to solve our problem, we will have to adopt a non-particularist, rather than particularist, theory. In the next chapter I will lay out the basic claims of the prominent particularist theories and give reasons for why they will not serve our purposes.

Chapter 5

Particularist Theories

In this chapter I am going to present several of the more prominent views in the literature on the causal issue. I will argue that each of these is particularist in nature, and, due to this feature, they won't help us solve the problem discussed in Part I. Although this alone is a good reason not to adopt any of the theories in the present context, there is a second reason why I find these accounts incomplete: with very few exceptions, they seem to me *highly artificial*. In some areas of scholarship, like literary and music studies, this wouldn't be troubling, since what matters most in those fields is textual content. Save for the most strident audience-response theories, the lives of the audience matter very little for interpretations of artworks. But here we are talking about the *value* of artworks—and specifically about the *moral* value of works—which suggests that life as lived should be as central to the conversation as the work itself. Unfortunately, when the scope of the discussion is limited to individual works or experiences, there is less opportunity to explore how this work stands to affect multiple lives—lives which, if considered in comparison to one another, would reflect mutually conflicting values and perspectives. Because works make very specific demands on their audience members, when one stands before the individual work, one often has to deal with it on *its* terms. Thus we often have to take up nearly identical stances to those of other audience members, which requires a great deal of bracketing off of our own particular perspectives.

When we strip away this artificial constraint, we can begin to discern new dimensions to aesthetic experience. When I am viewing a painting standing next to my wife, I may not be fully

in that impersonal stance described in the previous paragraph. When Jean is reading an Iris Murdoch novel, she may not be treating it as purely as a work of art. When John listens to Beethoven's *Sixth Symphony*, he may be treating it as a goad to creative thinking. There is no good reason to decry these impure stances toward works, even if we choose not to place the focus on those occasions when the works are not treated as art at all. Works of art are objects in our lives, and, as many objects tend to be, they are treated in ways both profane and sacred, depending on the circumstances. We should recognize this fact about our human lives, even if we ultimately decide that some stances toward artworks, irrespective of value, are not appropriate for the issue under discussion here. Therefore, independent of the failure of particularist views to present a solution to our moral problem, I contend that such views also fail to paint an accurate picture of the role aesthetic experience occupies in our lives.

I will begin by considering informational particularist views, followed by reasons why I find these to be inadequate for our purposes. Then I will consider a formational particularist view, followed by reasons why it too fails. This will set us up to consider what may be the most prominent view on the causal issue, that of Martha Nussbaum, which will be the sole topic in chapter 6.

Before we look at some of the popular informational particularist views, we have to make a further distinction. As we saw in chapter 4, responses to non-cognitivism can vary. Cognitivists can directly confront non-cognitivist attacks, as I did with Stolnitz's attack, and try to show that the argument fails, but they can also concede the non-cognitivist point, yet maintain that there is still cognitive value to be had from artworks, and that that value is primarily informational in nature. Within views adopting this latter approach, there are three relevant *kinds* of information that get prioritized, to varying degrees: (a) moral-theoretic, (b) moral-relevant, and (c)

phenomenal information. Moral-theoretic information is directly relevant to *types* of moral theories, such that the information gained points to principled frameworks underlying morality.¹ Moral-relevant information stands to serve as support for beliefs within particular ethical theories or application of these theories or anything appropriately bearing directly on moral decision-making, at either the practical or theoretical levels. Finally, phenomenal information has to do with perspective—i.e., what it’s like, with the “it” being a particular experience. If (a) and (b) concern knowing-that, (c) concerns knowing-what-it-is-like.² We’ll begin with moral-theoretic informational accounts.

1. Moral-Theoretic Informational Accounts

As I described the type of information earlier, moral-theoretic information has to do with moral theories. Of course, this is not necessarily information to be stated in ways characterized by moral philosophy. For instance, many of the parables in the New Testament are constructed in order to convey moral-theoretic information. In general, the information is related to the questions traditionally addressed in moral philosophy, such as How should one live? What is good? What is right?

Indeed, Colin McGinn thinks that one of the ways that artworks—fiction in particular—are cognitively and morally valuable is their functioning as parables of a sort. “Our moral

¹ This is what Hakemulder (*The Moral Laboratory*) calls a moral effect, an alternative to ethical and pre-ethical effects, defined as “the actual persuasion in favor of some moral position.”

² These options are not mutually exclusive. For example, (a) implies (b), since moral-theoretic information would be moral-relevant. Some instances of (c) would count as (b). Moreover, it’s likely that no informational account focuses *solely* on one of these types of information. Nonetheless, I am going to treat the following accounts as if they do, with this disclaimer that I recognize the highly idealized nature of my discussion.

understanding,” McGinn writes, “and the story form seem fitted for one another.”³ More specifically, *character* as an element of stories seems most relevant to moral understanding. McGinn discusses the “beautiful soul” as well as the evil character, and how these are illustrated in certain fictional works. We learn what it is to be and face down evil characters as a result of engaging with literary works.

Another influential theory in the moral-theoretic tradition is found in Martha Nussbaum’s *Love’s Knowledge*.⁴ She makes the bold claim that certain novels should themselves be considered works of moral philosophy. Her account is not strictly informational, however, since she claims that engaging with literature also cultivates in us a skill, which makes her account properly described as a hybrid account. We’ll look at her account in detail in chapter 6.

Perhaps the most influential informational particularist account is that of Richard Eldridge,⁵ intended to be a Kantian version⁶ of Nussbaum’s “Aristotelian” account found in *Love’s Knowledge*. Eldridge thinks that engagement with art—literature in particular—stands to teach us something about personhood and the importance of recognizing others and ourselves properly as persons. Drawing on Kant’s second formulation of the Categorical Imperative (“Always act in such a way that you treat others and yourself as an end and never merely as a means.”), he tries to make the case that narratives can demonstrate to us this importance. He argues that “autonomy and connectedness” converge in the unified phenomenon of personhood and that a large part of recognizing its moral importance centers on seeing their importance and the demands they place upon us. “Elaborating a perspicuous representation of our personhood will itself be a moral enterprise,” he writes, and “not a detached cognitive exercise. As we go on

³ *Ethics, Evil, and Fiction*, NY: Oxford University Press, 1997, p. 174.

⁴ NY: Oxford University Press, 1990.

⁵ *On Moral Personhood: Philosophy, Literature, Criticism, and Self-Understanding*, Chicago, IL: University of Chicago Press, 1989.

⁶ Conveyed in personal conversation.

in deploying our readings of cases and our strands of characterization of our personhood, taking our bearings as agents and evaluators from these developments, we will be making something of ourselves, taking up or abandoning one or another possibility of our personhood, in ways that will themselves be open to assessment.”⁷ Thus, in addition to the task of seeing the importance of personhood, we have the responsibility to determine what our individual personhoods will be. In this latter task, Eldridge says, philosophy and literature share a bond, but an uneasy one. Each draws from the other, with philosophy providing general principles and literature giving particular renderings. But describing them this way also gives a false impression, since neither does its respective task without effect from the other. Each is “implicated” in the other:

Narratives of the particular are always embedded within general presuppositions about persons and about what makes sense (literature is implicated in philosophy), and general presuppositions gain their sense and intelligibility only from the particular narratives that draw on them (philosophy is implicated in literature). [...]. Hence philosophy must both resist and embrace the perennial temptation to achieve full closure of understanding, to deny the potentially revolutionary importance of new cases, as it must both move toward general characterizations and retain its responsiveness to the particular. And hence literature must both embrace and resist its perennial temptation to concentrate solely on the particular, as it must respond to what is new and contingent while still drawing on general characterizations in order to focus our attention. And to say all this is to say that philosophy and literature must simultaneously embrace and resist one another.⁸

The reason for the tensioned embrace, claims Eldridge, is that, if T. S. Eliot is right that poetry stands to philosophy as that which makes the “Word Flesh,” the situation is indeed a dire one,

⁷ *On Moral Personhood*, p. 20.

⁸ *Ibid*, pp. 21-22.

because without the Flesh, “there can be nothing that we will honor as the Word,” and without the Word, “or a partial articulation of it in a general characterization of the person, there can be no coherent or recognizable poetry or narrative, no Flesh.”⁹ In other words, literature without philosophy would be too *particular*, while philosophy without literature would be too *general*.¹⁰

Thus, for Eldridge, the way art stands to enlighten us about personhood is twofold: (1) it augments and adds *flesh* to the general statements about personhood we might learn in other contexts from Kant and Charles Taylor, though this flesh constitutes additional knowledge about personhood; and in accomplishing (1) art also (2) gives us candidates, or a range of options, from which to choose or on which to base decisions regarding our own individual personhoods. It’s thus in this distinctive way literature works that it gives us this flesh and this new knowledge, though the new knowledge isn’t what might be considered distinctively artistic knowledge. Literature *brings home to us* what is important about personhood, even if it isn’t the exclusive source of learning of this importance. But as the Flesh needs the Word, Eldridge is not defending the view attacked by Stolnitz and Diffey that cognitive and moral value is constituted by *unique* knowledge.¹¹

⁹ Ibid, p. 21.

¹⁰ One is reminded of the comment made by Aristotle in *Poetics* IX that “poetry speaks more of universals, but history of particulars.” It’s curious that Eldridge’s claim about literature is almost the exact opposite of Aristotle’s claim about poetry.

¹¹ It might appear that I have mischaracterized Eldridge’s account as particularist, since he presupposes a number of instances of engaging with literature. In places, he does suggest that we should study the effects of literary narrative within the boundaries of life. He writes that the kind of criticism that we bring to literature is the “the very work that we implicitly and under various pressures of contingency carry out in our own lives as persons.” He says “Literary narratives come from our expressive acts and their narrative backgrounds; the narrative backgrounds of our acts are structured by assumptions that literary narratives help to make manifest.” To this extent, “literary narratives and personhood are, one might say, internal to one another.” (*On Moral Personhood*, p. 12) But once we get down to brass tacks, Eldridge is clear that we gain the relevant insights by engaging with *individual* and *appropriate* works. When we ask the questions that matter about personhood and about how to live, the answers lie, as he says, in “thoughtful and interpretative readings of narratives.” (Ibid, p. 67) That is, in *individual* readings of narratives. He demonstrates this by composing his book in such a way that the central chapters focus on (Kantian/Hegelian) elements of personhood and individual literary works—Conrad’s *Lord Jim*, Wordsworth’s *Prelude*, Coleridge’s “Frost at Midnight,” and Austen’s *Pride and Prejudice*.

Now that we've seen some examples of what I argue are moral-theoretic accounts, we can look at moral-relevant accounts.

2. Moral-Relevant Informational Accounts

In order to see the difference between moral-theoretic and moral-relevant information, consider the following example. Harold, upon hearing about the protests in the streets of Tehran in 2011, says to his friend, "Some people say that the U.S. needs to lend aid to the protesters, but I think that that's just plain wrong. After all, if the people of Iran didn't want a theocracy, they should have resisted it when the Shah was overthrown." Harold's argument is reflected in the following:

P1 The U.S. should aid protesters only in countries where the prevailing conditions were not assented to by the general population.

P2 The protesters in the streets of Tehran protest conditions to which the general population has assented.

C Therefore, the U.S. should not aid the protesters.

Now, in evaluating this argument, we can certainly take issue with P1. The Islamic Revolution occurred in 1979, while the protests occurred in 2011. How much time should pass before the conclusion is doubtful that the general population's assent at one point implies the assent at another point? Isn't thirty-two years a long enough passage? Thus P1 is implausible. But it's P2

that someone could really press Harold on. What reason does he have for thinking that the general population assented to the overthrow of the Shah? He might respond that it stands to reason: Revolutions are by the people, supported by the greater population. One doesn't have an uprising of minorities, unless the particular minority is powerful. But then the minority would already be *in* power, if it *had* power. Thus the Islamic Revolution had the support of the majority.

Suppose that someone responds to Harold by providing him statistics, perhaps accounts of how some of the population resisted the Islamists, of how many earlier protests had been staged, and other significant information, but these numbers fell on deaf ears, with Harold maintaining his position. He maintains his position, let's say, until happening upon Marjane Satrapi's graphic novel *Persepolis*, which gives Harold an illustration of how Iranian citizens were treated during the Islamic Revolution. He sees that it's not the case that the majority assented to the Revolution. He comes to reject P2 and eventually changes his mind about whether the U.S. should give aid to the protesters.¹² What Harold learns from *Persepolis* is moral-relevant information, since it bears directly on a decision that, while political, is also moral. This is a fairly unsophisticated example of moral-relevant information gained from a work of art, but the point should be clear.

There have been a number of well-known accounts focusing on moral-relevant information, the most famous of which are Noël Carroll's¹³ and Gregory Currie's.¹⁴ Carroll's account supports two central claims: in order even to understand many literary narratives, we have to bring our moral knowledge to bear on the text. By doing so, we subsequently achieve a

¹² There could, of course, be other reasons for why the U.S. should not aid the protesters. Thus, technically speaking, Harold rejects this particular argument for why the U.S. should not aid the protesters.

¹³ "Art, Narrative, and Moral Understanding," in *Aesthetics and Ethics*, ed. Jerrold Levinson, NY: Cambridge University Press, 1998.

¹⁴ "Moral Psychology of Fiction," *Australasian Journal of Philosophy* 73, 1995.

deeper *understanding* of the moral knowledge than we possessed prior to encountering the text. Thus one doesn't, strictly speaking, learn any new *knowledge*, but one better understands what one already knew. Carroll takes as his primary object of focus *narrative artworks*. He begins by noting how narratives are, by their very nature, incomplete. They have to be incomplete, since providing every detail would prevent the narrative's progress. "Every narrative makes an indeterminate number of presuppositions," he says, and "no storyteller portrays everything that might be portrayed about the story she is telling."¹⁵ There are gaps, by necessity, in any narrative. Sometimes, it makes sense to describe the blue sky that Emma Woodhouse sees upon her trip into town, but in other instances, when Austen describes her trip into town, there is no need to say anything about the sky's color. Does this mean that the sky is not blue when Emma makes this later trip? Do we assume that the sky lacks color in this instance? No. Unless we have been told that the sky portends rain, we as practiced readers might implicitly assume that the sky is blue in both instances—when we are explicitly told that it is blue and when we are not.¹⁶ This means then that in some instances readers fill in the unavoidable gaps. The storyteller, says Carroll, "must depend upon her audience to supply what is missing." This filling in is important not only to the storyteller, but to the reader. Filling in the gaps is "a substantial and ineliminable part of what it is to *understand* a narrative."¹⁷ If a reader were unable, or unwilling, to fill in the gaps, the narrative would likely make no sense. Carroll suggests then that successful communication from author to reader depends crucially on reader uptake of the narrative, and since uptake depends on the filling in, the reader must be able to fill in the gaps.

¹⁵ "Moderate Moralism," *British Journal of Aesthetics* 36, 1996, p. 227.

¹⁶ Of course, we might not assume anything about the color of the sky. Figuring out what features about the situation we assume and which things we leave undetermined would be an interesting project—perhaps best performed by empirical psychologists. The main point that Carroll is making, however, is that if we attend to certain features—such as the color of the sky—and the author has left these particular features under- or undetermined, then we as readers must fill in the gap.

¹⁷ "Moderate Moralism," p. 227, emphasis added.

What is required of the reader in filling in the gaps? In general the reader will have to bring to bear on the text items from his cognitive stock. That is, he will have to call on his existing set of background knowledge. There is more involved, however, than just items of knowledge. Emotional stock is often called upon as well. We as readers must have emotional responses to certain situations in order to complete the gappy narrative. The kinds of emotions we have will matter as well. Sometimes we will be called on to have *moral* emotions. “Many of the emotions that the audience brings to bear,” Carroll writes, “are moral both in the sense that many emotions, like anger, [...] possess ineliminable moral components, and in the sense that many of the emotions that are pertinent to narratives are frequently moral emotions, such as the indignation that pervades a reading of *Uncle Tom’s Cabin*.”¹⁸ Thus we have to have these emotions to make the narrative intelligible, and since these emotions are “moral emotions,” we are called on to make moral judgments in order to render the narrative coherent. Our moral powers, claims Carroll, are “activated” at this point. Once our moral powers are activated, there is more at stake than just rendering the narrative intelligible. In calling on our cognitive and moral stock in order to fill in the narratives, we also can acquire valuable learning, but learning of a specific sort. Carroll describes this learning in the following: “in mobilizing what we already know and what we can already feel, the narrative artwork can become an occasion for us to deepen our understanding of what we know and what we feel. Notably, for our purposes, a narrative can become an opportunity for us to deepen our grasp of the moral knowledge and emotions we already command.”¹⁹ He calls his view *clarificationism*, since readers are not acquiring any new knowledge, but making clearer what they already know. They begin to connect the dots between what they knew prior to coming to the work or, as Carroll puts it, they

¹⁸ Ibid, p. 228.

¹⁹ “Art, Narrative, and Moral Understanding,” p. 142.

“put together previously disconnected belief fragments in a new gestalt in a way that changes their moral perception.”²⁰ Thus a reader comes away from some works with deeper understanding, even if not having gained any new items of knowledge. The deeper understanding would, presumably, be applied to new situations outside the artistic context. In Carroll’s account, then, the moral-relevant information isn’t new information, but a new, more profound *grasp* of the information.

Currie argues that we can learn morally relevant information about *ourselves*, specifically in works of fiction. These works can help us determine what we value and, in some cases, lead us to reject certain previously held values. First, he makes a distinction between two kinds of imagination—imagination-as-theory and imagination-as-simulation. Imagination-as-theory occurs when we begin from a set of initial conditions a person might be in and then do some reasoning, perhaps from principles of folk psychology, concluding with an idea of what it’s like to be in that person’s situation. With imagination-as-theory we don’t engage in any sharing of the phenomenology of the person. Imagination-as-simulation, however, occurs when we try to represent the experience phenomenologically. For example, if I am trying to imagine what it’s like to be chased by a lion, I might begin from the initial assumption that a lion is giving someone chase and then use reason, based on principles of folk psychology that I possess, in order to conclude how it might feel. Or I might imagine *myself* being chased by a lion. That is, I begin by imagining the lion giving me chase and then, through running a simulation off-line (i.e., all behavioral mechanisms disengaged) I determine what kinds of mental states someone—based on the mental states I imagine I have within the simulation—being chased by a lion might possess.

²⁰ Ibid, p. 143.

Currie believes that being able to imagine by simulation is a great moral resource. By imagining in this way, “we can learn something about whether a goal is worth pursuing, for ourselves and for those we care about.”²¹ It’s quite easy to imagine how my actions will affect someone else, especially when the person I am considering is very similar to myself. Of course, our powers of imagination are not as reliable as our powers of perception, so we have to keep in mind that imagining the effects of my action might produce inaccurate results. For this reason, Currie believes that if we imagine by simulation, we inevitably trade off reliability for the relative ease of imagining in this way. This is where fictions come in. Fictions are “guides to the imagination.” They help readers get better at imagining. He writes:

They make it easier for us to weave together a pattern of complex imaginings by laying out a narrative; they give us, through the talents of their makers, access to imaginings more complex, inventive and colourful than we could often hope to construct for ourselves. Sometimes, by artfully withholding crucial bits of narrative information, they can, like inspiring sports coaches, bring us to the point where we can make imaginative leaps for ourselves.²²

Thus, in being forced to imaginatively fill in gaps in the narrative (reminiscent of Carroll’s account), readers exercise their powers of imagination.

Currie makes a further distinction—between primary and secondary imagining. It might be the case that, in, say, reading Iris Murdoch’s novel *A Severed Head* that I am prescribed to imagine that the character Georgie is distraught. Georgie’s experience, however, isn’t described by the narrator. It’s clear that I am supposed to imagine her being distraught, which is the primary imagining. It’s not likely, however, that I will be able to imagine her being distraught

²¹ “Moral Psychology of Fiction,” p. 53.

²² *Ibid.*, p. 53.

without placing myself in her situation. This would call for secondary imagining. Thus, *in order to* imagine primarily, I must imagine secondarily. In conducting secondary imagining we learn something about ourselves. Just as finding ourselves in real and novel situations informs us of whether we value the decisions we make and the outcomes that result, when we imaginatively put ourselves into fictional situations, by performing secondary imagining, we learn about what we value. With every instance of placing ourselves into these situations in which we would not likely find ourselves in real life, we acquire more and more knowledge about what we do and don't value.

Arguably, less robust accounts of how we learn morally valuable information most commonly take the pertinent information to be moral-relevant. When we turn to the criticisms of informational particularist theories in general, the point will apply across the board to moral-relevant informational accounts.

3. Phenomenal Informational Accounts

In order to see how phenomenal information differs from both moral-theoretic and moral-relevant information, let us consider the following fictional description by T. C. Boyle of a character who has been hit by a car driven by an American in the desert:

He'd had headaches before—his whole life was a headache, his whole stinking worthless *pinche vida*—but never like this. It felt as if a bomb had gone off inside his head, one of those big atomic ones like they

dropped on the Japanese [...] But that wasn't all—the throb was in his stomach too, and he had to go down on his hands and knees and vomit in the bushes [...]²³

This alone stands to give the reader some insight, since very few people have suffered such an event, but there is an additional dimension here—the character who has been hit, Cándido, is an undocumented Mexican immigrant in the U.S. He staggers away from the scene of the accident in order not to be detained by border agents. He and his wife América have no money and no place to live, camping in the desert for the time being. But now Cándido is very badly hurt. América tells him he needs a doctor, but there is no way that fetching one seems feasible:

He didn't need a doctor. He didn't need to put himself in their hands—his bones would knit, his flesh would heal. What would he say to them? How would he pay? And when they were done with him, the man from *La Migra*—the Immigration—would be standing there with his twenty questions and his clipboard. No, he didn't need a doctor.²⁴

Unfortunately, he *does* need a doctor, because the next day he lapses into a fever, and goes “out of his senses for three hours and more, gibbering and raving like one of those inmates of the asylum on Hidalgo Street.”²⁵ América informs him later that, in his delirium, he “gave a speech to the President of the United States, shouted out snatches of songs popular twenty years ago, spoke in a whisper to his dead mother,” “chanted, snarled, sobbed, screeched like a pullet with five fingers clamped round its throat; and finally, exhausted, [...] [fell] into a deep trancelike sleep.”²⁶ Sure that he was about to die and that she would not be able to make it alone in the new

²³ *The Tortilla Curtain*, NY: Viking Press, 1995, p. 16.

²⁴ *Ibid.*, p. 22.

²⁵ *Ibid.*, p. 23.

²⁶ *Ibid.*, p. 23.

country as an undocumented immigrant, América had “lost all hope and could think only of forcing her head under the surface of the creek till she drowned herself or leaping from one of the high crags and smashing her body on the rocks below.”²⁷ But just at that point, Cándido comes out of the fever and seems destined to survive.

It’s this second experience, as consequence of the first, that stands to transform one’s way of viewing the “illegal immigrant” problem. In today’s political world a common theme is what is called “self-deportation,” which is the ultimate goal of making life for undocumented immigrants so unpleasant, making the acquisition of even the basic necessities so difficult, that these immigrants “choose” to return to their native lands. But what Boyle’s description does is give us an entry into one of these “unpleasant” experiences, which is more than merely unpleasant. Cándido is so afraid of being deported that he chooses to roll the dice and try to recover on his own. Moreover, if he loses, if he doesn’t recover, América too will likely die, since she sees no way forward without him beside her. What this shows is the possibility that even likely death will not convince people to return to their own homelands. Thus gaining knowledge of this perspective stands to be politically valuable.

But there is a moral point here, too. A reader, perhaps imagining herself being run down by a car, assumes, first, that the driver will, if at fault, make restitution. Also, she assumes that medical attention would be provided for her, even if she happened to be uninsured; emergency rooms would still treat her for her life-threatening injuries. Finally, she might assume that she would be given access to a place of moderate comfort for recovery. These would be minimum consequences of being run over by a car. And none of these is provided to Cándido, because he is unable to lay claim to any of them. Perhaps he is owed these minimal considerations based solely on his being human. Nevertheless, the discussion of self-deportation should at least

²⁷ Ibid, p. 23

include concrete and particular descriptions of what it's like to be in this kind of situation. What these passages provide, in conjunction with other parts of the novel, is phenomenal information.

Phenomenal information, or knowing what it's like, received some spotlight in the philosophy of mind as it pertains to consciousness.²⁸ This kind of information is most closely associated with *experience*. When we talk about someone learning what *it's* like, the *it* is experience—what the experience of *x*, whatever *x* is, is like. It's been argued that this information is not different in kind from, but is in fact reducible to, propositional information. But can phenomenal information be reduced to a collection of learned propositions?

Take, for example, D. H. Lawrence's *Twilight in Italy*, a travelogue recounting Lawrence's voyages through the Alps in the early twentieth century. In order to determine whether phenomenal information is reducible to propositional information, we would have to answer the following two questions: (1) Is it possible for me learn what it was like for Lawrence to make that voyage? (2) Is such learning based on his relaying to me a set of statements about what was the case on his voyage? Answering (1) gets to the heart of the issue of whether phenomenal information is even relatable, while (2) focuses on the nature of this information.

Consider another example: hitting a home run to win the baseball game. Someone might tell me everything that was the case when she hit that home run, down to the way her legs were like rubber as she ran around the bases, how her hearing seemed to top out with the roar of the crowd, the way she worried all the way round the bases that she would miss one of them and have the run discounted. She could tell me everything she takes to be relevant to the experience.

²⁸ See Frank Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* Vol. 32, 1982, pp. 127-136; "What Mary Didn't Know," *Journal of Philosophy* Vol. 83, 1986, pp. 291-295; David Lewis, "What Experience Teaches," in *Philosophy of Mind: Classical and Contemporary Readings*, ed. David J. Chalmers, NY: Oxford University Press, 2002, pp. 281-294; Thomas Nagel, "What Is It Like to be a Bat?" *Philosophical Review* Vol. 83, No. 4, 1974, pp. 435-450.

Now, if phenomenal information is reducible to propositional information, the exhaustive list would, by implication, successfully relate the experience.

But we can imagine more or less engaging and affective descriptions, even if content-wise the descriptions are the same. It seems to me quite obvious that a monotonous and cold description of Lawrence's voyage would convey to me much less than his engaged and emotional descriptions do, even when the content in the two descriptions is identical. There is then something about the *way* the content is delivered that factors into phenomenal information, which results in new content. The form affects the content—in some cases, even adding to it. Thus phenomenal information is not reducible to propositional information.

The person most responsible for bringing phenomenal information and art together is Dorothy Walsh, who, in her book *Literature and Knowledge*,²⁹ argues that the acquiring of phenomenal knowledge is what makes literary experience valuable. To support this claim, she contrasts actual and virtual experience. When we observe people's actual experience, she notes, we see differing degrees of awareness. While commuting between work and home, for example, some people see the sun set out their windows, and perhaps pull the visor down to avoid being blinded, but they don't see the sunset. That is, they don't fully experience the setting of the sun, the bright luminance occupying a portion of the visual field, the dwindling rays dropping below the horizon occupying another portion. Perhaps the driver curses the glare on the dash, infers rain tonight based on the tint of clouds; but his perceptiveness is weaker than, say, a person who sits on a park bench and intentionally puts herself within view of the sunset in order better to perceive the event.

There are, of course, other ways in which people fail to fully experience the world around them. For instance, I might (mis)perceive my boss's way of looking at me as evidence that I will

²⁹ Middletown, CT: Wesleyan University Press, 1969.

be getting that raise after all, even though I have been late to work an inordinate number of times within the last month. I deceive myself by seeing my boss's face in a way utterly at odds with the way a disinterested observer would—I am deceiving myself, but not by telling myself a lie, suggesting that there is a sense of truth other than correspondence at play here. I am lacking what might be called an accurate experience, even though my experience corresponds to the facts. If I were to describe the experience to someone who was not present, sticking to the facts, this other person too might agree with me that I am in store for a raise. But that's because the experience itself is affected by my optimism and my refusal to recognize that my excessive tardiness won't be overlooked by my employer.

Walsh suggests that the sense of truth different from correspondence that we're looking for is *authenticity*. My experience is inauthentic, even if corresponding to the facts, because I should know better than to perceive my boss's looking at me as promising. I should experience the world in the light of what is the case, which is that I am habitually late. More than this, I should experience the world in light of what is the case in general—i.e., as open as possible to the way the world really is. Walsh says that having authenticity of experience requires “spiritual freedom.” “This freedom,” she writes, “can be something hard won by deliberate discipline, but it can also be the unsolicited gift of temperament or of happy circumstances. We have all encountered free spirits who seem to be naturally free. To say this is no more than to acknowledge the hard fact that what some of us must win, others just have.” Some of us possess the ability to have authentic experience, while some of us don't, but this shows that we recognize authenticity as a measure of experience already. If we already recognize authenticity and inauthenticity in actual experience, we should be able to recognize it in virtual experience, the

kind of experience that literature is best suited to convey and share. Indeed, one of the ways we evaluate literary works is by judging the degree of ability to depict authentic experience.

If Boyle's descriptions of the homeless undocumented immigrants come off as appearing to be written by a middle-aged white native of the U.S., the work's aesthetic and cognitive value is diminished. This mode of appreciation extends to the author as well. Walsh says that "it is appropriate to suppose that at least part of the talent of the talented literary artist is his greater perceptiveness with regard to the possibilities of experience." Literary works, through their depictions, provide us "something *to be* experienced," but they also do something above and beyond this—they provide us "with experience of experience, and this is different."³⁰ The value of literary art for Walsh is located primarily in the conveying of authentic experience. Thus literary truth—artistic truth, if we generalize—is not correspondence—an assumption Diffey implicitly makes—but authenticity, which Walsh glosses as the property underlying the comment "It rings true" and absent in the comment "It rings false."

But we might want to push for a stronger criterion here, by considering what might be a better test for artistic truth than what Stolnitz and Diffey assumed necessary—i.e., the listing of propositions learned from a work of art. Catherine Wilson,³¹ for example, argues that the relevant behavioral criterion for determining whether someone has acquired artistic truth is that he or she be able to apply the new knowledge. "Learning," Wilson says, "applies primarily to a modification of a person's concepts, which is in turn capable of altering his thought or conduct and not primarily to an increased disposition to utter factually correct statements."³² Thus we have to look at how someone behaves in certain relevant conditions after having had the

³⁰ Ibid, p. 124.

³¹ "Literature and Knowledge," in *Philosophy of Literature*, ed. Eileen John & Dominic Iver Lopes, Malden, MA: Blackwell, 2004, pp. 324-328.

³² Ibid, p. 327.

purported authentic virtual experience and see whether there is an appropriate shift. Wilson uses as an analogy a person who claims to have learned the principles of civil engineering.

Let's say that we ask this person to recite the principles she has learned. Possibility 1: she is able to do so. Would we necessarily take her to have learned the principles of civil engineering after having seen her recite these principles? Imagine that we place her before a proposed plan to build a bridge and she is unable to tell us how the principles are evident in the plan. Further suppose we give her a problem in which the developers were unable to proceed because of a fairly obvious (to people who know the principles of civil engineering) theoretical mistake and she was unable to tell us what the mistake is. If we find her unable to accomplish either practical task, we would likely deny that she had learned the principles, in spite of her ability to recite them from memory. Therefore, the ability to recite statements is not sufficient for learning.

Possibility 2: she is not able to do so. Try as she might, she can't get the principles right, she forgets many of them, misstates others. But when presented with the two tasks described above, she is able to accomplish them both effortlessly. We would likely say that she has learned the principles, even if she is unable to recite them. Therefore, the ability to recite the putatively learned principles is not necessary for learning to have occurred.

But what has been established, according to Wilson, is that being able to apply the principles, being able to implement the concepts, is necessary and sufficient for learning. It's this ability, and not the ability to recite, that is the one we should focus on. Thus one possible criterion for authenticity of shared experience is the ability to implement the phenomenal information learned. If I, a normal person with normal sensibilities, after having read *The Tortilla Curtain*, am still unable to see undocumented immigrants as having already faced severely difficult trials, or if I don't recognize the hardships of Dust Bowl farmers after reading

The Grapes of Wrath or I can't feel the horror of the Jewish concentration camp experience in Germany and Poland after having viewed *Schindler's List*, then the virtual experience is what we could call inauthentic.

But aren't *seeing that, recognizing that, feeling the horror that*, not really implementing the new knowledge insofar as I could still treat the undocumented immigrants I encounter on the street with scorn, even though I recognize they have had a tough struggle to get here. I might still cling to my preconceived notions that they get what they ask for, even if what they get is unpleasant. There has to be something deeper than mere recognition—perhaps what Stanley Cavell calls “acknowledgement.” John Gibson³³ discusses this notion in relation to literature. Gibson describes two figures—the Sadist and the Simpleton—who possess the relevant knowledge about a situation but fail to act on that knowledge. The Simpleton observes a person undergo an injury. He knows that the person is suffering, he knows that he is the only one who can provide aid, he knows that he is the only one who can call for an ambulance. He doesn't provide aid, doesn't call for an ambulance, but stands motionless and watches the person suffer. The Sadist watches a performance of *Medea* and laughs at her suffering, experiencing genuine amusement as a response to the pain and betrayal depicted on-stage. He knows what suffering is, knows that he is observing a (depicted) instance of it, and has an inappropriate response.

Gibson uses these two examples to show that merely having knowledge isn't, on its own, a moral benefit. What is lacking in both cases is the connection between knowledge and engagement with the world. There doesn't seem to be any contact between the knowledge of suffering possessed by the Sadist and the Simpleton and their desires. In other words, neither is motivated by the knowledge of suffering; neither of them *acknowledges* the suffering of others.

³³ *Fiction and the Weave of Life*, NY: Oxford University Press, 2007, ch. 3.

Thus what we want in cases of phenomenal information is, first, for the sharing of the virtual experience to be *authentic*, a determination of which can be made in light of the way the audience member implements the phenomenal knowledge gained, and, second, for this implementation to occur in part through acknowledgement of the similar experience of others in non-artistic life. But if the having of mere knowledge is not enough, if it doesn't create in the agent a desire to acknowledge the experience of others, what does? This leads us into the main problem for particularistic informational accounts—the problem of the missing desire.

4. The Problem of the Missing Desire

Gibson's statement of the problem of acknowledgement relates back to Blum's John/Joan and Theresa examples discussed in chapter 1. The Simpleton isn't able to recognize the suffering of others as bad for them and that the badness might commit him to helping alleviate that suffering, while the Sadist is an example of someone motivated purely by what Schopenhauer would call malice. The Simpleton is the extreme version of the person who hasn't yet become a person moved to act by empathetic concern, or what I called compassion in chapter 2. If art is going to begin to solve the problem for morality, as presented in part I, it's going to have to do two things: first, it should be able to transform the Simpleton, or less extreme versions of the Simpleton, into a person more sensitive to the claims of others; second, it should be able to address the problem of empathic accuracy by improving on the four areas I outline in chapter 3.

Notice that one reason for the Simpleton's moral failure is the lack of an appropriate desire to alleviate the suffering of a person when it's possible to do so. As I say in chapter 2, this

needn't require altruism, since desiring to alleviate suffering is consistent with judging at the same time that acting in such a manner is not the right thing to do. It's very difficult to see, however, a way for a particularistic informational theory to demonstrate the formation of this kind of lasting desire, since informational theories seem apt to change existing beliefs, but not desires.

One way to get around the worry that informational theories can show only that beliefs can be changed is to say that the Simpleton can come to have the right desire so long as he acquires the right beliefs in addition to the ones he already possesses. For example, I might possess the following set of beliefs:

(B1) a belief that people should not suffer unnecessarily

(B2) a belief that when people do suffer unnecessarily I should do my part to mitigate their suffering

(B3) a belief that Rose who is in front of me is not suffering

Given my set of beliefs, I don't see anything amiss with the world, so I have no desire to alleviate any particular suffering. Suppose, however, that (B3) is replaced with

(B4) a belief that Rose is suffering

Now, given that I have (B2), I might form the desire to mitigate Rose's suffering. New desires can then emerge from a new set of beliefs, which might give informational theorists hope.

One might worry, however, that a desire has been smuggled in, since there might be the following background desire obscured by the original circumstances

(D1) a desire that I mitigate suffering whenever I am in a situation in which I should mitigate suffering

If this is the case, then no new desires are formed, but a new object of an existing desire is found as a result of a new belief. But then the Simpleton won't have this desire prior to encountering any work of art, so the new belief wouldn't help. What the informational theorists must show in order to address the first part of the moral problem is how art can create in audience members new desires—that is, desires that would play a role in the formation or cultivation of the motive of compassion in audience members. None of the accounts that I have considered spend time addressing this requirement, but there are two ways that these accounts could do this: (a) make the case that individual works or experiences can create in audience members the right desire(s); (b) show that the relevant desire is there all along, waiting for the right beliefs, as in the Rose example.

Let's consider (a) first. How could a work produce a desire in someone? First, the right desire would have to be something like (D1)—i.e., a desire linked up to some norm such that one desires to do the action prescribed by the norm. If this is the case, however, then it's *this desire*, and not the moral-theoretic, -relevant, or phenomenal information that is doing the heavy moral lifting. What is putatively an informational account would actually be primarily a desire-centered account.

Second, it would appear that an undue amount of pressure is now put on the individual work or experience. While there is nothing logically amiss here, psychologically it is much less plausible that individual works can produce such robust desires in audience members. What is more plausible is that the aggregated effects of multiple experiences results in the eventual formation of a robust desire, but then this moves us away from the particularistic approach.

Let's consider (b) then. Perhaps the relevant desire is already in place, but the new knowledge, or deeper understanding, is sufficient for engaging this desire. But what exactly would the relevant desire be? Maybe it's something like the following:

(D2) a desire that one do what one should do

This is a pretty broadly described desire, however, and extends beyond morality to normativity in general. It might be said that it's not too broad, since it combines with the various pieces of information learned from the work and, provided that the information is concerning what one should do, we have what we need. In order for an informational account to succeed here, it would have to make the case that the relevant desire (D2) is not sufficient for engaging with the world in a morally worthwhile way and that it *must* be combined with the information learned from the work.

Take Carroll's discussion of *Uncle Tom's Cabin* as an example. Carroll argues that people might approach this novel believing that slaves are moral agents and that we should treat moral agents with a minimum level of respect, while still not getting that slavery is a violation of that respect. Once someone engages with the narrative, sees how slaves were treated, how slaves were fully human with complete lives, this person might understand in a new way just how

egregious slavery was. His epiphany extends beyond new beliefs, because he already had the relevant beliefs. But he now understands the immoral nature of slavery in a way distinct from having the appropriate beliefs. Now, in order for a reader to treat slaves in a moral way as a result of reading *Uncle Tom's Cabin*, there must be a desire already lying in wait for this new understanding. But if this desire is something like (D2) then it, combined with the already existing beliefs, should have resulted in treating slaves in the morally appropriate way, which would involve desiring that slavery be abolished. Carroll's claim, however, is that there is something previously lacking in people who finally got it when they read the novel: the understanding that slaves were human beings and not chattel. Thus, just as in the earlier option (a), what emerges from this new understanding must be a new desire—*viz.*, the desire that slaves be treated like human beings and not chattel. In other words, a desire more specific than (D2). If the reader had possessed this desire prior to reading the novel, it just seems downright irrational for the person not to actually treat slaves like human beings. What this example suggests is that even when an appropriate coarse-grained normative desire exists prior to the engaging with a narrative, it's nonetheless necessary that the work cultivate a finer-grained desire in the audience member. But then it's the desire-cultivation, and not the new information, that is the relevant moral change.

Therefore, informational accounts fail to address the need for the formation of the appropriate desire in the Simpleton. A second reason why these accounts won't help us solve the moral problem found in part I is that, because they are particularistic, they cannot address the problem of empathic accuracy. It's highly implausible that the formation of new beliefs alone will have any bearing on our ability to decenter, our acquisition of metaknowledge of our mindreading successes and failures, our motivation to mindread well, or the mechanisms

underlying mindreading.³⁴ Such changes require the passage of time, the process of habituation, and this possibility is forestalled by particularistic, informational accounts. We should thus turn our attention elsewhere.

5. Formational Accounts

As I noted earlier, formational accounts are more straightforward. Ironically, the number of prominent formational account is low. The most prominent formational account belongs to Currie as well.³⁵ Currie begins his discussion by noting how crucial the ability *to plan* is for good moral decision-making. He says, “Many, perhaps most, of our moral failures come about not because we choose the bad over the good, but because our choices, whatever they are, are not well planned.”³⁶ Being a good planner, of course, is not sufficient for being a good moral decision-maker. To see why we need only consider the meticulous serial killer, like Hannibal Lecter, or the well-organized kidnapper, like Clegg, in John Fowles’s *The Collector*. They are consummate planners, but pure evil. Is becoming a good planner as a result of engaging with art merely a pre-ethical effect then? Let’s see how the rest of the account takes shape in order to answer this question.

Currie has to show how planning relates to art—fictions in particular. He argues that someone can become a better planner as a result of engaging with a specific kind of fiction—that containing “realism of character.” The kind of planning involved in these fictions utilizes

³⁴ With the possible exception that new beliefs might be incorporated into an audience member’s store of folk psychological principles. In order not to get off-track, I will leave this issue open here.

³⁵ “Realism of Character and the Value of Fiction,” in *Aesthetics and Ethics*, ed. Jerrold Levinson, NY: Cambridge University Press, 1998, pp. 161-181.

³⁶ *Ibid*, p. 164.

imagination and is contrasted with planning using only theoretical inferences, the same distinction we saw him use in the earlier account.³⁷ Planning as theory occurs when we begin from a set of circumstances and beliefs about how people would generally behave in those circumstances. We then make inferences by plugging in various courses of action, determining how the courses would fare, given the initial circumstances and set of beliefs. Finally, we judge the preferability of the outcomes, “based on further beliefs we have about what kinds of outcomes we are most likely to find rewarding.”³⁸ (165).

But there is a second kind of planning, involving imagination, that Currie argues is more appropriate in some circumstances. He writes

Think of the planning we do when we construct our next chess move, settle on the most winning way to ask someone for a date, or decide how to move into the side street against heavy oncoming traffic. For one thing, the inferences involved in these sorts of cases would seem to be extraordinarily complicated, with a vast number of premises and intermediate steps required in order to get to the goal: a decision about what is best to do. That would certainly be the case where, as in my examples, what you decide to do depends on what you reckon someone else will decide to do; and a great deal of morally relevant action is of that kind. For another, the experience of planning—unless one does it in an unnaturally cerebral fashion—does not seem to fit very well with the model.³⁹

This kind of planning requires “imaginative projection,” which involves placing ourselves imaginatively into the initial circumstances and then, while still within the imagined situation, trying out different courses of action, imaginatively, to see which of them can get us to the preferred outcome. While planning like this trades reliability for ease, just as Currie argues in the

³⁷ See section 2.

³⁸ “Realism of Character and the Value of Fiction,” p. 165.

³⁹ *Ibid.*, p. 166.

earlier account, the ability to imaginatively project is morally beneficial. Take someone—Smith—who is miserable. I want to understand why Smith is miserable, so I imaginatively project myself into Smith’s perspective. “I imagine doing the things I know or guess Smith to have done, and the act of so imagining causes in me feelings akin to those of misery.”⁴⁰ If the imagining goes well, I now understand something about why Smith is miserable. But I know something relevant to my own plight as well: I know how I would feel in that situation. And if I know how I would feel in that situation,

I am starting to plan, since to plan is to have a view about which actions would have good or desirable outcomes and which would not. So projecting myself into the life of another has, potentially, the double function of telling me about his mental life, and about my own possible future course of action; whatever I do, I had better make sure that things don’t turn out *that* way for me.⁴¹

Imaginative projection is morally valuable then because it motivates one to empathize with others, which, Currie suggests, makes one more likely to take into account the interests of others. It could also have moral value because it guides us in planning our own lives.

But where does fiction come into the picture? Up to this point we have only considered imaginative projection *simpliciter*; it hasn’t been linked up yet to our experiences in fiction. Currie thinks that engaging with realist fiction teaches us how to see our own lives as potentially structurable in the ways fictions are structured. We imaginatively project into the lives of characters and by “living that life, or important moments in it, in imagination, we may learn something about how to see our own lives as connected wholes capable of being structured in a

⁴⁰ Ibid, p. 169.

⁴¹ Ibid, p. 169.

planned way.”⁴² We begin to see our own lives differently as a result of seeing the fictional lives in the same way. Thus we become better situated, after having adopted this perspective, and consequently, we can become better planners in our own lives. There are two features of fiction that facilitate our learning how to become better planners: (1) it allows us to disengage from real life and “to dedicate scarce mental resources to the task of reflecting on the result of the projection, whereas in real life we cannot easily take time off from our obligations as social agents;” and (2) through provided commentary in the work, the fiction can “encourage and guide that reflection.”⁴³

Currie completes his account by acknowledging that fiction isn’t the only activity that could improve our ability to plan. If someone has a particularly robust faculty of imagination, then she might be able to plan using imagination and consider a range of possible courses of action. But good fictional narratives “give us, through the talents of their makers, access to imaginings more complex, inventive, and instructive than we could hope to make for ourselves,”⁴⁴ which for most of us will compensate for our less than robust faculty of imagination. This gives us the opportunity to exercise our faculty of imagination while also seeing how certain plans of action turn out in specific sets of circumstances.

We can infer from Currie’s account the general way that most formational accounts would work. In order to engage with artworks, we have to perform certain actions or activate certain cognitive systems, which requires the possession of certain abilities. These abilities would play an important role in moral reasoning. Engaging with artworks thus makes someone a better moral reasoner by compelling her to cultivate abilities crucial to moral decision-making.

⁴² Ibid, p. 170.

⁴³ Ibid, pp. 170-171.

⁴⁴ Ibid, p. 171.

6. The Problem of Application

But skills alone have very little power to improve someone's moral reasoning, if the individual doesn't know *when to utilize which skills*. For example, if Currie is correct that I can learn better how to plan using imaginative projection, I might be better off in situations that call for effective planning, but there are going to be moral situations in which planning is not necessary, and may be even be obstructive, so I will need to have the necessary knowledge to determine which kind of situation I find myself in. Without this knowledge I am not necessarily better prepared for moral decisions and may even be worse off as a result. In other words, formational accounts rely on there already being moral-relevant information available, which casts doubt on the claim that the skill is the crucial component acquired as a result of aesthetic experience.

One might think that Currie builds into his account an informational component, since he says that, using imaginative projection, we can learn things about the lives of others and that it can help us in determining what factors to consider in planning. Also, he claims that fictions help us to see that our lives can be structured in planned ways. This suggests that the information we learn already tells us when to plan and when not to plan. But notice that these pieces of information don't give us insights about situations calling for planning, but how *best to plan*. In other words, this is not moral-relevant, but planning-relevant, information.

What we have here is what might be called a problem of application: even if we learn new or cultivate existing skills, we won't necessarily know when to apply these skills in ways that make us better moral reasoners. At best, particularist formational accounts are able to show that art produces pre-ethical, rather than ethical, effects.

7. A Hybrid Account?

I've raised worries about the most prominent particularist informational accounts and particularist formational account. Regarding the former I described the problem of the missing desire, which was that it seems that informational accounts don't have a way to show how individual aesthetic experiences can cause the requisite desire necessary to prompt the Simpleton to acknowledge the suffering of others. Regarding the latter, I described the problem of application, which was that the formation or cultivation of a moral skill does not constitute an ethical effect when there is lacking the requisite knowledge of when to implement this skill. At best, art would produce a pre-ethical effect in these cases. At this point I am tempted to conclude that what explains the failures of both of these kinds of account is their particularistic nature. The emergence of the appropriate desire can be explained when we take into account the long-term, aggregative effects of aesthetic experience, while we can acquire practical wisdom that will allow us to know when to use which skills. But there is one more path that the particularist can pursue: defending a hybrid account, which consists of both claims that individual aesthetic experiences can cultivate moral skills and that such experiences can produce the relevant knowledge of when to apply these skills. The most famous hybrid account is Martha Nussbaum's, which will be the topic of the next chapter.

Chapter 6

Nussbaum's Hybrid Particularist Theory:

Learning Jamesian Perception

At the end of chapter 5 we considered the possibility that a hybrid account might work, whereas the purely informational and purely formational particularist theories discussed there failed to show what they purported to show. In this chapter I will consider what is arguably the most famous theory of how art can morally improve audience members, Martha Nussbaum's theory laid out in her book *Love's Knowledge*.¹ Her account is complex, interwoven with commentary on great, "high" novels like Henry James's *The Golden Bowl* and *The Ambassadors* and Marcel Proust's *Recherche a la Temps Perdu*. She argues, in places, the broad point that novels themselves can be works of moral philosophy, and, consequently, we shortchange ourselves as ethicists and teachers when we fail to include this medium in our investigations. I interpret her claims to be about the power of rare, *individual* works, which is why her account is described as particularist.² The passages of *Love's Knowledge* most germane to the issue addressed here are

¹ NY: Oxford University Press, 1992.

² Interpreting her account as particularist isn't a straightforward matter, however, since she says things like this:

[W]hat I am calling for [is not criticism that simply mines the work for a set of propositional claims, but rather] an investigation of that which is expressed and 'claimed' by the shape of the sentences themselves, by images and cadences and pauses themselves, by the forms of the traditional genres, by narrativity, themselves. It seems unlikely that this richer ethical task can be carried out by someone who is not in *the habit* of attending to these things. (*Love's Knowledge*, pp. 191-192, my emphasis)

That she alludes to "the habit" of performing certain acts of attending to the narrative devices suggests that she has in mind continued, sustained engagement with literary art. But in other places she states clearly that she believes that certain works, on their own, bear the potential to transform readers, in the same way that readers can be transformed by moral philosophy. Thus, for the purposes of my discussion, I am going to stipulate that there is at least a particularist strand in her work.

found in her chapters discussing Henry James's *The Golden Bowl*. As I read her she claims that James's novel gives us three valuable moral insights/benefits: (1) The novel alerts us to the fact that possessing knowledge of moral rules is not sufficient for moral decision-making; in some circumstances, this knowledge can lead us to act without regard to other persons as individuals. What is also necessary is attention to the concrete particulars, which attention is effected by a highly refined kind of *perception*. (2) The novel also shows us *which* perceptions to make and which not to make, such that we can become "finely aware" of the world around us. (3) Finally, the novel shows us what to make of these perceptions, how to interpret them, and how to implement them in our moral reasoning, such that we can become "richly responsible" moral agents. In the end we have to determine whether her particularist theory shows what it purports to show. I will argue that, because of its very particularist focus, it does not.

1. Rules and Moral Ideals

In the chapter titled "Flawed Crystals: James's *The Golden Bowl* and Literature as Moral Philosophy" Nussbaum recounts an incident in the novel in which the character Maggie tries to cling to a pre-conceived moral ideal of valuing individuals. When this ideal is carried to its logical conclusion, however, it results in treating individuals as objects—specifically, "*objets d'art*". Nussbaum claims that the novel shows us that Maggie, holding a strict moralism based on rule-possession and -following, "with its extreme emphasis on flawless living and, because of this, on consistency and harmony," espouses an ideal that "generates an extraordinary blindness

to value and ends by subordinating the particular claim of each commitment and love to the claims of harmony.”³ This ideal, Nussbaum argues, which has “very deep roots in the moral tradition of our culture” is, in the end, “childish.”⁴ It “commits the holder to a systematic neglect of certain features of persons—namely, both their separateness and their qualitative uniqueness—on which their specific personal virtue might be thought to rest.”⁵ The “neglect of [...] their separateness and [...] qualitative uniqueness” runs deeper than the phrasing suggests. It’s not *just* the separateness and uniqueness that Maggie fails to take into account, but the *particularities* of persons emerging from these general features. Maggie fails to note both *that* Charlotte is an individual and Charlotte’s *individuality* itself. Bringing in the particular details of Charlotte’s individuality represents for Maggie a threat to the “watertight” nature of her moral vision of the world. And this watertight nature of her vision is what leads her to treat individuals, not as persons, but as objects.

The ideal that Nussbaum claims has “very deep roots in the moral traditional of our culture,” emerges from rule-focused morality. We saw in chapter 1 that a commonly held conception of moral reasoning is that the agent approaches situations requiring a moral decision armed with a set of rules or one rule or one test for determining what to do in any situation bearing the same kind of features. This oversimplified characterization of most consequentialist and deontological theories is fairly accurate as far as it goes, as views of this kind depict moral reasoning as a matter of confronting the world, consulting our rule(s), and finally making our moral decision. Good moral reasoning would just be a matter then of the possession of the right rules and the ability to subsume the situation one confronts under the appropriate rule.

³ Ibid, p. 131.

⁴ Ibid, p. 132.

⁵ Ibid, pp. 132-133.

The moral particularist asks whether this is an apt description. We can be charitable to rule-based ethical theories and say on their behalf that, as with any theoretical enterprise, certain assumptions are made, *ceteris paribus* clauses are held implicitly, and even if, where the rubber hits the road, there is messiness, we can still see that it's conformity to the rules that makes an action right. It's the very assumption of watertightness that allows us to see what otherwise we wouldn't be able to see. This is not something to be lamented, as it's just what we need in order to be able to make the right decisions in the real world. Nussbaum's objection to this way of looking at moral reasoning is that Henry James's depiction of Maggie as someone who clings to this ideal allows us to see that this is exactly what we *don't* need and that it shouldn't be an ideal, attainable or not. When a rule-theorist says that it's an ideal that no one will actually reach, the objection still stands, since it's the ideal itself that leads Maggie astray.

Nussbaum uses a second metaphor—in addition to the metaphor of the watertight container—to push against rule-based theories: that of seeing straightness where there are only angles. Describing another of James's characters, Mrs. Newsome in *The Ambassadors*, as needing everything in the world to be “straight,” that is, as admitting of “no softening in the light of present circumstance” of the rules, Nussbaum generalizes to a type of individuals who “come to a situation determined that it should not touch them, holding their general and rather abstract principles fixed and firm.”⁶ But, of course, if the world is not composed of “straight lines” and the principles to which the world is supposed to conform don't bend, then there will never be a fit. What James shows us is the need for “ethical flexibility” such that perhaps the rules bend to fit the non-straight concrete particulars of the world and not the other way round. Perhaps this is why Nussbaum claims that “concrete perceptions [are] prior”⁷ to moral principles. Sticking with

⁶ Ibid, p. 177.

⁷ Ibid, p. 156.

this metaphor, Nussbaum concludes that people like Maggie and Mrs. Newsome, holding to the moral ideal of trying to use “some antecedent general principle held firm and inflexible for the occasion” are like “an architect who tries to use a straight ruler on the intricate curves of a fluted column.” This architect needs, not a straight and rigid ruler, but a “Lesbian rule” that bends to the shape of the structure to be measured. “Good deliberation,” Nussbaum writes, “like the Lesbian rule, accommodates itself to the shape that it finds, responsively and with respect for complexity.”⁸

At this point we have before us an insistence that the perception of concrete particulars be considered as part of the moral reasoning process, with the suggestion that this perception may possess “theoretical normative priority.”⁹ But we haven’t yet looked at what this perception is supposed to be. This is what we focus on in the next section.

2. Finely Aware

It’s clear that Nussbaum is arguing, in part, that James’s readers stand to acquire *moral-theoretical* information. We get a fuller description of what this information is in the famous chapter of *Love’s Knowledge* titled “‘Finely Aware and Richly Responsible’: Literature and the Moral Imagination.” There she restates the lesson learned from James: “Moral knowledge [...] is not simply intellectual grasp of propositions; it is not even simply intellectual grasp of particular

⁸ Ibid, p. 70.

⁹ See Roger Crisp, “Particularizing Particularism,” in *Moral Particularism*, ed. Brad Hooker & Margaret Olivia Little, NY: Oxford University Press, 2000, pp. 23-47, where Crisp makes the correct point that the priority isn’t temporal, but normative. The question then becomes “Are the true principles true to the extent that they summarize the particular perceptual judgements of virtuous agents, or are the particular judgements correct in so far as they conform to true general principles?” (p. 103)

facts; it is perception. It is seeing a complex, concrete reality in a highly lucid and richly responsive way; it is taking in what is there, with imagination and feeling.”¹⁰ Moral agents should seek to attain what she calls “perceptive equilibrium” defined as “an equilibrium in which concrete perceptions ‘hang beautifully together,’ both with one another and with the agent’s general principles; an equilibrium that is always ready to reconstitute itself in response to the new.”¹¹ The moral task is to become people “on whom nothing is lost.” But what exactly does “perception” mean in this context and what does it mean to be someone “on whom nothing is lost”?

The phrase “finely aware and richly responsible,” which appears in Nussbaum’s chapter title, originates in one of James’s prefaces, collected in the book *The Art of the Novel*:

This in fact I have ever found rather terribly the point—that the figures in any picture, the agents in any drama, are interesting only in proportion as they feel their respective situations; since the consciousness, on their part, of the complication exhibited forms for us their link of connexion with it. But there are degrees of feeling—the muffled, the faint, the just sufficient, the barely intelligent, as we may say; and the acute, the intense, the complete, in a word—the power to be finely aware and richly responsible. It is those moved in this latter fashion who ‘get most’ out of all that happens to them and who in so doing enable us, as readers of their record, as participators by a fond attention, also to get most.¹²

Note the ambiguous phrase “get most,” one interpretation of which is to *see* the entirety of the situation before one, given any situation. Of course, the word “see” is used metaphorically here, as it is in Murdoch’s account discussed in chapter 1, but it is nonetheless closely tied to, among the other senses, the sensory activity of sight, which implies the utilization of perceptual, but also

¹⁰ *Love’s Knowledge*, p. 152.

¹¹ *Ibid*, p. 183.

¹² Chicago: The University of Chicago Press, 2011, p. 62.

cognitive, resources. Nussbaum uses the term “perception” to refer to “the ability to discern, acutely and responsively, the salient features of one’s particular situation.”¹³ The word “salient” here is vague, since it could refer to any number of contexts, but this is not necessarily an oversight, since her intention here is to get us to see the importance of context in moral reasoning. What is salient in one context won’t necessarily be so in another, and what is salient in moral contexts might not be so in non-moral ones.

Being what Nussbaum calls a “Jamesian perceiver” involves both the ability to see what is there—that is, be open to the details in front of one’s eyes, so to speak—and the ability to see what is there *as* a particular situation. When one looks at a widow, dressed in black, weeping, hunched over a coffin, one must not only see the black dress, the tears, and posture, but the woman *as* a grieving widow. Sometimes this is easier said than done, since doing so requires a keen sense of imagination. And in moral contexts, the need for imagination is all the more significant. Moral imagination, Nussbaum writes, “is subtle and high rather than simple and coarse; precise rather than gross; richly colored rather than monochromatic; exuberant rather than reluctant; generous rather than stingy; suffused with loving emotion rather than mired in depression.”¹⁴ This description implies that there is a *right* way of perceiving.

One of the ways that James’s book benefits its readers, according to Nussbaum, is its showing the reader the right way to perceive. And if the reader is able, she too can engage in the right way of perceiving the details of the scenes in the novel. James suggests that this occurs when we are “participators by a fond attention” and Nussbaum argues that “our own attention to [James’s] characters will itself, if we read well, be a high case of moral attention.”¹⁵ The “ideal” reader will be someone keenly alive in thought and feeling to every nuance of the situation,

¹³ *Love’s Knowledge*, p. 37.

¹⁴ *Ibid*, p. 152.

¹⁵ *Ibid*, p. 162.

actively seeing and caring for all the parties concerned—and therefore safely right in the perfection of his or her attention.”¹⁶ This is because “this novel calls upon and also develops our ability to confront mystery with the cognitive engagement of both thought and feeling.” “To work through these sentences,” Nussbaum says, “and these chapters is to become involved in an activity of exploration and unraveling that uses abilities, especially abilities of emotion and imagination, rarely tapped by philosophical texts. But these abilities have, at the very least, a good claim to be regarded as important parts of the moral assessment process.”¹⁷ Therefore, in addition to the moral-theoretic information we can learn from James’s novels, we can also learn the *skill* of perceiving in the right way, which is why it’s most appropriate to characterize this account as hybrid.

The importance of perception for Nussbaum lies in three distinct points:

1. Without being able to perceive the concrete particulars of a situation, one wouldn’t know which rules and standing commitments are operative. In other words, one cannot approach a moral decision already armed with the relevant moral principles and relevant commitments.
2. Even if one has the right moral principles in hand, if one isn’t highly attuned to the particularities of the case, one might not act rightly with these principles in hand. Having the right moral principles isn’t sufficient for acting rightly.
3. There are always potentially new and surprising elements in a concrete situation that requires one to use the standing moral principles in different ways. Without being attuned to the particularities of the situation one would misapply moral principles when

¹⁶ Ibid, p. 143.

¹⁷ Ibid, p. 143.

confronted with the new elements. “The fine Jamesian perceiver” says Nussbaum, “employs general terms and conceptions in an open-ended, evolving way, prepared to see and respond to any new feature that the scene brings forward.”¹⁸

Nussbaum’s main claim about being finely aware is that James’s novel shows us (a) the right kind of perception, mediated through the narrator’s attention or through the perspective of a particular focalized character, and (b) also how to *be* the right kind of perceiver. He does the latter, she argues, by often requiring it of us as readers. We will, to be sure, not reach this ideal on the first read, or perhaps even subsequent reads, but within the novel the ideal sits in wait, and this ideal, rather than those held by Maggie and Mrs. Newsome, is one Nussbaum clearly thinks is worth pursuing.

Being the right kind of perceiver is itself a moral achievement, according to James and Nussbaum. Similarly, being the wrong kind of perceiver is a moral failure. By extension, if an artist, in engaging the audience member in the attention illustrated in the work, evokes the wrong kind of perception in the audience member, then the artist is morally culpable. “The artist’s task *is* a moral task,” Nussbaum writes. “The whole moral content of the work expresses the artist’s sense of life; and for the excellence of this the novelist is, in James’s view, rightly held (morally) responsible.”¹⁹ But then there seems to be more weight placed on the element of *seeing as* rather than *seeing of*. In other words, the *salience* to which Nussbaum alludes is key to the moral vision on offer here. In order to understand this, let’s consider a case in which the artist may fail to perform the moral task to which Nussbaum refers.

¹⁸ Ibid, p. 157.

¹⁹ Ibid, p. 163.

3. On Being Finely Aware of Lolita's Charms

Being finely aware is not sufficient for moral decision-making, which is of course going to lead us into the second half of James's quoted phrase—being richly responsible. But it will help to see why fine awareness is insufficient by considering a literary work that is at least as drenched in high attention as *The Golden Bowl*, but focused primarily on details that are not morally salient, or if they are salient, they would serve as reasons *not* to perform the actions depicted in the work. Consider Vladimir Nabakov's *Lolita*.

Nussbaum and James hold the line that an artist is morally culpable for artistic decisions—he should be praised for improving audiences and blamed for corrupting them. This claim can be tested in the example of *Lolita*. The novel has been condemned because of its pedophilic narrator Humbert Humbert, usually because the narrator's proclivities are purported to “rub off” on readers. If this is the case, we can ask why it occurs and whether Nabakov himself should be condemned for it.

First, we can reject a way commonly used to determine whether audiences are corrupted by *Lolita*. Sometimes people produce examples of pedophiles who, on getting caught, point to their reading the book as the reason they began to sexually desire children or to act on those desires. But producing such examples alone won't settle the matter. We have to ask whether these people had read the novel in the intended way. Of course, Nabakov can be held accountable for unintended consequences, but the burden is on the critic to prove that the unintended consequences were due to Nabakov's error and not the pedophilic reader's. In most cases these readers had not paid due attention to the subtleties of the novel, which can serve to mitigate some of the ham-fisted criticisms of the novel.

Even though we can reject this first way of determining whether Nabakov is morally at fault for some aspect of the effects on readers of *Lolita*, we can still attend to these subtleties and ask whether he has committed some wrong even given the finer details. It's been suggested that pedophilia is prescribed in the novel, even if the prescription isn't found at the surface level beyond which the pedophilic readers never explore. The narrator Humbert pines for Lolita, and provides lucid and engaging descriptions of her as he indulges in machinations intended to facilitate the satisfying of his immoral desires. Even careful readers have detected indications that Humbert is to be thought of as a hero with whom we are supposed to commiserate. If we are being persuaded to pull for Humbert and to hope that he achieves his goals of successfully seducing Lolita, there is a case to be made that the novel is immoral.

But it isn't as straightforward as that. Berys Gaut,²⁰ for example, argues that the narrator with whom we commiserate is in fact the Humbert who is in prison, waiting to be executed, and is for the most part penitent, while the figure that molests Lolita is Humbert years ago, prior to being imprisoned. It's the latter Humbert who is the moral monster, but not the former Humbert, who is trying to atone for his earlier sins. Even if there are passages in which Humbert seems to be describing his actions in ways suggestive of approval of what is being done, it's the later, penitent Humbert *remembering* the thoughts and emotions of earlier, shameless Humbert, a subtlety lost on the careless reader. As this is the case, Gaut suggests, the novel doesn't prescribe active pedophilia.

If Gaut is correct, then readers who think that *Lolita* does prescribe pedophilia have misinterpreted the attitude of the actual narrator—Humbert who relates the account from the prison cell. A different way to find the immoral prescription nonetheless, however, is to

²⁰ *Art, Emotion, and Ethics*, NY: Oxford University Press, 2007.

determine whether the author—either the actual or implied author²¹—commends pedophilia, even if the actual narrator doesn't. Indeed, Brandon Centerwall²² argues that the work doesn't condemn pedophilia and is in fact an expression of Nabokov's closeted pedophilia. The attitude of the work, which is commendatory of pedophilia, would be discoverable only after extensive and nuanced interpretation, again indicating that the careless pedophilic reader lacks justification, even if the novel does in fact prescribe pedophilia.

James Phelan²³ argues that there are two figures: Humbert as character and Humbert as narrator, and that the narrator is writing years after the incidents being recounted, but that it's not a case of penitent Humbert sitting down to atone for his sins. The work is actually being focalized through *both* Humbert the character and Humbert the narrator. In discussing the events just after Humbert first has sex with Lolita, and, at the end of chapter 29, he claims that he will not dwell on the details of the sexual act, but at the beginning of chapter 30, he luxuriates in such a description nonetheless.²⁴ Phelan says the following:

Nabokov uses this sequence and the dual focalization to add a significant layer to the narrative: the ethical struggle of Humbert the narrator. The struggle, at the most general level, is about whether he will continue to justify and exonerate himself or shift to admitting his guilt and accepting his punishment.²⁵

²¹ See Wayne Booth, *The Rhetoric of Fiction*, Chicago: University of Chicago Press, 1983. The distinction between the actual and the implied author is that the actual author is the human being who wrote the book, while the implied author is the figure constructed by the reader as a result of gathering clues in the text.

²² "Hiding in Plain Sight: Nabokov and Pedophilia," *Texas Studies in Literature and Language* Vol. 32, No. 3, pp. 468-484.

²³ "Dual Focalization, Retrospective Fictional Autobiography, and the Ethics of *Lolita*," in *Narrative and Consciousness: Literature, Psychology and the Brain*, ed. Gary D. Fireman, Ted E. McVay, Owen J. Flanagan, NY: Oxford University Press, 2003.

²⁴ See Text 1 for this description.

²⁵ "Dual Focalization, Retrospective Fictional Autobiography, and the Ethics of *Lolita*," p. 141.

On Phelan's reading, the issue isn't as clearly resolved as Gaut claims. It's not as if Humbert, filled with contrition, sits down to write a confessional in order to seek forgiveness for his sins, but it's also not a matter of Humbert's seeking *to pose* as an atoner, when he is really just trying to garner sympathy. If the latter were the case, then Centerwall's interpretation might gain some traction, since Nabakov too could be posing as apologist for Humbert the character's ghastly actions. Phelan believes, instead, that in scenes like the ones at the end of chapter 29 and beginning of chapter 30 we are seeing "Humbert's *shift* to condemning himself" which is not something pre-planned at all. It turns out to be "an unplanned shift that develops as he writes."²⁶

All of this is preliminary to looking at the case of *Lolita* as one in which we might have the right kind of perception, but perfect examples of non-virtuous action. One way to avoid even beginning this exploration is to argue that the perception being demonstrated is not the right kind of perception at all, and that showing us this is one of the intentions found in the work.²⁷ It's doubtful, however, that this is the case, since Humbert appears to be engaging in what Nussbaum would call fine perception.

Consider Humbert's description of Charlotte, *Lolita's* mother, who provides him no pleasure, but an opportunity nonetheless to show that at least nothing *physical* is lost upon him:

I think I had better describe her right away, to get it over with. The poor lady was in her middle thirties, she had a shiny forehead, plucked eyebrows and quite simple but not unattractive features of a type that may be defined as a weak solution of Marlene Dietrich. [...] Her very wide-set sea-green eyes had a funny way of traveling all over you, carefully avoiding your own eyes. Her smile was but a quizzical jerk

²⁶ Ibid, p. 141. my emphasis.

²⁷ If this is the case, then we have the flip side to Henry James's novels, which Nussbaum claims show us how to perceive. *Lolita* would be a demonstration of how *not* to perceive.

of one eyebrow; and uncoiling herself from the sofa as she talked, she kept making spasmodic dashes at three ashtrays and the near fender [...] whereupon she would sink back again, one leg folded under her.²⁸

Humbert doesn't just see Charlotte, but he sees her *as* a hopeful widow who, worried about her best days being behind her, has set her sights on this handsome and winsome potential lodger as a potential suitor. And this is, of course, exactly what Charlotte turns out to be. Moreover, he doesn't have to say that this is an apt characterization of her, because we as readers, seeing Charlotte as he sees her, recognize that it is apt. We too see her as the widow rapidly approaching her sell-by date. We see Charlotte with Humbert, and we might even feel complicitous in his evaluative description.

This feeling of complicity becomes troublesome, however, when we encounter Lolita and witness Humbert's initial description of her, reminiscent of his childhood love, Annabel Leigh:

[...] the same frail, honey-hued shoulders, the same silky supple bare back, the same chestnut head of hair. A polka-dotted black kerchief tied around her chest hid from my aging ape eyes, but not from the gaze of young memory, the juvenile breasts I had fondled one immortal day. And, as if I were the fairy-tale nurse of some little princess [...] I recognized the tiny dark-brown mole on her side. With awe and delight [...] I saw again her lovely indrawn abdomen where my southbound mouth had briefly paused; and those puerile hips on which I had kissed the crenulated imprint left by the band of her shorts [...] The twenty-five years I had lived since then, tapered to a palpitating point, and vanished.²⁹

One might take this description as a reminiscence of an adult man of his younger days when he enjoyed time with his young sweetheart. But the main object of focus here is undoubtedly the body of Lolita, and only secondarily its resemblance to the body of Annabel. One might also take

²⁸ Vladimir Nabakov, *Lolita*, NY: Vintage, 1991, p. 37.

²⁹ *Ibid.*, p. 39.

it as an inappropriately sexual description of a female whose age has been overlooked, but when we replace the ellipses at the beginning of the passage with the original phrase—“It was the same child ...”—it’s clear that Humbert doesn’t just see a sexual being, whom he fails to recognize as too young for such descriptions, but from the beginning, a *child* as a sexual being. It’s not that he has momentarily forgotten that the abdomen on which he imagines placing his mouth and the hips which he envisions kissing belong to a child. He recognizes she is a child all along. Can it be that he fails to realize that it’s inappropriate—morally wrong—to see Lolita both as a child and as a sexual object? But what else could he see that would forestall his sexualizing her?

The mystery deepens as we witness Humbert’s plans to “seduce” Lolita develop and fizzle, even as he renders the scenes along the way in wondrous and brilliant detail. In a real sense the reader is seduced as well. We finally get to the scene in which Humbert looks finally poised to achieve his goal, but only with the aid of sedatives. He describes the scene thus:

When the dessert was plunked down—a huge wedge of cherry pie for the young lady and vanilla ice cream for her protector, most of which she expeditiously added to her pie—I produced a small vial containing Papa’s Purple Pills. As I look back at those seasick murals, at that strange and monstrous moment, I can only explain my behavior then by the mechanism of that dream vacuum wherein revolves a deranged mind; but at the time, it all seemed quite simple and inevitable to me. I glanced around, satisfied myself that the last diner had left, removed the stopper, and with the utmost deliberation tipped the philter into my palm. I had carefully rehearsed before a mirror the gesture of clapping my empty hand to my open mouth and swallowing a (fictitious) pill. As I expected, she pounced upon the vial with its plump, beautifully colored capsules loaded with Beauty’s Sleep.

“Blue!” she exclaimed. “Violet blue. What are they made of?”

“Summer skies,” I said, “and plums and figs, and the grape-blood of emperors.”

“No, seriously—please.”

“Oh, just Purpills. Vitamin X. Makes one strong as an ox or an ax. Want to try one?”

Lolita stretched out her hand, nodding vigorously.³⁰

The plan is to drug her, take her back to the hotel room, visit the lobby long enough for the pills to take effect, return to the room, and have his way with her. In the above description, we see the fine details, again we see Lolita, when presented with the pills, display her pre-adolescent innocence and excitement. We see her naivety, and if we're honest, we feel Humbert's thrill when Lolita stretches out her hand and nods vigorously. Unfortunately for Humbert, however, he returns to the room later to find that the drugs have not had the effect he had hoped—the doctor had not prescribed what Humbert had asked for. Thus, while heavily sedated, she is not completely unconscious. Humbert therefore chooses not to implement the rest of his plan. Why not?

It's here that we might place some stock in Humbert the narrator's claims about Humbert the character. In his defense he says, "I am not, and never was, and never could have been, a brutal scoundrel. The gentle and dreamy regions through which I crept were the patrimonies of poets—not crime's prowling ground."³¹ The insinuation here is that, even if he was prepared to take liberties with an unconscious Lolita, he was not going to do so with her conscious, even in a twilight state. We can take Humbert the narrator at his word, since in fact Humbert did hold off completing his plan. We can thus attribute some moral sensibility to Humbert the character. But there are multiple reasons for why he might have done this. Imagine that Humbert the narrator had said, not that Humbert the character had halted his plan because he is not a "brutal scoundrel," but that he "is not, and never was, and never could have been, someone accused of common rape"? This alternate reason would not be a moral one, even if the original moral reason

³⁰ Ibid, pp. 121-122.

³¹ Ibid, p. 131.

was incomplete; the reason would be based on Humbert's pride. We might even imagine a different novel, one with Humbert the character performing the same actions and observations as in *Lolita*, but with a very different Humbert the narrator. Imagine this different novel—let's call it *Juanita*—contained a narrator—Hubert, let's say—who wasn't pleading a case before a jury and trying to evoke sympathy, but was describing his crimes to a cellmate, enjoying the remembered pleasure and hoping to produce pleasure in his audience, as he constructs his fine descriptions of what he had done. We can see that this novel, absent any other factors that might give us reason for con-attitudes toward the actions of Hubert, would in fact be morally corrupt. And if readers were to take up the pro-attitude toward pedophilia, a strong case could be made for the culpability of the author. But notice that, *ex hypothesi*, the perceptions of Hubert the character would be the same as those of Humbert the character.³²

But we don't even need to imagine the alternate novel. We shouldn't be so quick to let *Lolita* off the hook here. The novel does make use of fine perceptions, as we have seen that Humbert recognizes Lolita as a child, and he still performs vicious actions. Humbert the narrator says that he was not a brutal scoundrel, since he was not prepared to perform his actions on a semi-conscious Lolita. He says "Had I reached my goal, my ecstasy would have been all softness, a case of internal combustion of which she would hardly have felt the heat, even if she were wide awake."³³ Even Humbert the narrator doesn't see that the primary wrongness of what he planned wasn't in the pain it would have caused Lolita, but in the way he would have treated her, a way to which she wouldn't have, or couldn't have—because of her age—consented. There is then a morally pernicious attitude there already. As Phelan says, one struggle Humbert the

³² This works only if we don't hold to Phelan's interpretation that some of the perceptions of the scenes belong to Humbert the narrator. If this is the case, then it's unlikely to find the same perceptions of the scenes in *Lolita* and *Juanita*, since the narrator's observations are blended into these scenes.

³³ *Lolita*, p. 131.

narrator is undergoing is about “vision”. Can he see himself as a man who “was aware of Dolores’s pain at the time of the action but refused to attend to it”?³⁴ And if so, does this realization exonerate him?

I am not trying to show that *Lolita* is an immoral work. I think the novel is too complicated and rich to make such a case here. What I am considering is whether the kind of perception Humbert the character implements is the kind Nussbaum and James prescribe. And if so, whether Humbert the character, who is at least *one* immoral figure in the work, is a counterexample. John McDowell articulates the worry I am considering in the following:

That may seem to open the following possibility: a person’s perception of a situation may precisely match what a virtuous person’s perception of it would be, although he does not act as the virtuous person would. But if a perception which corresponds to the virtuous person’s does not call forth a virtuous action from this non-virtuous person, then the virtuous person’s matching perception—the deliverance of his sensitivity—cannot, after all, fully account for the virtuous action which it does elicit from him. Whatever is missing, in the case of the person who does not act virtuously, must be present as an extra component, over and above the deliverance of the sensitivity, in a complete specification of the reason why the virtuous person acts as he does. That destroys the identification of virtue with the sensitivity.³⁵

If being finely aware is the sensitivity McDowell is talking about, is being “richly responsible” that extra component?

³⁴ Ibid, p. 141.

³⁵ “Virtue and Reason,” in *Virtue Ethics*, ed. Roger Crisp & Michael Slote, NY: Oxford University Press, 1997, pp. 144-145.

4. Richly Responsible

With our reading of *Lolita* we have an example in which being finely aware appears to be insufficient for virtue. Merely seeing fine details is not enough, as Humbert has the capacity to see these details and commits vile actions nonetheless. For this reason it was suggested by the quote from McDowell that something besides perception is necessary. But does Humbert really have the power of Jamesian perception as described by Nussbaum? Return to the scene in which Humbert feigns the taking of the pill and then patiently awaits Lolita's eager entreaty to give her one. He knows she will do so. Why? Part of the reason is that by now he knows her quite well and is aware of her personality, but a significant part of the reason is that he knows that she, because she is a child, has childlike curiosity and gullibility, both of which are typical of children and why societies have decided to make the age of consent older than twelve. For Humbert the gullibility serves as a crucial component to his plan. While he's trying to pull off his scheme, he never questions his exploiting of this characteristic. It's clear that he never believes his attempt is morally justified, but this determination is made at a coarse level. It simply never occurs to him that he is committing a moral wrong by trying to take advantage of her gullibility—which means that he never *recognizes* that her gullibility and his trying to exploit it are morally significant features of the situation. But perhaps this is also part of the reason why what looks like fine awareness never leads Humbert to decide not to pursue his plan to drug and take advantage of Lolita. McDowell would think a case like this “destroys the identification of virtue with the sensitivity,” but what this probably shows us instead is that Jamesian perception has a recognition-component that Humbert doesn't possess. It's not just seeing the details, but recognizing that they are—seeing them *as*—morally significant. Being finely aware then appears

to *include* being richly responsible. Or, to put it a different way, perhaps becoming finely aware, under the Jamesian conception, seamlessly leads a rational person to be a richly responsible moral agent. Our attempt to break them apart has treated them as temporally distinct, but perhaps they come in a package. Being a Jamesian perceiver is being a richly responsible agent. Ergo, Humbert is not a Jamesian perceiver and what he engages in throughout the book is not fine Jamesian perception.

Jamesian perception then turns out to be what was referred to in chapter 1 as *weak moral perception*, which occurs when an agent can see a moral situation as a moral situation. and not *strong moral perception*, which is when an agent *just sees* what action to perform. It's not strong moral perception because Nussbaum holds that, although the concrete situation is theoretically prior to the abstract moral principle, the principle has a role to play still. It's weak moral perception because perception is used to recognize the morally salient detail *as* morally salient.

Once we have made this clarification to Nussbaum's account, we can see that if Nussbaum is correct, her account, more than the others discussed in chapter 5, holds great promise in solving the moral problem described in Part I. The ultimate changes prescribed in Part I were the actualizations of accurate axiological and weak moral perception, but even axiological perception was thought to be preparatory to moral perception. If reading novels like *The Golden Bowl* improve our moral perception, then we don't even have to address the need to improve axiological perception. Moral edification happens in one step—from the reading of the novel to being a better perceiver.

But we have to look more closely at *how* readers are supposed to become better perceivers, on Nussbaum's particularist account. If the moral payoff of reading high novels like *The Golden Bowl* is, in part, becoming better perceivers through perceiving with James as he

makes his perceptions, the process is one of *modeling*. James models for us the right kind of perception and, through our engagement with the narrative by virtue of our sharing in his perception, we engage in the same perceptions, thus making us better perceivers.

Learning through modeling takes time, however, and often proceeds through trial and error. If I am modeling the critical reading of a philosophy text for my students, initially some of them may fix on the wrong features. Suppose one of them notices that the first few times I isolated a passage in a text and critically evaluated it, the passage was less clearly written than the passages surrounding it. This student may infer, incorrectly, that we critically evaluate a passage based on lack of clarity, rather than issues having to do with content. It takes a few observed instances for this student to notice that later passages are not as unclear as the earlier instances and eventually realizes that what looked like a conscious decision on my part was actually just a coincidence.

Because modeling takes time and usually progresses through trial and error, if readers of these great novels are supposed to learn through modeling, it's not clear how this can happen in isolated experiences. Recall that Nussbaum's account, at least on one reading, is particularist and the claims she makes are about the power of individual novels to provide moral education. It's hard to see how these individual novels—even when read multiple times—can have the effects she claims without taking into account how readers blend the experience of reading these novels with experiences in life. In short, learning through modeling requires trial and error, which itself requires the passage of time and the intersecting of fictional and real life experiences, but Nussbaum's account forestalls these possibilities.

Nussbaum would probably respond by saying that this is precisely why she chooses “high” novels as her prime examples, rather than merely good literature. They are great, she

might say, because they are rich enough to provide a reader enough opportunity to model the author's perception in one work or one experience. But in most cases, if the author is consistent within a work, the modeling will appear to the reader to be more of the same, since the perspective on the situations arising in the narrative, and the kinds of perceptions made, will remain the same. If the perspective changes, we might have two models of perception and then the reader would have to choose which model to adopt.

In response Nussbaum might say that she never intended to claim that a naïve reader might sit down and read *The Golden Bowl* and be able to gain the kind of learning she describes. First, it takes a fairly sophisticated and dogged reader even to be able to read such a novel. Therefore, the appropriate reader comes armed already with a kind of practical judgment that would lead her to make decisions about what aspects of the author's perceptions are the ones to adopt in one's own reasoning. Second, there is a *class* of "high" novels that she recommends for such learning, in addition to *The Golden Bowl*, such as Proust's *Recherche a la Temps Perdu* and other James novels like *The Ambassadors*. There is no reason, she might say, to think that an individual novel, read by just anyone, has the effects she describes.

If she provides this kind of response, then she would seem to be drifting slowly away from some of the strong claims made in *Love's Knowledge*, and the interpretation of hers as a particularist theory would no longer be tenable. It would be very difficult, however, for her to contend, on the one hand, that a novel like *The Golden Bowl* has the tremendous power to effect moral transformations and, on the other hand, to concede that readers might have to read other similar novels in order to be transformed in the way she describes. At this point she would be a dithering particularist, searching about for a way not to be one any longer.

All of these challenges would go away, however, if she were to stop dithering and re-focus her account. In part III, when we expand our scope of investigation, no longer limit ourselves to individual artworks and experiences, rejecting the particularist focus and re-casting the central question as whether living a certain kind of life committed to aesthetic experience can transform us in a moral way, we will see connections and relations that are obscured by the narrow focus possessed by the theories considered in the last and present chapter.

Chapter 30 (*Lolita*)

I have to tread carefully. I have to speak in a whisper. Oh you, veteran crime reporter, you grave old usher, you once popular policeman, now in solitary confinement after gracing that school crossing for years, you wretched emeritus read to by a boy! It would never do, would it, to have you fellows fall madly in love with my Lolita! Had I been a painter, had the management of The Enchanted Hunters lost its mind one summer day and commissioned me to redecorate their dining room with murals of my own making, this is what I might have thought up, let me list some fragments:

There would have been a lake. There would have been an arbor in flame-flower. There would have been nature studies--a tiger pursuing a bird of paradise, a choking snake sheathing whole the flayed trunk of a shoat. There would have been a sultan, his face expressing great agony (belied, as it were, by his molding caress), helping a callypygean slave child to climb a column of onyx. There would have been those luminous globules of gonadal glow that travel up the opalescent sides of juke boxes. There would have been all kinds of camp activities on the part of the intermediate group, Canoeing, Coranting, Combing Curls in the lakeside sun. There would have been poplars, apples, a suburban Sunday. There would have been a fire opal dissolving within a ripple-ringed pool, a last throb, a last dab of color, stinging red, smearing pink, a sigh, a wincing child.

pp. 134-135

Text 1

Part III

Aesthetic Maturity

and the Life of the Reader

Chapter 7

Aesthetic Maturity and Narrative Artworks

I described in Part I a serious moral problem: currently there is a widespread phenomenon of diminished caring on the parts of moral agents for those around them, in part because they are naturally poor at identifying and properly weighting the interests of others. I suggested in chapter 1 that art is the solution to this problem, but in Part II I argued that accounts that purport to show how art produces ethical effects—and specifically how art is the solution to the problem laid out in Part I—fail when the focus is narrowed to individual artworks or aesthetic experiences. In Part III I will construct an account without such a narrow focus that is intended to show two things: (1) that living a life in which aesthetic experience is taken seriously can produce improvement in the four areas presented at the end of chapter 3, which are

1. decentering
2. metaknowledge regarding our empathic accuracy
3. motivation to mindread
4. the ability to mindread (heuristic mindreading, deliberative mindreading, and mindwriting);

and (2) that living this kind of life can result in audience members feeling more concern for those around them—i.e., an expansion in the capacity to care. Inasmuch as the right kind of motive for moral agents to have is empathetic concern—as I argued in chapter 2—aesthetic experience does result in a genuine ethical, rather than merely pre-ethical, effect.

1. What Kind of Art?

Since we are after the improvement of empathic accuracy and interpersonal relations, the appropriate moral relations to consider will be those between actual *agents*. What kind of art deals primarily with this kind of relation? Narrative art. The reason narrative art is the most appropriate is that it displays in the fullest way interactions between moral agents, a central requirement of learning anything about how moral agents should behave toward other moral agents. Moreover, narratives are inherently temporal, which allows us to see relationships develop over time—sometimes long periods, sometimes short and focused periods. But our attention is best rewarded within the context of interpersonal understanding when we are allowed to see these unfolding relationships under a relatively tight control. Because my focus will be on narrative art, I will not be discussing static artforms, like sculpture, architecture, and most paintings. Nor will I be discussing forms that don't allow for character development, such as many forms of poetry and absolute music.¹ My focus will be trained on novels, short stories, film, some poetry, opera, and drama.

What about biographies, documentaries, well-told non-fictional accounts? As Lamarque² argues, it's a myth that all narratives are "*au fond* fictional." So should we include non-fictional narratives? An obvious way to answer this question in the negative is to say that non-fictional narratives are not works of art, and, since we are engaged in a discussion of *art's* relation to morality, we can reject them out of hand. This answer won't work, however, since many historical novels (e.g., James Michener novels) and documentaries (e.g., the infamous *Triumph of the Will*) are commonly accepted as works of art, though the (intended) nature of the narrative

¹ Of course, some agency might be inferred of absolute music, but this occurs only at the interpretative level.

² "On Not Expecting Too Much from Narrative," *Mind & Language* Vol. 19, No. 4, 2004, pp. 393-408.

is non-fictional. So, if we want to limit the discussion to fictional narratives, we must have a reason other than that non-fictional narratives aren't art. There would have to be something about the very *fictionality* of fictional narratives that justifies our decision.

2. What Kind of Narrative?

There is a reason to limit this discussion to fictional narratives: the *virtuality* of fictional narrative space. I will have more to say about this aspect of fictional narratives later, but I will give a brief description here. Narrative artworks are especially interesting to look at because they give us a *semblance* of the real world, but, more than that, since their authors are able to give extended *histories*³ while retaining a high degree of control, conditions can be adjusted in ways unlike anything possible in the real world. As we saw in Currie's informational account in chapter 5, we can look at fictional narratives as a kind of thought-experiment, performed under highly idealized conditions.

Of course, this idealization is not specific to fictional narratives. We know that the descriptions of the streets of Boston and New York City in *The Autobiography of Malcolm X* are stylized and not intended to be absolutely accurate. But in non-fictional narratives, we tend to *excuse* this kind of stylization, whereas in fictional narratives, we expect and often luxuriate in it. Dorrit Cohn says that what distinguishes fictional from non-fictional narrative is that "(1) its references to the world outside the text are not bound to accuracy; and (2) it does not refer

³ I take this conception of narrative from Ira Newman, "Narrativizing Shame: A Study in Moral Psychology and Literary Structure," American Society of Aesthetics Annual Meeting, 2011.

exclusively to the real world outside the text.”⁴ Cohn’s view differs from those of Diffey and Beardsley discussed in chapter 4, since she thinks that fictional works *can* refer to the real world, but needn’t do so. Yet Cohn’s (1) might return us exactly to the point at which Diffey had left us. If fictional works aren’t bound by standards of accuracy, how do we know whether the text refers to the real world if we don’t have this information already? Diffey’s objection isn’t pertinent here, however, since Cohn’s (1) is part of why we approach fictional narratives as virtual space. While we are reading and processing the narrative, we aren’t tasked primarily with determining how closely the text reflects the real world, so we have some freedom to allow interpretation to run on further than we would if the text was intended to refer to the real world. But there are *some* standards of interpretation.

First, there are *internal standards*. If the elements of the text of the narrative just render some interpretations impossible, these interpretations won’t work. Critics try to avoid constructing interpretations that, while consistent with the details of the narrative, have no strong textual support.

Second, there are *external standards*, like the time period of construction. For example, interpreting *Plan 9 from Outer Space* as avant garde filmmaking doesn’t work, as Noël Carroll⁵ makes clear. This film cannot be avant garde, because it was made prior to the conceptual framework that makes avant garde filmmaking possible. Unfortunately, Ed Wood’s movie is not groundbreaking, it’s just bad.

Finally, there are *reference standards* of interpretation, but the relevant reference relation isn’t to the real world. Narrative theory has discovered an underlying feature of the structure of narrative: the distinction between the representation (the narrative) and the represented (the

⁴ *The Distinction of Fiction*, Baltimore, MD: Johns Hopkins University Press, 1999, p. 15, emphasis in original.

⁵ “Art, Intention, and Conversation,” in *Intention and Interpretation*, ed. Gary Iseminger, Philadelphia: Temple University Press, 1992.

world of the narrative). This distinction probably has its roots in Aristotle, but gained most of its traction in the early twentieth century with the Russian formalist terms “*fabula*” and “*syuzhet*.” Theorists have provided their own terms since then, but the most commonly accepted terms come from Seymour Chatman: *story* and *discourse*.⁶

Discourse is the narrative itself, which can take on any of a number of structural possibilities. It can run from the past to the present; from the present to the past; begin in the future, move to the past, then back to the present; etc.. *Story* is the chain of events themselves that are being represented in the discourse. This chain runs as all chains do in the real world—from the past to the present to the future. While the discourse will underdetermine the story, the story still has a determinate nature, even if not fully known by the reader of the discourse.⁷

In fictional narratives, the story doesn’t (modally) actually exist, in the same way the real world exists alongside non-fictional narratives. This is what leads Beardsley to argue that fictional works don’t refer at all. Discourses do refer, however, but only to a non-actual world called the story, which, *qua* theoretical entity, exists. While we are engaged with the narrative, our believing or make-believing⁸ that the story exists doesn’t really commit us to anything extravagant.

Therefore, Diffey’s criticism that we must antecedently know that the work’s propositions refer to the real world isn’t apt here. Everything turns on how we approach the narrative and what we hope to accomplish by doing so. If we approach it as we would a thought-experiment, as Currie advocates, then we would have already recognized the constraints that

⁶ *Story and Discourse: Narrative Structure in Fiction and Film*, Ithaca, NY: Cornell University Press, 1978.

⁷ We might have to allow for some novels, like Martin Amis’s *Time’s Arrow*, that depict a storyworld that really does run from the future to the present to the past. If so, then we would need to add the disclaimer that *traditional* narratives depict a storyworld with temporal laws like our own.

⁸ Kendall Walton, *Mimesis as Make Believe: On the Foundations of the Representational Arts*, Cambridge, MA: Harvard University Press, 1990.

Diffey highlights. Currie describes the situation in which we encounter a fully formed thought-experiment, with the conditions set beforehand, and we place our own conscious perspectives into the situation and allow the conditions to work upon us. But we could also think of narrative space as offering us the opportunity to set up the thought-experiment ourselves, as we make interpretations. Marisa Bortolussi & Peter Dixon argue that there are two elements that must be distinguished in order to understand how interpretation occurs: textual features and reader constructions. A textual feature is “anything in the text that can be objectively identified,” including “narration styles [...], speech styles [...], techniques for marking discourse and story time, and aspects of characterization.” Reader constructions are subjective, “events and representations in the minds of readers,” such as “changes in readers’ attitudes or beliefs, and affective reactions.”⁹ Interpretations are often a combination of textual features and reader constructions.

The kind of thought-experiment I envision, made possible by the virtuality I have been describing, would be primarily a reader construction, while Currie’s vision would be primarily thought-experiment as textual feature. Of course, textual features will act as constraints, under my description, but every instance of the kind of reading I am talking about will be a different reader construction. The justification for any insight learned from the thought-experiment will be two-staged: (a) we need a justification for the claim that the interpretation is consistent with the discourse (external and internal standards) and story (reference standards); and (b) we need a justification for the claim that the storyworld is similar enough to the real world to justify any inferences made beyond the work itself. In most cases, people devote their interpretative efforts to justifications belonging to (a), and arguably the possibilities are more open in (a), since this is

⁹ *Psychonarratology: Foundations for the Empirical Study of Literary Response*, Cambridge: Cambridge University Press, 2003, p. 28.

where the virtuality aspect comes to the fore. But it must also be recognized that for our purposes—tying insights gained from artistic experience to our extra-fictional moral and cognitive lives—we would also need to attend to (b).

For the reasons I have given, the kind of artworks we will consider in relation to the question of how to solve our moral problem will be fictional, narrative artworks.

3. The Aesthetic Adult

In addition to the kind of artwork we will consider, I should also give a description of the kind of audience member I have in mind. In the introduction I quoted Ronald Moore, who claims that there is a significant relationship between moral and aesthetic adulthood. He describes being an aesthetic adult as being committed to “a course of valuation (which may entail appreciation, preservation, protection, encouragement, among other involvements) regarding both artifactual and natural objects,” which involves having “worked one’s way through the various levels of objective and subjective valuation that one’s aesthetic culture presents to the point where one determines that these things count enough to have established a stake in one’s life.”¹⁰ The audience member I have in mind is Moore’s aesthetic adult, but in order to gain a fuller understanding of what this entails, we have to see that as important as *becoming* an aesthetic adult is *remaining* one. But this involves a continual pursuit of aesthetic development.

To see this point, consider Robert, who as a young man decides to perform a complete investigation of what the artworld offers. Based on his *prima facie* conception of the artworld, he

¹⁰ *Natural Beauty*, p. 232.

seeks out aesthetically great works. After surveying the aesthetic landscape, he finds a staggering number of different artforms, genres, works, movements, styles, and media for artistic expression. Undismayed, he begins to expose himself to the array, starting with novels and films. He reads the classics, while learning about literary styles and emerging schools. He views films, beginning with foreign classics, like those of Ingmar Bergmann, Jean-Luc Godard, Luis Buñuel, Robert Bresson, Vittorio de Sica, Akira Kurosawa. He moves on to English-language films, like those of Alfred Hitchcock, Sidney Lumet, Orson Welles, Woody Allen, Martin Scorsese. He reads film criticism, notes devices and mannerisms, and how they can give insight into interpreting the films. Before long, Robert also becomes interested in music—primarily classical music. He educates himself on music history and some of the underlying machinery of musical works. This leads him to opera, where he finds his interest in literary and filmic works mesh with his attraction to music. After he does this, he begins to travel the world, and visits world class museums, seeing in person history's most noted paintings. He reads up on how painters in history have approached the central task of applying paint to canvas. Of course, he doesn't engage with all of these arts and artworks coldly and with academic interest only. He does so openly, warmly, and receives many of the intellectual, emotional, and spiritual rewards of doing so. Let's imagine that at the age of forty, he emerges from this decades long enterprise in full command of the offerings of the artworld, an aesthete, with a robust awareness of the value of art and granting art "a stake" in his life. Behold the aesthetic adult!

But now suppose that Robert looks at his life thus far and realizes that his investigations have occupied so much of his time that he has overlooked some of the more mundane diversions. For example, he has allowed these last twenty years to pass without having exposed himself to any of the television shows his friends and family have talked about so much—he's never even

seen an episode of *Seinfeld*! Not wanting to shortchange himself, he commits himself to getting an education in pop culture, just as he had with the arts. Along the way, he discovers he really enjoys watching sports as well. He's endlessly impressed with the fact that he can pay a subscription to his satellite service company and ensure that there will be some game available for viewing at any time of the day. He discovers that now he prefers these forms of entertainment to the arts, opting to watch sitcoms rather than to read Solzhenitsyn, to see baseball games rather than to listen to Beethoven, to view romantic comedies rather than to view Monet. Art no longer has the stake in his life that it once did, other institutions occupying the role it had once played. Behold the aesthetic adolescent!

Robert's life is far from ideal. Even the most committed aesthetic adult wouldn't view aesthetic experience as a responsibility, as Robert originally does. But what this example suggests is that, in contrast to biological maturity, aesthetic maturity isn't something that someone achieves and possesses thereafter without further development. Robert *devolves* aesthetically because he is no longer *developing* aesthetically. We don't achieve aesthetic maturity and stop, but rather maintain it or lose it. We have to grow in order not to atrophy.

How does one maintain aesthetic growth? It's similar to muscle growth, which occurs only when the current musculature is pushed to the boundaries of its ability.

Aesthetic growth occurs only when aesthetic and intellectual sensibility is pushed to its boundaries. Thus as one works "one's way through the various levels of objective and subjective evaluation that one's aesthetic culture presents," this is just a series of challenges presented at each level. Difficulty of challenge is relative to each audience member, since, at any given time, each audience member will be at a different stage of aesthetic maturation. What is challenging earlier in the process will likely not be so challenging later. Aesthetic development starts with

works that are relatively simple, but present some degree of challenge. The young audience member adjusts and adapts in order to meet the challenge. Once these adjustments are made, aesthetic growth is spurred. This growth is not destined to happen, as it is with biological maturation, evident in this world full of aesthetic children; indeed, maintaining aesthetic maturity is hard work. What exactly is this hard work? That is, in what form do these challenges appear to the aesthetically maturing person? These challenges manifest themselves primarily in the task of making sense of narratives.

4. Making Sense of Narratives

Making sense of narratives is a fundamental challenge for every reader. Of course, there are differing levels at which one can be said to be trying to make sense of the narrative, and doing so at each level will utilize different skill- and knowledge-sets. At one basic level, we try to understand the language or imagery used in the narrative, while at other, deeper levels, we try to understand the events unfolding, the actions of characters within these events, and even more substantial intra- and cross-textual implications of these actions and events. As noted, the degree of difficulty of making sense of narratives at each of these levels will be relative to the individual. I find it very challenging to make sense of Abdelrahman Munif's *Cities of Salt*, while someone raised within the Bedouin culture might not. Perhaps most of us find the idiosyncratic lexicon of Anthony Burgess's *A Clockwork Orange* difficult to find our way in, but someone used to moving from one linguistic culture to another might not. Still, the challenge of making sense of narratives, manifest at these different levels, is a large part of why we find engaging

with narratives *worthwhile*, and our responding to and overcoming the challenge is a substantial part of what makes it aesthetically rewarding. For this reason the degree of difficulty presented by making sense of narratives cannot be excessively large. Challenge beyond a certain point becomes altogether forbidding. Many people reject the possibility of living the life of the reader because they are forced to confront artworks that are too challenging, and they assume that this is what aesthetic experience is like: finding works to be intractably hard to engage with. Of course, people often seek out simpler works with too little difficulty of challenge, for various reasons, but their goals in these circumstances are not often aesthetic. I watch films that present little in the way of challenge, but I generally walk away from them with very little sense of awe, bewilderment, or fascination—most often, with the film having made very little impression on me at all. In these cases I should not be said to have experienced the artistry of the film. Maybe it was there, and I just didn't detect it. Maybe there was little or no artistry in the film, because the work was intended only to be *entertaining*.¹¹

A natural question to ask is *how* do we make sense of narratives? To answer this question, we need to get a fairly detailed picture of what some of the challenges are. One challenge is filling in those gaps described by Carroll in chapter 5. There Carroll concludes that moral understanding could be deepened as a result of actions performed by readers who try to render the gappy narrative coherent. He argues that two actions central both to understanding narrative and deepening moral understanding are the (re)organizing of our beliefs in the appropriate way and the activating of certain moral emotions. I believe he is correct to include these two responses in his discussion, but there is a third response that it is equally important having to do with our understanding the actions of characters.

¹¹ Of course, art can be entertaining, and some of the aesthetic enjoyment we get from it is often blended with the entertainment we receive. Nonetheless, a work of art that is *merely* entertaining lacks the kind of aesthetic value we reserve for works that go beyond entertainment.

As readers, we often confront narratives with relatively little background information—about both the narrative world itself and the intentions of the author. We usually have to make basic assumptions right at the start. For example, we encounter the description of a character *behaving* in a certain way, but if we want to go further than this, and interpret the behavior as *action*, we have to assume intentionality behind the behavior. In other words, the character is presumed not to be a robot, operating under instructions encoded in a program, or an animal, merely responding to stimuli. We confront the fictional character just as we do real individuals, making some of the same background assumptions. Once we get past this basic level, where we interpret the behavior as intentional action, we have to make interpretations about what the particular intentionality is. That is, we try to make sense of *the actions* themselves. Here's my claim, one that isn't overly bold but pivotally important: *This should be familiar to us as the same goal we have when we engage in mindreading.*

Some of the very same challenges exist in both contexts. The challenges aren't *quite* the same, however, since in many cases, with fictional narratives, readers are just *told* what characters think and feel. For instance, in some narratives, the narrator is presented as omniscient, having privileged access to the thoughts and emotions of all the characters. In these cases there appears no need to make sense of the characters' actions, since the sense is made for us. In other narratives, we lack this privileged access, however, requiring us to perform some of the interpretation ourselves. Some narrators present themselves as having access to one, or only some, of the characters. In these cases the task of making sense of the actions of other characters would lie both with the narrator and the reader. Occasionally, narrators have *no* privileged access to the thoughts of characters and present themselves to readers as merely reporting on the actions of characters. Films and drama, in particular, utilize this mode of narration, since it's very

cumbersome to enter the minds of characters, given these media.¹² In these cases, again, we would need to draw on some of the same resources we do with real people.

But we shouldn't even pass over the case of the omniscient narrator as quickly as we have, since the determination of omniscience is itself only an interpretation based on the evidence presented in the text. We have, for example, the possibility of unreliable narrators. Sometimes the narrator is trying to deceive us, as in many of the events related by Humbert (the narrator) in *Lolita*. We glimpse clues that we aren't being told the straight story—clues probably intentionally included by the author in order to advise us not to take the words of the narrator as gospel truth. The unreliability isn't always because the narrator is disingenuous. Sometimes the narrator downplays certain details, and focuses on others, because he is presented as missing the importance of the former and overattributing importance to the latter. This happens often in the Sherlock Holmes tales, where Watson fails to attribute the appropriate salience to what Holmes later reveals to be clues essential to the solving of a mystery. Occasionally the narrator is self-deceived, as in Nabokov's *Pale Fire* and Gide's *The Immoralist*. In these cases we see something that the narrator doesn't, but this determination is often made late in the game, based on textual details appearing along the way.

Even when the narrator isn't unreliable, omniscient narrators aren't omnipotent. They aren't able to provide *all* the, or even the *most essential*, details relevant to understanding the actions of characters. The narrative is essentially incomplete, as we've seen, and things will inevitably be left out. It's thus up to the reader in many cases to fill in these gaps. Alan Palmer points out, "There is a large number of constraints on the amount of evidence that the narrator can make available in the discourse for any one character, even a protagonist." Since these constraints limit the narrator in ways that limit a full presentation of the mental life of the

¹² Although the chorus is often useful in drama for accomplishing this.

character, “it is part of the competence of the reader to construct, both from this written text and from the unwritten implications that comprise the gaps within the written text, a continuing consciousness for that character.”¹³ And this constitutes a significant challenge to any reader in making sense of the narrative—more so, when the narrative itself has been constructed in order to compel the reader to take up the challenge.

But is this challenge as prevalent as I have claimed? We can recognize the possibility of unreliable narrators and gappy narratives without concluding that the task of understanding the actions of fictional characters is importantly similar to understanding the actions of real people. And if we don’t draw the latter conclusion, is mindreading fictional characters practiced often enough to warrant my claim that some of the same capacities used in mindreading with real people are applied to the narrative? For example, when the reader sees the following opening sentences of chapter eight of Dickens’s *Our Mutual Friend*, presented with several statements about the thoughts and affective states of a character made by a narrator, does the reader really need to engage in mindreading to confirm them?

The minion of fortune and the worm of the hour, or in less cutting language, Nicodemus Boffin, Esquire, the Golden Dustman, had become as much at home in his eminently aristocratic family mansion as he was likely ever to be. He could not but feel that, like an eminently aristocratic family cheese, it was much too large for his wants, and bred an infinite amount of parasites; but he was content to regard this drawback on his property as a sort of perpetual Legacy Duty. He felt the more resigned to it, forasmuch as Mrs. Boffin enjoyed herself completely, and Miss Bella was delighted.¹⁴

¹³ *Fictional Minds*, Lincoln, NE: University of Nebraska Press, 2004, p. 41.

¹⁴ NY: Oxford University Press, 1998, p. 307.

While it may seem obvious that the Dickens passage could be read in the same way that one would read a sentence in a history book, things are not quite that simple. First, it's pretty widely accepted that if we buy into the story/discourse distinction, then we have to posit a narrative figure who constructs the discourse for us, based on his or her knowledge of the story. There are varying positions about how this figure obtains this knowledge, but we can ignore that issue for the time being. What follows from this, however, is that every narrative—or at least every fictional narrative—has a narrator—that is, a narrator-as-constructer-of-the-narrative—even if the narrator-as-reporter-of-the-narrative is so thin as to seem non-existent.¹⁵ Some figure is presumed to have been privy in some way to the events occurring in the story, and to be able to construct the discourse reflecting these events.

If there is a presumed narrator, even in those narratives which give evidence of a very thin narrator, what is our relationship to this narrator? One model of this relationship, called the text-as-communication model, has it that the narrator has encoded messages in the text, and it is the job of the reader to find the objective messages so encoded. There are several problems with this model, however. First, it implies one and only one correct interpretation, which runs against the grain of conventional aesthetic practice. Second, this seems more appropriate to the relation between the reader and the *author*, but then the role of narrator seems mitigated to the point of uselessness. Third, it leaves out stylistic and formalistic elements thought to be central to

¹⁵ This possibility seems least plausible when considering film and drama. But there are still decisions that have been made prior to the final presentation of the images and events that imply narratorial decision-making. For example, in Hitchcock's *Rope*, the narrative begins with a scream heard offstage, followed quickly by the image of David already dead. A natural question to ask is why the death itself, which looms so large throughout the film, isn't directly presented. But a theoretically more interesting question is who made this decision, and why did this person decide this? A quick answer is that Hitchcock made this decision, but then we would be supposing that Hitchcock, a person existing in this world, somehow knew enough about the world of the story to construct the discourse. This presents problems, however, since, for all intents and purposes, the story takes place in a non-actual, possible world. But as Kripke (*Naming and Necessity*, Cambridge, MA: Harvard University Press, p. 44) makes clear, gaining knowledge about a possible world isn't like using a strong telescope. How, then, does Hitchcock know enough about the story to construct the discourse? Maybe this theoretical problem is a proof by contradiction that the story/discourse paradigm is misguided, but if we choose to retain the paradigm, we have to reject the possibility that the filmmaker stands in the role of narrator in films. There would be an implied narrator then even in film.

approaching narrative artworks *qua* art. If the ultimate goal is to communicate a message, then everything else is subordinated to that goal. Surely there are other goals with equivalent, if not greater, importance.

Another model, called the conversation model, envisions the relation as between two conversational partners. This has the advantage of bringing in Paul Grice's notion of conversational implicature, something that narrative theorists believe is central to how we approach narratives. There is, of course, the obvious flaw of the model that conversations are inherently two-directional, and narratives are one-directional—from narrator to reader. This leads some narrative theorists, like Bortolussi & Dixon,¹⁶ to suppose that readers treat narrators *as if* they are partners engaged in a conversation with the reader. When we treat narrators in this way—as conversation partners—we take on many of the same commitments we do in everyday conversation. For example, we adopt the principle of charity, such that we interpret ambiguities in the way that makes the most sense of the ambiguous statement, given our best determination of the narrator's purposes. But we also treat the narrator of a fiction in ways significantly different from how we treat everyday conversation partners or even a narrator of a non-fiction: we *source-tag* the statements of narrators.

5. Source-Tagging Narrative Statements

Leda Cosmides and John Tooby argue that decoupling, the process in which we handle and manipulate propositions without necessarily ascribing truth values to them, is evolutionarily

¹⁶ *Psychonarratology*.

advantageous. One way we accomplish this goal of decoupling is *source-tagging*. We source-tag statements and propositions when we make the utterer of the statement or the offerer of the proposition explicit. We append to the statement or proposition an explicit statement of who or what the source is. This allows us to operate with the statement or proposition even if we are unsure of whether it's true.¹⁷ "Source tags are very useful," they say, "because often, with contingent information, one may not have direct evidence about its truth, but may acquire information about the reliability of a source."¹⁸ We treat this information as conditionally true—that is, conditional on finding out later that the source is reliable.

Statements made by narrators in fictional narratives are source-tagged as such. Thus when we read "Happy families are all alike; every unhappy family is unhappy in its own way" at the beginning of *Anna Karenina*, we tag that statement as deriving from the narrator.¹⁹ We may agree with the statement, but within the context of the fictional narrative, we process it in much the same way we would if we had heard someone say, "I heard someone say today 'Happy families are all alike; every unhappy family is unhappy in its own way.'" We may recognize the value of source-tagging narrative statements in the opening chapter of a novel or the first scenes of a film or play, but it might seem strange to think that we source-tag beyond these initial stages, after we have made determinations about the reliability of our source, the narrator. And Cosmides and Tooby note that "[b]ecause source tags, and especially derivations, are costly to maintain, mechanisms should monitor for sufficient corroboration, consistency with architecturally true information, or certification by a trusted source." They even say that when a

¹⁷ The advantage of this for logic was recognized by Gottlob Frege (*Conceptual Notation and Related Articles*, Oxford: Clarendon Press, 1972) where he implemented in his system devices for remaining neutral about the truth value of a proposition and for asserting the proposition.

¹⁸ "Consider the Source: The Evolution of Adaptations for Decoupling and Metarepresentations," in *Metarepresentations: A Multidisciplinary Perspective*, ed. Dan Sperber, 2000, p. 69.

¹⁹ This is assuming, of course, that we recognize that *Anna Karenina* is a fictional narrative.

threshold is met, “the system should no longer expend resources to maintain source information, and it should fade.”²⁰ Does this mean that at some point in the narrative, after we have judged the narrator reliable, we stop source-tagging statements made by the narrator and begin to treat them as true?

Cosmides and Tooby think not. Source tagging, they claim, is constitutive of the fictionality of the narrative. They write

There is a difference between false beliefs and fiction. In the first case, the agent *believes* that the propositions recounted are true. In the second case, the agent is *providing* propositions regardless of their truth value. And, in fiction explicitly labeled as such, the teller *intends* that the hearer *believe* that the teller is providing false representations, but that the teller intends them to form a coherent narrative.²¹

Because it’s not assumed that the narrator believes the propositions, the reader of the fiction should source-tag the narrative statements. But isn’t this strange? On the one hand, the reader is supposed to treat the propositions as truth-value neutral, which is a reason to source-tag, while, on the other hand, the reader recognizes the narratorial intention that she “believe that the [narrator] is providing false representations,” which suggests that the reader not source-tag the statements.

A second reason to source-tag narrative statements provided by Cosmides and Tooby is that readers recognize the need to suspend disbelief in order to enter fictional worlds.

“Representations of stories require hierarchical levels,” they write, “so that inferences are restricted in their scope of application to one level at a time. When applied to the content of the story itself, we are able to make all the ordinary implicit inferences necessary to understand the

²⁰ “Consider the Source,” p. 70.

²¹ *Ibid*, p. 91.

goals, intentions, beliefs, and motivations of the characters in the story. When applied superordinately to a representation of the agent who is telling the story, one can make inferences about the real world and the agent.²² These hierarchical levels are interesting, and will be the topic of our discussion in Chapter 10, where we will consider different perspectives readers adopt toward fictional narratives. It's not clear, however, that readers of fictions do typically suspend disbelief.²³ Thus I find Cosmides and Tooby's second reason to source-tag narrative statements strange as well.

Despite casting doubt on their reasons why we should continue to source-tag narrative statements, I am in agreement with Cosmides and Tooby that we do and should continue to source-tag. Source-tagging narrative statements is at the heart of the virtuality of narrative space I have argued is so valuable to narrative experience. At one level, we are trying to render the narrative intelligible. At another level we are engaged in interpretation, which is related to, but distinct from, the project of making the narrative coherent. Both of these are part and parcel of making sense of the narrative. We often are doing both at the same time; and in these cases, we are constructing interpretations "on the fly," since we are building on a foundation that itself is being simultaneously constructed. We try to understand the actions of a particular character as, say, those of a jilted lover and we test this hypothesis by gathering subsequent evidence, based on the actions of the character herself as well as those around her. This necessitates a large degree of freedom to try out different possibilities, since we are often comparing our working hypothesis to competing hypotheses. This freedom is maintained only by resisting the temptation

²² Ibid, p. 92.

²³ See Jerrold Levinson, "Emotion in Response to Art: A Survey of the Terrain," in *Emotion and the Arts*, eds. Mette Hjort & Sue Laver, NY: Oxford University Press, 1997, p. 23, where he argues that the suspension of disbelief thesis "unacceptably depicts consumers of fiction as having both a tenuous grip on reality and an amazing ability to manipulate their beliefs at will."

to ascribe truth-values to some of the narrative statements—even those indexed to the story-world, since these statements too can be false.

Cosmides & Tooby say one of the purposes of source-tagging is to update and revise currently held information. “To the extent that the [...] information is sourced, or its grounds and derivation are preserved in association with the data, then new data about the grounds can be used to correct or update its inferential descendants.”²⁴ Since interpretation depends crucially on being open to revision, we are always in the state of openness made possible by source-tagging. Because we are engaged with a virtual space, we are granted more license than in a non-fictional context, but with this license comes the responsibility to pursue whatever interpretations present themselves as live options. It would be very difficult to engage in the act of making sense of narrative, as I have been describing, without being in a sustained state of source-tagging narrative statements. Thus, as a precondition for overcoming the challenges at the heart of aesthetic maturation, source-tagging is essential.

6. The Life of the Reader

For the rest of the dissertation I will be dealing with fictional narrative artworks in relation to aesthetically mature audience members. These audience members live a particular kind of life that is set apart from those who maintain only a casual acquaintance with art—I will call this *the life of the reader*.²⁵ In chapters 8, 9, and 10 I will show how living the life of the

²⁴ “Consider the Source,” p. 69.

²⁵ Strictly speaking, the term should be *life of the narratee*, but there is a sense in which narrative artworks—even films and performances of plays—are *read*, so the term “life of the reader” should be understood to include this latter sense of “reader.”

reader can improve the capacity to mindread. Part of what maintains aesthetic maturity is the willingness to confront challenging narrative artworks, with the degree of challenge indexed to the audience member's stage of maturity. One of these challenges is making sense of narratives; one of the main challenges to making sense of narratives is understanding the actions of characters, especially in light of the fact that narrative statements are often source-tagged. Therefore, even when the audience member is told by the narrator what actions a character is performing and why, the audience member will still need to draw on her ability to mindread. We will see ways in which mindreading fictional characters is different from mindreading real people, and how these ways stand to improve everyday mindreading as a result. In Chapter 11 I will argue that living the life of the reader makes the aesthetic adult more inclined to care for those around her, an indisputable ethical effect.

Chapter 8

Defamiliarization, the Aesthetic Trajectory, and Decentering

At the end of chapter 3 I outlined four areas improvement in which could make us more accurate mindreaders, and thus put us on the way to being able to access and properly weight the interests of those around us. In the present chapter I will begin to construct an account that shows how each of these areas can be improved by a sustained and regular engagement with fictional narrative artworks of an appropriately challenging nature, found in the life of the reader. Here I will argue that this kind of engagement can improve on the ability to decenter, which turns out to be key to improving the other areas. In chapters 9 and 10 I will present evidence of how living the life of the reader can result in improvements of the other areas.

1. Sula's Reaction

Consider the following scene from the novel *Sula* by Toni Morrison. It occurs fairly early in the novel, the two teenagers Nel and Sula spending the afternoon by the river when the young boy Chicken Little approaches.

They stood up, stretched, then gazed out over the swift dull water as an unspeakable restlessness and agitation held them. At the same instant each girl heard footsteps in the grass. A little boy in too big knickers was coming up from the lower bank of the river. He stopped when he saw them and picked his nose.

“Your mamma tole you to stop eatin’ snot, Chicken,” Nel hollered at him through cupped hands.

“Shut up,” he said, still picking.

“Come here and say that.”

“Leave him ‘lone, Nel. Come here, Chicken. Lemme show you something.”

“Naw.”

“You scared we gone take your bugger away?”

“Leave him ‘lone, I said. Come on, Chicken. Look. I’ll help you climb a tree.”

Chicken looked at the tree Sula was pointing to—a big double beech with low branches and lots of bends for sitting.

He moved slowly toward her.

“Come on, Chicken. I’ll help you up.”

Still picking his nose, his eyes wide, he came to where they were standing. Sula took him by the hand and coaxed him along. When they reached the base of the beech, she lifted him to the first branch, saying, “Go on. Go on. I got you.” She followed the boy, steadying him, when he needed it, with her hand and her reassuring voice. When they were as high as they could go, Sula pointed to the far side of the river.

“See? Bet you never saw that far before, did you?”

“Uh uh.”

“Now look down there.” They both leaned a little and peered through the leaves at Nel standing below, squinting up at them. From their height she looked small and foreshortened.

Chicken Little laughed.

“Y’all better come on down before you break your neck,” Nel hollered.

“I ain’t never coming down,” the boy hollered back.

“Yeah. We better. Come on, Chicken.”

“Naw. Lemme go.”

“Yeah, Chicken. Come on, now.”

Sula pulled his leg gently.

“Lemme go.”

“OK, I’m leavin’ you.” She started on.

“Wait!” he screamed.

Sula stopped and together they slowly worked their way down.

Chicken was still elated. “I was way up there, wasn’t I?” Wasn’t I? I’m a tell my brovver.”

Sula and Nel began to mimic him: “I’m a tell my brovver; I’m a tell my brovver.”

Everything seems innocent enough. We’re not led to think Nel or Sula’s actions are backed by any sadistic motivation—perhaps boredom, but not malice. Sula even seems interested in ensuring that he enjoys himself. But then the following happens:

Sula picked him up by his hands and swung him outward then around and around. His knickers ballooned and his shrieks of frightened joy startled the birds and the fat grasshoppers. When he slipped from her hands and sailed away out over the water they could still hear his bubbly laughter.

The water darkened and closed quickly over the place where Chicken Little sank. The pressure of his hard and tight little fingers was still in Sula’s palms as she stood looking at the closed place in the water. They expected him to come back up, laughing. Both girls stared at the water.

The event of Chicken Little’s *sinking* comes unexpectedly. The reader is caught as unaware as Nel and Sula. If the reader were an observer within the storyworld, she would expect either Nel or Sula to jump into the water in an attempt to save the boy or, if perhaps neither of them knew how to swim, one of them to run for help. They don’t do this, however.

Nel spoke first. ‘Somebody saw.’ A figure appeared briefly on the opposite shore.

After Nel notes the figure, Sula immediately makes her way to the house on the other shore. Neither Nel nor Sula considers the possibility of rescuing, or seeking help for, Chicken Little. Instead, their first reaction, which they don't reject, is to find out who saw them and what he planned to do about it. They were seeking to save their own figurative skin, rather than saving Chicken Little's literal skin. Sula finds the house on the opposite shore to belong to Shadrack, a shell-shocked veteran known around town for his eccentricities. Clearly scared of him, Sula approaches and prepares to ask him whether he had seen. Before she utters a word, he says, "Always" as if answering a question she wasn't going to ask. She turns and runs away. When she gets back to Nel, who is still on the shore next to where Chicken Little had sunk and was still not visible, Sula is nearly hysterical and crying.

Nel quieted her. "Sh, sh. Don't, don't. You didn't mean it. It ain't your fault. Sh. Sh. Come on, let's go, Sula. Come on, now. Was he there? Did he see? Where's the belt to your dress?"

Sula shook her head while she searched her waist for the belt.

Finally she stood up and allowed Nel to lead her away. "He said, 'Always. Always.'"

"What?"

Sula covered her mouth as they walked down the hill. Always. He had answered a question she had not asked, and its promise licked at her feet.¹

If the natural expectation after the event of Chicken Little's sinking is that Nel and Sula would try to find some way to save him, the natural expectation is thwarted here. Their behavior prior to the event isn't obviously malicious, so why is their reaction to attempt to evade blame for the accident? Of course, a common initial response to accidents like this is for the person most directly responsible to desire to seek evidence of their innocence, or to try to devise some

¹ NY: Plume Publishing, 1982, pp. 62-63.

way to conceal their guilt, but that response generally gets superseded by the urgency of the situation. People rarely allow that first response to guide their actions, and those that do are, at best, criticized for being self-absorbed and, at worst, tried for criminal negligence. But these two girls—not so young that they could be excused for not seeing the moral implications of their decision—don't appear at this stage to be the sorts of individuals who would seek to prevent harm to themselves at the cost of harming a boy like Chicken Little. Why do they do it then?

The reader spends much of the rest of the novel trying to figure out the character Sula and this scene begins to take on more and more significance the further the narrative runs. Even here, at this early point, we know we have witnessed something fairly significant, because we are left with unanswered questions. Are these just self-interested teenagers who have such little value for life that they would rather save themselves from blame than save Chicken Little? Is there some prevailing culture of fear in the town Medallion that could explain why they acted the way they did? Were their motives more malicious than they appeared in the scene in the tree? Seeking answers to these questions is part of the pleasure of reading the novel. But for our purposes, we might consider *how* the novel succeeds in demonstrating the significance of the scene and raising so many questions. Imagine that in this scene, Nel and Sula immediately jumped in after Chicken Little to save him, to no avail. One implication of this alternate scene would be feelings of helplessness and sadness remaining in Nel and Sula, but it's likely that the scene would bear little weight in our attempts at interpretation, other than perhaps the causal implications appearing later in the narrative. So what is it about the actual scene that allows it to assume so instantly such significance?

Undoubtedly it has to do with the thwarted expectation we discussed earlier. What is clearly a fact about us as humans—that our first instincts are aimed toward self-preservation—is

also usually banished to the background. We don't get the opportunity to see upclose a case of people acting so immediately on their first instincts. This catches us off-guard, presents a novel phenomenon for us to consider, and arouses in us both wonder and curiosity. Something generally unfamiliar to us is treated as, and consequently made, familiar. This is not by accident.

2. Defamiliarization

This phenomenon of highlighting and developing an unfamiliar element is called *defamiliarization*.² Discussions of it begin in Coleridge and Shelley, but it enjoyed most prominence in the Russian formalist literary movement of the early twentieth century.³ The Russian formalists held that defamiliarization is what distinguishes literary works as *art*. We're not interested in the ontology of art here, so we don't have to evaluate this claim, but we should look closely at defamiliarization as a *narrative* feature, since making sense of narratives is often made more challenging because of it.

Victor Shklovsky highlights defamiliarization as a literary device in his essay "Art as Technique."⁴ There he begins with the fact that our observations of the everyday world are habitual and automatic, so much so that we eventually cease to notice what is happening around us. This "habitualization" has dire consequences for us. He quotes Tolstoy, who says "If the whole complex lives of many people go on unconsciously, then such lives are as if they had never been." Shklovsky agrees and adds, "And so life is reckoned as nothing. Habitualization

² The other side of the coin applies, too—what is *familiar* could be treated as *unfamiliar*.

³ The Russian term used is *ostraneniye*, which literally means "making strange."

⁴ In *Russian Formalist Criticism*, eds. Lee T. Lemon & Marion J. Reis, Lincoln, NE: University of Nebraska Press, 1965.

devours works, clothes, furniture, one's wife, and the fear of war." We become *numb* to life as it happens around us. Art rescues us from this numbness. Art exists, he says

that one may recover the sensation of life; it exists to make one feel things, to make the stone *stony*. The purpose of art is to impart the sensation of things as they are perceived and not as they are known. The technique of art is to make objects 'unfamiliar,' to make forms difficult, to increase the difficulty and length of perception because the process of perception is an aesthetic end in itself and must be prolonged. *Art is a way of experiencing the artfulness of an object; the object is not important.*⁵

The object is not important, but how we experience the object is. Really seeing what is before one is vital to being human, while being able to do things with or know information about the object is secondary. Thus what is most essential to art, according to Shklovsky, is not choosing the right object to depict, but presenting whatever object is chosen in the right way. This is why form supersedes content for Shklovsky and also why, when we find form in art, we find defamiliarization. Defamiliarization is the "artistic trademark":

we find material obviously created to remove the automatism of perception; the author's purpose is to create the vision which results from that deautomatized perception. A work is created 'artistically' so that its perception is impeded and the greatest possible effect is produced through the slowness of the perception.⁶

What is automatic becomes "deautomatized" and this is a move toward regaining one's humanity. While appropriate in many artistic works, defamiliarization would undermine the

⁵ Ibid, p. 12, emphasis in original.

⁶ Ibid, p. 22.

purposes of non-art, since most non-artistic written accounts would suffer from perception's being "impeded."

We can begin to see how the effect was achieved in the scene from *Sula*. Since we were expecting Nel and Sula to jump into the water after Chicken Little, had this happened, our eyes would have moved effortlessly along the page, with near-automaticity. When the unexpected happened, and we were compelled to dwell on the unfolding of the unexpected response, what resulted was "the slowness of perception." We noticed, and as a result of such noticing, we granted more weight to the scene than we would have had our expectations been borne out.

In art the effect of the slowing down of perception, the deautomaticizing of the automatic, is achieved through form, says Shklovsky, and in literary art, through forms of speech. Thus

we can define poetry as *attenuated, tortuous* speech. Poetic speech is *formed speech*. Prose is ordinary speech—economical, easy, proper, the goddess of prose is a goddess of the accurate, facile type, of the 'direct' expression of a child.⁷

Lest we think that Shklovsky would accuse novels, like those of Tolstoy, Gogol, Turgenev, Dostoevsky, of being "economical, easy, proper" and "the 'direct' expression of a child," we can see that "poetry" isn't to be distinguished by verse, but by the very kind of speech he is discussing: "formed speech." Indeed, "roughened form and retardation" is "the general *law* of art." Novels displaying such form, speech, and retardation would count as well. Shklovsky focuses on defamiliarization as the effect of poetic speech, but there are various ways of achieving this effect. One of these ways will be relevant to our discussion: foregrounding.

⁷ Ibid, p. 23.

3. Foregrounding

If defamiliarization is the effect of formal devices, then one of these devices and causes of defamiliarization is *foregrounding*. In its literal sense foregrounding is a device used in visual art. The space of the canvas is divided into back- and foreground, and visual attention is naturally held fast by those elements in the foreground. Visual artists use this device to direct the viewer's attention. In its metaphorical sense the effect is still the same, because whatever is brought to the fore in cognitive or linguistic space directs the audience member's attention. The history of foregrounding, in the metaphorical sense, is interesting,⁸ but we should focus our efforts on just a couple of moments. Jan Mukarovsky, of the Prague Circle linguistic movement, described foregrounding in ways reminiscent of Shklovsky's discussion of defamiliarization:

Foregrounding is the opposite of automatization, that is, the deautomatization of an act; the more an act is automatized, the less it is consciously executed; the more it is foregrounded, the more completely conscious does it become. Objectively speaking: automatization schematizes an event; foregrounding means the violation of the scheme.⁹

The language used here is a bit more precise than what we find in Shklovsky. One important element is the term "scheme." The habitualization of which Shklovsky spoke can now be understood to be an effect of our constructing schemes to which events are supposed to conform. It's when events don't so conform that "deautomatization" occurs.

⁸ See Willie van Peer, *Stylistics and Psychology: Investigations of Foregrounding*, Dover, NH: Croom Helm, 1986, for a comprehensive historical account of foregrounding.

⁹ "Standard language and poetic language," in *A Prague School Reader on Esthetics, Literary Structure, and Style*, ed. P.L. Garvin, Washington, DC: Georgetown University Press, 1964, p. 19.

David Miall¹⁰ elaborates on Mukarovsky's description, locating the notion of "schemes" in what is called "schema theory." Schema theory holds that in our everyday experience, we operate according to various "schemata" that have developed over time. These schemata are "stereotyped processes of behaviour by which we orient ourselves and know what to expect—as, for example, when we enter a restaurant."¹¹

When I enter a restaurant, I expect a general script, to which all of the players—maitre'd, wait staff, chefs, other diners, my dining partner, myself—will conform. First, I will be shown to my table. Second, I will order a drink. Third, I will order food. Fourth, I will be served my food. Fifth, I will be presented with the bill. Sixth, I will pay the bill. Finally, I will leave the restaurant. We often don't even pay attention to the order, because it's a presupposition that, modulo a detail or two, this is how it always happens.

It's when the script is disrupted, or arranged out of order, that our attentional mechanisms are engaged. If I walk into the restaurant and the maitre'd hands me a bill, an itemized list of drinks and entrees that I hadn't yet ordered, I would not know what to do. Do I pay it now? Will I have any choice of what to order or will the list be what I will receive? Or if the restaurant is organized in such a way that all diners sit at one large table and converse together, I would notice new things in this context. I would go off auto-pilot and on high alert, because my schema would no longer be applicable and thus not helpful. My habits will have been undermined, at which point I experience a state of groundlessness—what Miall calls "dehabituation."

Literary works, in particular, are apt to cause in readers this state of dehabituation, says Miall, through the device of foregrounding. Foregrounding, "whether at the phonetic level (e.g., alliteration, rhyme), the grammatical level (e.g., inversion, ellipsis), or the semantic level (e.g.,

¹⁰ *Literary Reading: Empirical and Theoretical Studies*, NY: P. Lang, 2006.

¹¹ *Ibid*, p. 17.

metaphor, irony),”¹² is the act of bringing into full view within the artistic space an element that is normally operating only in the background. Because what is usually in the background is now in the foreground, we are not sure how to respond. We seek for ways of responding, which naturally results in slowness of perception. “From the point of view of schema theory,” Miall writes, “a literary narrative will often be found not to be ‘well formed,’ whether in terms of overall plan, coherence, completeness, or conventions.” This happens in the following way:

1. Schema identification is necessary to allow the work of understanding to begin, but the application of schemata is likely to be thwarted or disrupted in a variety of ways by a range of textual features.
2. Providing a causal account of states and events is often problematic, due to uncertainties about how text elements relate to multiple potential relationships within the text. The status of narrative elements as states or events may itself be indeterminate.
3. A goal directed account of characters is often inadequate. Goals may be multiple, ambiguous or conflicting, so that their status becomes a focus of narrative interest.¹³

Dehabituation happens in step 3, where the reader loses her footing due to the loss of the trustworthy schema she held in step 1. It is at this step, I believe, that we find ourselves when we try to understand the scene in which Sula makes her way to Shadrack’s cabin, rather than trying to keep Chicken Little from drowning. The groundlessness resulting from dehabituation is the problem to be solved. How do we go about trying to solve it? How do we make sense of now seems senseless, given our useless schema? We need a new schema.

Finding a usable schema is hard, especially as the reader is in the narrative’s mid-stream.

“The initial schemata are likely to contradict each other in subtle ways,” says Miall, “providing

¹² David S. Miall and Don Kuiken, “Foregrounding, Defamiliarization, and Affect: Response to Literary Stories,” *Poetics* 22, 1994, p. 390.

¹³ *Literary Reading*, pp. 50-51.

the reader with signals of their inadequacy and impelling her to recognize that they have only provisional status. The *primary* work of the reader is thus to interpret the unfolding sentences of the text for clues to a more adequate schema.”¹⁴ Thus after losing her footing, a *major* component of making sense of the rest of the narrative will be to find the appropriate schema.

Miall & Kuiken argue that this is at the core of the pleasure we feel when we are engaged in literary interpretation. Defamiliarization prompts the reader to seek “refamiliarization.” This can happen at the various levels at which foregrounding occurs:

[T]he reader may review the textual context in order to discern, delimit, or develop the novel meanings suggested by the foregrounded passage [...] At the phonetic level, the reader may reconsider the context that enables identification of the feeling connotations of alliterative or assonant passages. [...] At the grammatical level, the reader may reconsider the context that helps to identify the ‘absent’ referent of an ellipsis. At the semantic level, the reader may recall other passages that extend or embellish a metaphor.¹⁵

What guides the reader through this process, they claim, is feeling. “[T]he feelings accentuated while reading foregrounded passages sensitize the reader to other passages having similar affective connotations” and “to other ‘texts’ (e.g., personal memories, world knowledge) having similar affective connotations.” Using these connections, the reader eventually “‘refamiliarizes’ or ‘thematizes’ the textual subject matter.”¹⁶ Thus, according to Miall & Kuiken, we have a predictable stagewise process of responding to foregrounding: (1) “foregrounding prompt[ing] defamiliarization;” (2) “defamiliarization evok[ing] affect;” (3) “affect guid[ing] ‘refamiliarizing’ interpretative efforts.”¹⁷

¹⁴ Ibid, p. 53, my emphasis.

¹⁵ “Foregrounding, Defamiliarization, and Affect,” pp. 394-395.

¹⁶ Ibid, p. 395.

¹⁷ Ibid, p. 404

This description of the process is consistent with the “aesthetic trajectory” discussed by Fitch, et al.¹⁸ They argue that successful art, in every medium, evokes the same sort of cognitive response that results in aesthetic delight. The aesthetic trajectory runs through three temporal points: familiarity, surprise, resolution. The first stage of familiarity is encountering the work with firm schemata in place. It’s marked by “a feeling of recognition and familiarity that invites and binds the viewer’s attention to the piece.” The piece, however, includes a defamiliarizing element in it, so that “as the eye¹⁹ plays over it, the painting begins to seem more adventurous.”²⁰ In the second stage dehabituation (in Miall’s framework) occurs, causing the viewer to consider new schemata. She begins to consider them one by one, attending to them in order to evaluate their chances at success. “The initial familiarity has evolved into novelty, uncertainty, and ambiguity. This creates a palpable tension between the reality of the object itself (strokes of paint on canvas) and the image it represents (a row of poplars by the water).”²¹ She is now groundless, seeking a way toward refamiliarization.

With the first stage came “a movement away from one’s initial sense of familiarity,” with the second came “uncertainty and ambiguity, with a heightened emotional tension.” With the third stage we get “a (partial) resolution of the tension caused by the second stage of the novelty,” with “an integration or synthesis of the earlier two, and not simply a return to the initial familiarity.”²² That is, we become refamiliarized, but not by returning to the original schema, which turned out to be useless.

¹⁸ W. Tecumseh Fitch, Antje von Graevenitz, and Eric Nicolas, “Bio-Aesthetics and the Aesthetic Trajectory : A Dynamic Cognitive and Cultural Perspective,” in *Neuroaesthetics*, eds. Martin Skov and Oshin Vartanian, Amityville, NY: Baywood Publishing, 2009.

¹⁹ They use visual art as an example, but the point generalizes.

²⁰ *Ibid*, p. 65.

²¹ *Ibid*, p. 66.

²² *Ibid*, p. 66.

In stage 3 we get only a *partial* resolution. Why just partial? Fitch et al. think that this is the key to aesthetic pleasure. Even at the end of the trajectory, there is a sense of incompleteness that “gives us pleasure” and “keeps us coming back for more.” An aesthetic experience, they say, “will often consist of multiple passes through the [...] three stage process.”²³ Thus an artwork’s richness is constituted, in part, by its propensity to reward attention, but not with a sense of an ending. Each trajectory leads into a new one.

On this view, a large part of what constitutes aesthetic pleasure is the *process* of moving from familiarity to defamiliarization to refamiliarization. This is part and parcel of confronting and overcoming the challenge presented by foregrounding. I want to slow down now a bit and look at this passage from defamiliarization to refamiliarization.

4. De- and Re-centering Perspectives

Making our way to the refamiliarization stage is one of the greatest challenges to making sense of narratives. For this reason we have to make adjustments to *ourselves*, since the narrative is fixed and stable. We have to adapt in order to be able to reconceptualize what seems for the time very alien to us. In many cases, like that of *Sula*, our challenge will have to do with explanation of character action. We are then faced with a decision: make these adaptations with the goal of understanding the narrative or recognize that we may have only a partial or errant grasp of what is happening. In general people choose the former inasmuch as choosing the latter often leads to giving up completely on the narrative itself. When we opt for the former choice,

²³ Ibid, p. 66.

and begin to make adjustments to ourselves in order to render coherent situations marked by defamiliarization within the context of character action, one of the adjustments we make is exactly the kind of adjustment necessary for improving empathic accuracy.

One of the reasons we find defamiliarization in the context of character action so unsettling is that we naturally project our own perspectives into those characters said to be *focalized*. People often speak about points of view in fiction. Thus the first-person point of view is marked by the use of “I” by the narrator, while third-person point of view is marked by the use of “he”/”she”/”it”. Using these terms, it has been argued, tends to obscure the fact that a narrative could be centered on a character and her experiences without the narrative being actually told from her point of view. What’s interesting is whose perspective is being presented, not how it’s presented. Thus Gerard Genette²⁴ suggests using the term *focalization* and Seymour Chatman²⁵ recommends *narrative voice*, rather than point of view. I will use the term “focalization.”

Bortolussi & Dixon describe three dimensions in which focalized character perspectives are presented: (1) descriptive reference frames (“a set of axes that determine how spatial and relational information in a perceptual description is portrayed”); (2) positional constraint (“constraint on the location of an agent who might have perceived the information”); (3) perceptual attribution (“perceptual knowledge that may also be attributed to characters in the story world.”).²⁶ As we saw in chapter 3 we naturally project our own reference frames, positional and perceptual, onto those around us, even when we rationally recognize that the perspectives of others will be markedly different from our own.

²⁴ *Narrative Discourse : An Essay in Method*, trans., Jane E. Lewin, Ithaca, NY: Cornell University Press, 1980.

²⁵ *Story and Discourse : Narrative Structure in Fiction and Film*, Ithaca, NY: Cornell University Press, 1978.

²⁶ *Psychonarratology : Foundations for the Empirical Study of Literary Response*, Cambridge: Cambridge University Press, 2003, pp. 186-187.

When we do this in narratives, however, we often run into problems fairly soon thereafter, unlike in real life where the effects of the disparity might not be significant enough to note. In many cases of defamiliarization, the very strangeness felt by the reader is a direct result of egocentric projection. Not only this, but we generally get immediate feedback that shows us that the perspective of the character is different from our own. Thus we have to constrain our natural propensity to project our own perspectives onto characters and in many instances we must actively reject these perspectives *vis-à-vis* the narrative context.

Therefore, one common requirement for moving through the aesthetic trajectory to arrive at the refamiliarization stage is decentering. The reader sheds her own perspective, and all the quotidian concerns that are attached to it, and becomes intimately associated with the perspectives of the focalized characters. Miall quotes Birkerts, who says ‘Our awareness, our sense of life, gets filtered into the character, where it becomes strangely detached from us’ (p. 93). Miall adds that to “emerge from the absorbed state of such reading can seem a distinct change of state, like awakening from a dream; there is a momentary and disorienting sense of engaging the gears of our own daily life once again.” But the reader doesn’t just decenter, shedding her own perspective to become free-floating. Refamiliarization will also draw on her ability to *recenter* on the perspective of the focalized character. Recentering doesn’t require loss of self, however. This is evident in the fact that as absorbed as readers become, they don’t *strictly identify* with characters, contrary to what some theorists have argued in the past. Readers always maintain distinctness of self from character. Otherwise, engaging with narrative would be as unpleasant as actually experiencing some of the negative events in narratives. But recentering, as described earlier, doesn’t imply loss of self. Imagining experience is itself an experience different from the experience being imagined. That is, there is a qualitative difference between

an unimagined experience and an imagined one. Recentering is imagining the experience of another, while retaining one's own identity. Someone who loses his own identity in the process of imagining the experiences of another is more often than not suffering from a pathological disorder.

Given that de- and recentering are common requirements in making one's way to refamiliarization, does this mean that one's ability to achieve these in real life is improved? Miall thinks so. Through the process of defamiliarization "we prepare ourselves for encountering experience in ways that are potentially (although not necessarily) more productive, thus enhancing the flexibility of our responses to the environment or our social interactions."²⁷ By getting better at re-centering and enhancing our "flexibility" within literary contexts, we become better at re-centering and more flexible in real contexts. Miall concludes that such training benefits us as "an 'offline' way of experimenting with emotions or experiences that might have dangerous or unpleasant consequences in the real world, gaining insight into their implications so that we know better how to act when similar situations occur in reality."²⁸

Inasmuch as the ability to recenter requires both the abilities to decenter and imagine, and engaging with appropriately challenging narratives requires all three abilities, successfully navigating these narratives presupposes all three. Further, we've seen that successfully navigating the waters of social understanding requires these abilities. We are thus on our way to seeing how reading narratives can begin to improve our empathic accuracy. While de- and recentering are not all that is required for becoming better mindreaders, these abilities are key to being able to realize the other areas of improvement, which are the topics of the next two chapters.

²⁷ *Literary Reading*, p. 198.

²⁸ *Ibid*, p. 17.

Chapter 9

Heuristic Mindreading and Metaknowledge—An Opportunity to Step Back and Reflect

In the last chapter we saw how engaging with appropriately challenging fictional narrative artworks often requires de- and re-centering perspectives, one of the areas of improvement for mindreading discussed in chapter 3. Possession of this ability is crucial for the improvement of the other three areas: the working order of mechanisms underlying mindreading, the gaining of metaknowledge regarding empathic accuracy, and increasing motivation to mindread. In this chapter I provide evidence of how such experience can improve *heuristic* mindreading—i.e., the mode of mindreading used in everyday encounters with others. Along the way we will see how readers can also gain metaknowledge about their empathic accuracy. In the next chapter I provide evidence of how *deliberative* mindreading and mindwriting can be improved, as well as how living the life of the reader can increase the motivation to mindread well. In the final chapter I argue that living the life of the reader can result in an increase in the capacity to care for others.

1. Mindreading at the Subject Level: What's the Story?

Readers utilize heuristic mindreading at two levels. The first is what Peter Lamarque and Stein Haugom Olsen¹ call the “subject level,”—that level at which readers are simply trying to

¹ *Truth, Fiction, and Literature : A Philosophical Perspective*, NY: Oxford University Press, 1994, pp. 286-287.

piece together the narrative events. Part of this piecing together is determining the nature of character behavior, which requires having some idea of the mental states underlying the behavior. This particular role for mindreading is different from those discussed in chapter 7—*viz.*, the filling in the gaps of the narrative and the checking of the reliability of narrators. Here the role of mindreading is located in the process of *attributing* mental states to characters. Mindreading plays a role at a deeper level, as well, called by Lamarque and Olsen the “thematic level,” where the reader attempts to “construe the subject under a perspective.” This is where the underlying themes of the narrative are determined. In this section I will discuss how mindreading occurs at the subject level and in the next section I will address its role at the thematic level.

At the subject level mindreading plays roles in understanding the narrative events themselves—i.e., answering the question “what is the story?”. While it’s often true that readers are given privileged epistemic access to the thoughts of some characters, this doesn’t show that mindreading is unnecessary. On the contrary, despite such access, readers engage in mindreading in fictional contexts all the time.

First, whenever there is little doubt that the narrator is unreliable, readers have to engage in mindreading. In chapter 7 I discussed the need for source-tagging narrative statements, but there I was primarily dealing with the possibility that the narrator is unreliable. In many narratives the narrator is clearly not to be trusted. In these situations everything the reader is told must be taken with a grain of salt and the reader must make sense of character behavior in any way open to her. One of these ways is mindreading fictional characters.

Second, when characters themselves sincerely report on their own mental states, either in internal narration or to other characters, they may fail at introspection and incorrectly report on the content of their own mental states. When this happens, the reader is told what the character

claims to believe and desire; but these reports conflict with the actions of the character. The reader then has to gather evidence, perhaps simulate the experiences of the characters, in order to get an accurate determination of these characters' mental lives.

Third, readers are often given *incomplete* epistemic access to the minds of characters, in which case they have to fill in the gaps through mindreading. This incomplete access could be the result of the gappy nature of narratives, in which case there will always be some incompleteness; but having full and complete epistemic access is not necessary to understanding narratives. How much is enough will depend on the particular narrative and making this determination will be the responsibility of the reader, which will require mindreading in some cases, depending on how large the epistemic gap is. Incomplete access could also be the result of an authorial decision to create suspense, tension, or conflict. The author may grant access to the minds of one or more characters, getting the reader familiar with the character's mental life, and then suddenly withdraw this access, prompting the reader to investigate why the character behaved the way he did. Finally, incomplete access manifests itself in situations where a focalized character's thoughts are open to the reader, but our access to non-focalized characters is as limited as our access to real people.

These are three ways in which readers are required to mindread because the epistemic access to the minds of characters is inadequate. There is also a role for mindreading to play in response to the *richness* of epistemic access found in certain artworks. In these cases it's not the gathering of content that is achieved by mindreading, but knowing to whom to attribute the content represented in the text. Lisa Zunshine² argues that the richness of access, evident in detail and complexity, explains why modernist, psychological novels of the early twentieth century are so difficult to process—they push the reader's mindreading ability to its *limit*. She

² *Why We Read Fiction: Theory of Mind and the Novel*, Columbus, OH: Ohio State University Press, 2006.

contrasts the cognitive ability to process long causal chains, like “A gave rise to B, which resulted in C, which in turn caused D, which led to E, which made possible F, which eventually brought about G, etc.,” with the ability to process sequences of nested mental state attributions, like “A wants B to believe that C thinks that D wanted E to consider F’s feelings about G.” She cites a study by Dunbar et al. that finds “that people have marked difficulties processing stories that involve mind-reading above the fourth level.” Once we get beyond *D* above, we begin to lose track of which states belong to whom.

Modernist novels exploit this limitation, forcing readers to go beyond the fourth level, driving them out of their comfort zones, and tapping their cognitive resources to degrees over and above what is required in any real life situation. Zunshine discusses Virginia Woolf’s *Mrs. Dalloway*, arguing that what makes this novel so difficult to understand is that it regularly pushes the reader’s mindreading capacity beyond the fourth and even up to the *sixth* level. She presents a passage from the novel as an example, saying the following about it:

Woolf *intends us to recognize* [by inserting a parenthetical observation, ‘so Richard Dalloway felt’] that Richard is *aware* that Hugh *wants* Lady Bruton and Richard to *think* that because the makers of the pen *believe* that it will never wear out, the editor of the *Times* will *respect* and publish the ideas recorded by this pen (6th level).³

In order even to understand Woolf’s narrative, the reader has to continually navigate through passages displaying this kind of richness and complexity.

This kind of thing happens in post-modernist novels, as well. For example, in Mario Vargas Llosa’s *Conversation in the Cathedral*, two characters are having a conversation in a bar

³ Ibid, p. 33, Zunshine’s bracket

called The Cathedral, and one of these characters Ambrosio is relating an account about the past to the other character Santiago Zavala. Zavala has to process the story as told by Ambrosio, thus grasping the latter's mental states now, his mental states in the past, and the mental states of those characters appearing in the account. The reader has to process the story of the novel by grasping Zavala's mental states, too.

The mindreading that occurs at the subject level is primarily important for attributing mental states of characters. The story itself will be largely incomprehensible if the character's behavior is inexplicable. Thus mindreading is used all the time at the subject level.

2. Mindreading at the Thematic Level: Putting Together the Picture of Dorian Gray

Mindreading occurs less often at the thematic level, mainly because much of the work for mindreading has already happened at the subject level. Nonetheless, mindreading does occur at the deeper level and sometimes determining a coherent theme crucially relies on mindreading. In this section I am going to give an example of how this works. This will be an extended commentary on Oscar Wilde's *The Picture of Dorian Gray*.⁴

In terms of richness and complexity no one would hold *The Picture of Dorian Gray* up against modernist novels like those of Woolf, Faulkner, and Joyce, but rendering a thematic interpretation of *Dorian Gray* presents some challenges the overcoming of which requires

⁴ If the reader wishes to skip the commentary, accepting that mindreading often happens at the thematic level, she may move on to section 3.

mindreading. I will lay out the interpretational problem below and show how mindreading helps the reader to solve it. We must take note, however, of Wilde's cautionary comment in the preface to the novel: "Those who go beneath the surface do so at their peril."⁵ Thus forewarned, down we go.

The plot of *Dorian Gray* is well-known: a young man is introduced to a life devoted to the appreciation of beauty by Basil Hallward and Lord Henry Wotton. He meets an actress Sibyl Vane, falls in love, proposes marriage, but after witnessing a particularly bad performance by her, he tells her the wedding is off. Sibyl then commits suicide, and when Dorian discovers this, he decides to pursue a life filled with hedonistic and egoistic pursuits. All the while he is living this life, a painting of him by Basil is continually changing, the image of Dorian becoming an older and visibly more rotten person, even as Dorian himself never ages or changes. Eventually, when Basil tries to gain access to the painting, Dorian stabs and kills him. In order to hide the evidence Dorian enlists the help of Alan Campbell, blackmailing him into disposing of Basil's body. Wracked with guilt, Campbell commits suicide. In the end Dorian stabs the painting, in an attempt to atone for his malicious actions, but as his servants come into the room afterward, they find the withered and horrible body of Dorian Gray, full of knife wounds, at the age he should have been, with the painting alongside him returned to its original form.

While Wilde is quite explicit that moral truths are not to be sought in artworks—that "[v]ice and virtue are to the artist materials for an art,"⁶—many commentators have nonetheless interpreted the novel to be a statement about the dangers of being over-obsessed with beauty, pleasure, or youth. On these interpretations the book would be a cautionary tale about striking a balance between lasting, internal beauty and ephemeral, external beauty. This kind of

⁵ *The Picture of Dorian Gray and Other Fantastic Tales*, NY: Fall River Press, 2010, p. 4.

⁶ *Ibid*, p. 4.

interpretation possesses some textual support. For instance, at the beginning of the story, Lord Henry Wotton initiates a period of influence on Dorian in the following exchange:

“You really must not allow yourself to become sunburnt. It would be unbecoming.”

“What can it matter?” cried Dorian Gray, laughing, as he sat down on the seat at the end of the garden.

“It should matter everything to you, Mr. Gray.”

“Why?”

“Because you have the most marvelous youth, and youth is the one thing worth having.”

“I don’t feel that, Lord Henry.”

“No, you don’t feel it now. Some day, when you are old and wrinkled and ugly, when thought has seared your forehead with its lines, and passion branded your lips with its hideous fires, you will feel it, you will feel it terribly. Now, wherever you go, you charm the world. Will it always be so?”⁷

From this conversation, it would appear that Lord Henry is showing Dorian the value of *youth*, but he then adds the following bit about *beauty*:

“You have a wonderfully beautiful face, Mr. Gray. Don’t frown. You have. And Beauty is a form of Genius—is higher, indeed, than genius, as it needs no explanation.”⁸

A reasonable response to Lord Henry’s comments then is that he is stressing the value of beauty, a prerequisite to which is youth, and, further, that since Dorian has both, he should give every effort to preserve them.

⁷ Ibid, p. 27.

⁸ Ibid, p. 27.

The story is about *Dorian's* obsessions. So if the tale illustrates the danger of the preoccupation with beauty and youth, there would be a pivotal moment when Dorian throws caution to the wind and issues a full-throated yawp declaring beauty and youth to be the only things of value. There *is* a pivotal moment, but when we look closely at it, and use our mindreading skills, we see that the obsession isn't exactly with youth and beauty. This moment comes at the end of Chapter 7, after Dorian has been under Lord Henry's influence for awhile, and has decided to buy out the contract of and propose marriage to the actress Sibyl Vane. On this night, he brings his friends Lord Henry and Basil Hallward to the theater to watch her performance as Juliet.

Unfortunately, she has a very bad night, and her performance is awful. Lord Henry and Basil cannot bring themselves to stay for the entire performance, leaving at the intermission. Dorian, humiliated, talks with Sibyl backstage after the play.

"How badly I acted to-night, Dorian!"⁹

she says to which he responds

"Horribly! It was dreadful. Are you ill? You have no idea what it was. You have no idea what I suffered."¹⁰

No, she says, she isn't ill. She tries to explain what had happened.

"Dorian, Dorian," she cried, "before I knew you, acting was the one reality of my life. It was only in the theatre that I lived. I thought that it was all true. I was Rosalind one night and Portia the other. The joy of

⁹ Ibid, p. 97.

¹⁰ Ibid, p. 97.

Beatrice was my joy, and the sorrows of Cordelia were mine also. I believed in everything. The common people who acted with me seemed to me to be godlike. The painted scenes were my world. I knew nothing but shadows, and I thought them real. You came--oh, my beautiful love!--and you freed my soul from prison. You taught me what reality really is. To-night, for the first time in my life, I saw through the hollowness, the sham, the silliness of the empty pageant in which I had always played. To-night, for the first time, I became conscious that the Romeo was hideous, and old, and painted, that the moonlight in the orchard was false, that the scenery was vulgar, and that the words I had to speak were unreal, were not my words, were not what I wanted to say. You had brought me something higher, something of which all art is but a reflection. You had made me understand what love really is. My love! My love! Prince Charming! Prince of life! I have grown sick of shadows. You are more to me than all art can ever be. What have I to do with the puppets of a play? When I came on to-night, I could not understand how it was that everything had gone from me. I thought that I was going to be wonderful. I found that I could do nothing. Suddenly it dawned on my soul what it all meant. The knowledge was exquisite to me. I heard them hissing, and I smiled. What could they know of love such as ours? Take me away, Dorian--take me away with you, where we can be quite alone. I hate the stage. I might mimic a passion that I do not feel, but I cannot mimic one that burns me like fire. Oh, Dorian, Dorian, you understand now what it signifies? Even if I could do it, it would be profanation for me to play at being in love. You have made me see that."¹¹

Sibyl's explanation makes it clear that her poor performance on this night is not an anomaly, caused by some momentary imbalance, but that her love for Dorian has *transformed* her, "freed [her] soul from prison." For the first time she "became conscious" that all the tools of the stage were nothing but artifice, that acting is nothing but dancing with shadows, associating with puppets. She once could *mimic* love in drama, because she had not felt love before, but now, after genuine passions has been inspired within her breast, she will never be able to mimic love like that again. A door has been shut, never to be opened again.

¹¹ Ibid, p. 98.

Sibyl perceives the shut door as a good thing, but Dorian responds in a very different way.

“You have killed my love.”¹²

She can't believe what he says, so he elaborates.

"Yes," he cried, "you have killed my love. You used to stir my imagination. Now you don't even stir my curiosity. You simply produce no effect. I loved you because you were marvellous, because you had genius and intellect, because you realized the dreams of great poets and gave shape and substance to the shadows of art. You have thrown it all away. You are shallow and stupid. My God! how mad I was to love you! What a fool I have been! You are nothing to me now. I will never see you again. I will never think of you. I will never mention your name. You don't know what you were to me, once. Why, once . . . Oh, I can't bear to think of it! I wish I had never laid eyes upon you! You have spoiled the romance of my life. How little you can know of love, if you say it mars your art! Without your art, you are nothing. I would have made you famous, splendid, magnificent. The world would have worshipped you, and you would have borne my name. What are you now? A third-rate actress with a pretty face."¹³

This appears to be a drastic response to a bad night's performance, but Dorian is responding to more than Sibyl's acting—he recognizes the nature of her explanation. She no longer has her art, she has nothing, she *is* nothing. And Sibyl's transformation has “killed” Dorian's love. But what exactly then did he love in her before his love had been killed? This is a pressing interpretative problem and is the point in the novel at which the reader is called most strongly to mindread. Dorian himself seems uncertain about what to make of his reaction. The reader will have to forge ahead alone, if she wants to know how to make sense of Dorian's rejection of Sibyl.

¹² Ibid, p. 99.

¹³ Ibid, p. 99.

Dorian's words shatter Sibyl. He leaves her, saying that she is such a disappointment that he will never see her again. He returns home. It's informative to attend to what Dorian attends to on his walk home: the sky that "hollowed itself into a perfect pearl; "huge carts filled with nodding lilies rumbl[ing] slowly down the polished empty street; "the air [...] heavy with the perfume of the flowers"; the flowers' "beauty seem[ing] to bring him an anodyne for his pain"; "a long line of boys carrying crates of striped tulips, and of yellow and red roses [...] threading their way through the huge, jade-green piles of vegetables;" "a troop of draggled bareheaded girls, waiting for the auction to be over;" drivers asleep on sacks; "iris-necked and pink-footed [...] pigeons [running] about picking up seeds;" the silent square on which his home is located; blank, close-shuttered windows; staring blinds; the pure opal sky; the roofs of the houses glistening like silver; a thin wreath of smoke rising from a nearby chimney—"a violet riband, through the nacre-coloured air."¹⁴

The eye that sees these scenes, that notices the fine detail, the color and shapes of the mundane constituents of the city, is not the mundane eye. It's an *artist's* eye. The words cohere into what could be considered a literary depiction of a scene painted by someone like Brueghel the Elder. And when the reader picks up on this highly artistic sense of perception, she naturally asks: why is Dorian so focused on the superficial features of the individuals in the city? Why is there an apparent wall between him and the rest of humanity?

When he gets home, he finds the portrait Basil had painted of him—the picture that Dorian had earlier called "part of myself." He thinks there is something changed about the painting and recalls his earlier wish:

¹⁴ Ibid, pp. 101-102.

He had uttered a mad wish that he himself might remain young, and the portrait grow old; that his own beauty might be untarnished, and the face on the canvas bear the burden of his passions and his sins; that the painted image might be seared with the lines of suffering and thought, and that he might keep all the delicate bloom and loveliness of his then just conscious boyhood.¹⁵

He also begins to reflect on his treatment of Sibyl. A “feeling of infinite regret came over him,” as he “remembered with what callousness he had watched her.” He tries to rationalize the way he treated her: “His life was well worth hers. She had marred him for a moment, if he had wounded her for an age. Besides, women were better suited to bear sorrow than men.”¹⁶ All the while, the painting continues to haunt him. “It held the secret of his life, and told his story. It had taught him to love his own beauty.” It watches him, he thinks, “with its beautiful marred face and its cruel smile.” He begins to feel “infinite pity,” not for himself, but “for the painted image of himself.” “It had altered already, and would alter more. Its gold would wither into grey. Its red and white roses would die. For every sin that he committed, a stain would fleck and wreck its fairness.”¹⁷ Why should he feel pity for a lifeless object like a painting? He hasn’t projected personhood onto it. He feels “infinite regret” for the way he treated Sibyl, but feels “infinite pity” for the painting. Why?

The reader doesn’t get any explanation for this behavior, so, if she wants one, she will have to construct it herself. At the subject level, the scene is intelligible as one in which he feels bad for what he’s done to Sibyl, and is approaching crisis, which is somehow related to the presence of the painting. But there is something deeper going on here. The reader then arrives at the most pivotal part of this pivotal scene. If every sin that Dorian committed “would fleck and

¹⁵ Ibid, p. 103.

¹⁶ Ibid, p. 104.

¹⁷ Ibid, p. 104.

wreck [the painting's] fairness," then, for the sake of the painting, he should try to change what he had wrought, right whom he has wronged:

But he would not sin. The picture, changed or unchanged, would be to him the visible emblem of conscience. He would resist temptation. He would not see Lord Henry any more--would not, at any rate, listen to those subtle poisonous theories that in Basil Hallward's garden had first stirred within him the passion for impossible things. He would go back to Sibyl Vane, make her amends, marry her, try to love her again. Yes, it was his duty to do so. She must have suffered more than he had. Poor child! He had been selfish and cruel to her. The fascination that she had exercised over him would return. They would be happy together. His life with her would be beautiful and pure.¹⁸

But now we should ask is whether Dorian really is setting out to make amends with Sibyl for the sake of the painting, or does the painting represent himself, or is he coping in the best way he can with the fact that he had behaved so badly? Or is there a non-obvious complexity that can be perceived only through extensive interpretation?

In the following chapter Dorian learns that he will never get to make amends with Sibyl, because she kills herself the night of the bad performance. Her suicide begins for Dorian a downward spiral that results in a self-centered headlong immersion in hedonism. This is why the end of chapter 7 is crucial for understanding the thematic content of the book. And as we can see, the task of understanding the end of chapter 7 is complicated by the fact that the access we have to Dorian's thoughts is not complete. We must engage in mindreading. We must *go back* to earlier parts of the book in order to construct an explanation for Dorian's behavior in chapter 7. It is perhaps most appropriate to re-visit those scenes in which *beauty* is discussed, because Dorian's rejection of Sibyl has something to do with beauty, though it's not clear yet what.

¹⁸ Ibid, pp. 104-105.

Consider how Basil describes Dorian to Lord Henry, prior to Lord Henry's meeting

Dorian:

“Dorian Gray is to me simply a motive in art. You might see nothing in him. I see everything in him. He is never more present in my work than when no image of him is there. He is a suggestion, as I have said, of a new manner. I find him in the curves of certain lines, in the loveliness and subtleties of certain colours. That is all.”¹⁹

Dorian is described in impersonal terms—as a “suggestion [...] of a new manner.” And later Dorian recognizes that Basil thinks of him in this way:

“[...] You like your art better than your friends. I am no more to you than a green bronze figure. Hardly as much, I dare say. [...] I am less to you than your ivory Hermes or your silver Faun. You will like them always. How long will you like me? Till I have my first wrinkle, I suppose. I know, now, that when one loses one's good looks, whatever they may be, one loses everything. Your picture has taught me that. Lord Henry Wotton is perfectly right. Youth is the only thing worth having. When I find that I am growing old, I shall kill myself.”²⁰

Here Dorian talks about youth, as though that is what is important, but clearly what he is objecting to is that Basil is valuing him only for his appearance. It's more than this too, however, because he also seems to be unhappy with Basil's treatment of him now, even when he does still have his “good looks.” Dorian seems not to have discovered yet that it's not Basil's focusing on looks that he is objecting to, but in the way he treats Dorian as an object—in particular, an *aesthetic* object. *A work of art.*

¹⁹ Ibid, p. 16.

²⁰ Ibid, p. 32.

Lord Henry is also guilty of treating Dorian as an object, but in ways less noticeable to him. For example, Lord Henry likes to see evidence of his influence in Dorian's behavior, much like a dog trainer is pleased when a dog starts to display obedience:

And how charming he had been at dinner the night before, as with startled eyes and lips parted in frightened pleasure he had sat opposite to him at the club, the red candleshades staining to a richer rose the wakening wonder of his face. Talking to him was like playing upon an exquisite violin. He answered to every touch and thrill of the bow. . . . There was something terribly enthralling in the exercise of influence. no other activity was like it. To project one's soul into some gracious form, and let it tarry there for a moment; to hear one's own intellectual views echoed back to one with all the added music of passion and youth; to convey one's temperament into another as though it were a subtle fluid or a strange perfume: there was a real joy in that—perhaps the most satisfying joy left to us in an age so limited and vulgar as our own, an age grossly carnal in its pleasures, and grossly common in its aims.²¹

Lord Henry's objectification of Dorian goes even deeper, as he treats him as an experimental test subject: "certainly Dorian Gray was a subject made to his hand, and seemed to promise rich and fruitful results." His stance toward Dorian would disturb Dorian to the same degree as Basil's stance, if Dorian had had knowledge of it. What is especially interesting is that, beyond youth and conventional physical or sexual beauty, what is being valued about Dorian, above all else, is his *aesthetic* beauty, which grants him the same status as Basil's "ivory Hermes or [...] silver Faun."

Notice now the parallelism between Basil's treatment of Dorian (and by extension, Lord Henry's as well) and the description of Sibyl Dorian communicates to them both. He says to Lord Henry

²¹ Ibid, p. 43.

“You said to me once that pathos left you unmoved, but that beauty, mere beauty, could fill your eyes with tears. I tell you, Harry, I could hardly see this girl for the mist of tears that came over me.”²²

“Mere beauty,” is now understood to be aesthetic beauty, even if Dorian is not yet aware of it. Perhaps he is using “beauty” in the physical or sexual sense, but this is unlikely, as we see how he describes her to Basil:

“She had all the delicate grace of that Tanagra figurine that you have in your studio, Basil.”²³

It certainly looks like Dorian is now thinking of Sibyl in the very same way to which he had earlier objected about Basil and himself. This brings to mind our earlier discussion of Nussbaum’s description of Maggie Verver’s thinking of her friends as *objets d’art*. There we saw that she wanted to handle them with extreme care and not to see the flaws in them that might threaten her “watertight” compartmentalizing way of approaching the world. What we see here with Dorian runs even deeper, since the entirety of the value of individuals is thought to reside in the aesthetic enjoyment one gets out of them. This way of looking at individuals is especially egregious as it manifests itself in Dorian’s rejection of Sibyl. Let’s return to the “horrible” performance she gives on the night he ends his relationship with her. The first part of her performance is described in the following:

Yet she was curiously listless. She showed no sign of joy when her eyes rested on Romeo. The few words she had to speak [...] with the brief dialogue that follows, were spoken in a thoroughly artificial manner.

²² Ibid, p. 59.

²³ Ibid, p. 86.

The voice was exquisite, but from the point of view of tone it was absolutely false. It was wrong in colour. It took away all the life from the verse. It made the passion unreal.²⁴

Dorian, Basil, and Lord Henry reserve judgment until the balcony scene, however, because she could still salvage the performance with a success here:

She looked charming as she came out in the moonlight. That could not be denied. But the staginess of her acting was unbearable, and grew worse as she went on. Her gestures became absurdly artificial. She overemphasized everything that she had to say. [One of the] beautiful passage[s] [...] was declaimed with the painful precision of a schoolgirl who has been taught to recite by some second-rate professor of elocution. When she leaned over the balcony and came to those wonderful lines [...] she spoke the words as though they conveyed no meaning to her.²⁵

Physically, she is stunning. Her performance is still woeful. The narrator's conclusion then is that

It was not nervousness. Indeed, far from being nervous, she was absolutely self-contained. It was simply bad art. She was a complete failure.²⁶

There are two ways we might read this statement that she was a complete failure. (1) As a component of the *performance*, which is the artwork, she fails, and because of her failure, the artwork is bad art. She performs poorly; consequently, the performance, the art object, is bad. But a second way is more coherent with the earlier scenes in which Dorian was treated himself

²⁴ Ibid, p. 94.

²⁵ Ibid, p. 95.

²⁶ Ibid, p.95.

as an artwork: (2) the artwork is not the performance, but Sibyl herself. As a bad actor, she is not aesthetically beautiful. She is a complete failure because she is a failure *qua* artwork. She has lost all value to Dorian, because (a) the only value Dorian had for her was aesthetic and (b) all of her aesthetic value has been stripped.

When Basil and Lord Henry rise to leave before the end of the play, Lord Henry says to Dorian, “She is quite beautiful, Dorian [...] but she can’t act.”²⁷ *Quite beautiful*. Certainly not aesthetically beautiful, since her acting would be her art. Lord Henry is, of course, not impressed with the kind of beauty she embodies, so he is clearly thinking of conventional physical beauty. And in spite of her having *this* kind of beauty, it’s not enough for Dorian. Thus Dorian rejects Sibyl.

This brings us back to the moment when Dorian decides to atone for his treatment of Sibyl, when he is said to have made this decision for the sake of a painting, which he is treating as a person. This can now be seen to be highly colored with irony. The act for which he must atone is treating a person as a work of art. But he also seems to have made this decision based on how his earlier behavior had affected the painting, a work of art he is treating like a person. To see the irony, however, we have to know that this is all going on inside his head. We weren’t able to read Dorian’s mind straight off the particulars of the scene at the end of chapter 7, having to return to earlier scenes. In this whole process we are constructing a thematic interpretation of the work, but as we do this, we are also interpreting Dorian’s behavior, made possible by our capacity to mindread. Without being able to enter into Dorian’s mind and determine what he’s feeling and thinking, we might be led to make all sorts of conjectures that get us nowhere. By using mindreading, by drawing on our ability to simulate the situation and on a set of principles that would help explain the behavior, we are able to construct a theme to the entire novel.

²⁷ Ibid, p. 96.

3. Folk Psychological Theory

Mindreading is thus utilized when readers try to make sense of narratives at both the subject and thematic levels. But does this utilization stand to improve our mindreading skills? Demonstrating that this is the case is, after all, part of the project undertaken in Part III. In this section I will argue that the kind of aesthetic engagement found in the life of the reader can add to the store of the reader's folk psychological theory. Recall that in chapter 3 I stipulated that heuristic mindreading is a combination of theoretical inference and transference simulation, understood as a kind of mental simulation where the source feeds initial states into the simulating mechanism, runs the simulation, and then transfers the output states to the target, under the assumption that the target is relevantly similar to the source. Here I want to begin the argument that engaging with fictional narrative artworks can improve heuristic mindreading by showing how it can augment the store of folk psychological principles.

Some theory-theorists—called nativists—hold that humans are born possessing a full set of hard-wired folk psychological principles. A relevant question that nativists should answer is whether these principles ever get amended or replaced. I don't intend to enter this particular debate, opting to make a relatively non-contentious assumption: we learn things about human behavior through the accumulation of experience and observation of human action and motivation. Whatever the nature of our folk theory, at some level it would have to get refined or amended or distorted; otherwise, experience would add nothing to our understanding of human behavior. We add to, remove from, and revise the set of folk principles, and this process can be informed not only by *actual* experience, but *virtual* experience, as well.

When we approach fictional works, we must make a determination of how true to life the work is. There are, of course, novels that cast aside nomological laws of the actual world, where transport vessels can travel faster than the speed of light, two concrete objects can occupy the same point in space, people can travel back in time, etc. These novels are not best described as realism, since the worlds in which the actions occur is very different from the actual world. A different kind of non-realism is possible, marked by a vast difference between sets of psychological laws of the actual and fictional worlds. It's up to the reader to determine whether or not the work is grounded in the psychological laws of the actual world. If not, it's often also necessary to determine how distant the fictional world is from the real world *vis-à-vis* its psychological laws. Arguably, there can't be too much distance or the reader would find it very hard to recognize actions of characters as intentional, rather than merely responses to stimuli. But what is key here is that making such determinations will draw on the reader's existing set of folk psychological principles.²⁸ Bortolussi and Dixon note that one of the ways that readers understand the behavior of characters is bringing to bear existing knowledge of the psychology of real people, such that interpreting character action is roughly the same kind of process as interpreting the actions of real persons. We thus apply our currently held principles to the actions of characters.²⁹

In other cases, however, readers go beyond mere application of already existing knowledge. For instance, sometimes readers take advantage of the virtuality of the fictional space and try out interpretations that would be too outside the norm to be justified in real life. If the reader has already determined that the work is fairly realistic in nature, she receives feedback on the new interpretation, feedback that can then be used to decide whether the principle adopted

²⁸ There are different strategies, such as using *scientific* psychological principles, evident in Freudian, Lacanian, and Jungian interpretations. But these are distinctions cultivated for, and pertinent in, academia.

²⁹ *Psychonarratology*, p. 141.

in the new interpretation can be generalized to both fictional and real persons. The same kind of feedback can be gathered by applying existing principles to character behavior interpretations, the results of which can be used to determine whether to continue to hold the tested principle. We can use the virtual space of fiction to confirm or disconfirm novel or existing folk principles. In addition to the reader's hypothesizing principles to explain character behavior, authors too can provide textual evidence of new folk principles, the evaluation of which readers can make in the course of interpreting the work. The reader needs to maintain a kind of reflective equilibrium in the course of such attempts. On the one hand, she must continually make determinations of the authenticity of the descriptions found in the work. If it turned out that the work makes several false assumptions about human nature, then any resulting findings will be unreliable. On the other hand, she can apply existing or new principles in explaining character behavior, testing both kinds against subsequent events. But she will also perform these actions when determining how realistic the novel is. Thus she must keep one eye on the reliability of the work and one eye on the events following her attempt to explain character behavior. The degree of benefit received from such attempts is directly tied to the virtuality of the fictional space and the degree of realism of the work itself. What gradually becomes clear, however, is that when the work's realism meets a minimum threshold and the reader has come to learn something about human behavior as a result of observing and trying to interpret the actions and intentions of characters, the store of folk psychological principles has been changed. Has it changed for the better? It depends on the ability of the reader to determine how much the artwork resembles the real world. But there is a great likelihood that the aesthetic adult will possess this ability.³⁰

³⁰ Notice here that I am not claiming that the aesthetic adult will always get it right—I am arguing only that the aesthetic adult *can* get it right and thus produce a warranted change in the store of folk principles.

4. What's Hecuba to Him, or He to Hecuba?

It might be objected here that *no* artwork resembles the real world enough to warrant a decision to add to one's store of folk principles as a result of observing character behavior. One reason is that we always relate to fictional characters in ways radically different from how we relate to real people. Hamlet seems to hold this view, finding it mysterious how an actor can exhibit such *passion* for the character of Hecuba in a play, when the actor must recognize that the character doesn't actually exist. He asks, "What's Hecuba to him, and he to Hecuba?" We might ask the same kind of question here: why think that any kind of mindreading that occurs in fictional contexts is the same kind that occurs in real contexts?

While it's not clear why it happens, it's clear *that* it happens. There is ample evidence that we do relate to fictional characters in similar ways to how we relate to real people. Bortolussi & Dixon offer the view that "even though literary characters and real people are ontologically distinct, they are processed in much the same way [...] literary characters are processed *as if* they were real people, and real people are processed in terms analogous to the categories brought to bear on the interpretation of literary characters."³¹ We can perhaps generalize to how viewers process fictional characters in film as well.

First, readers often have to use the relationships of characters with other characters to construct a mental life for them that allows an understanding of their behavior, but "real individuals too are perceived as deriving at least a part of their identity from their sphere of personal and social interactions and relationships, as well as from the social circles and general

³¹ *Psychonarratology*, p. 140.

context in which they conduct their lives.”³² Second, even though we often interpret fictional characters through the use of conventionalized codes and stereotypes, we do much the same in real life. Bortolussi & Dixon argue that “we are no less guided by conventionalized knowledge and stereotypical pattern of thought” in real life “than we are in our response to literary characters;” and in some cases, the stereotypes we use in real life are supplied by literature, such as when we describe people as having an “Oedipus complex” or as “quixotic, faustian, bovaesque, and donquanesque.”³³ Third, as I mentioned above, they argue that when we try to understand the behavior of literary characters “we draw on the same knowledge base of experience with people and situations.” We possess a large set of principles of the hypothetical form—e.g., “If a person blushes, then she’s embarrassed”—which we apply to observations of characters and, when we find the antecedent satisfied, we infer the consequent. Even if I am correct that we are often benefited by learning new principles as a result of engaging with fictional narratives, we still often bring real life psychological principles to bear on character behavior descriptions.

These are three features of character construction that are analogous to behavior explanation in real life. There is also neuroscientific evidence that we process character behavior and real person behavior in similar ways. In one study it was found that the same region of the brain thought to be crucial to mindreading—the temporo-parietal junction—is highly active when subjects read stories with descriptions of the mental states of characters. This region was not as active when subjects read stories containing descriptions of landscapes or stories with *physical* descriptions of characters, but no descriptions of the characters’ mental states. This

³² Ibid.

³³ Ibid., pp. 140-141.

suggests that the mechanisms underlying mindreading in real life are also at work when readers try to interpret character behavior.³⁴

So what's Hecuba to him? In some respects the same as Hamlet is to him.

5. Heuristic Mindreading

Thus far in this chapter I have argued that mindreading happens at the subject and thematic levels, that this mindreading isn't relevantly different from mindreading in everyday contexts, and that reading certain kinds of narratives can improve the set of folk psychological principles. Now it's time to focus more narrowly on the question of whether readers can become *better* heuristic mindreaders as a result of engaging with fictional characters. I will argue that readers can become better mindreaders. The most obvious place to start my argument is with that kind of character that requires the most of the reader's heuristic mindreading skills: the *epistemically opaque character*.

Epistemically opaque characters—that is, those characters whose mental states are totally obscured from audience members—are most commonly found in films, rather than literature, since most films don't enlist the device of the on-screen narrator. Whatever access film viewers have to the minds of the characters is a combination of filmmaking devices and the reader's capacity to mindread. Sometimes films can help viewers out with scenes of intimacy, in which a

³⁴ See R. Saxe & N. Kanwisher, "People Thinking about Thinking People: The Role of the Temporo-Parietal Junction in 'Theory of Mind,'" *Neuroimage* 19, 2003. See also Raymond A. Mar & Keith Oatley, "The Function of Fiction is the Abstraction and Simulation of Social Experience." *Perspectives on Psychological Science* Vol. 3, No. 3, 2008, pp. 173-192, for a comprehensive review of the body of literature suggesting that we process situations involving fictional characters in much the same way we process situations involving real people.

character is alone or communicating to another character information not widely shared with other characters. Even in these scenes, however, it's up to the viewer to determine *what* lies behind the character's solitary behavior or *whether* the information being communicated is genuine.

Some films exploit the epistemic opacity of characters, providing little evidence of what lies behind characters' behavior. The film *The American*, for example, details the last assignment of an assassin, Jack (played by George Clooney). Much of Jack's on-screen time is solitary, showing him silently preparing for his attempt to leave the life he has led for so long, with very few indicators of his goals and intentions. The filmmaker Anton Corbijn was so concerned that Clooney's representation of Jack's implacable expression would be too static for viewers that he decided to add to Jack's character traits a habit of chewing gum, so that there was at least some facial movement, even if it revealed little of what Jack was thinking. Other films foregrounding solitary experience, *Into the Wild*, for example, and television programs, like *Mad Men* and *Breaking Bad*, attempt to do the same thing. Viewers have to work hard to figure out what the characters are thinking. Lest we think this is limited to drama, we can also consider slapstick silent comedies, like those of Charlie Chaplin, where viewers have to process the pantomime behavior in order to know what is going on.

Very few narratives feature completely epistemically opaque characters, because this would radically limit the possibilities in the story. While the medium of film lends itself most easily to the use of epistemically opaque characters, literary narratives feature epistemic opacity as well.³⁵ When we find this kind of character in a literary narrative, usually we are prompted to

³⁵ See Joan Peskin & Janet Wilde Astington, "The Effects of Adding Metacognitive Language to Story Texts," *Cognitive Development* No. 19, 2004, pp. 253-273, where the authors conducted a study in which Kindergarten children were exposed to either a story with terms describing characters' mental states (the experimental group) or a story with no such descriptions (the control group)—i.e., with epistemically opaque characters. The researchers were

ask why the character is presented in this way, since literature is better suited than film to provide access to the mental states of characters. Many of the stories and novels of Ernest Hemingway are clear cases in which readers are not supposed to have wide access to the minds of characters. The story “Hills Like White Elephants,” for instance, limits reader access to the minds of the male and female characters, giving only a small number of pieces of behavior and the dialogue, which itself is obscure and cryptic. Piercing through the obscurity, however, is key to figuring out what the conversation is about. More than this, the obscurity is part of the conversation, which explains why Hemingway chooses to present the two characters in this way.

Some of the fiction of William Faulkner makes the same demands on readers. The story “Barn Burning,” for example, features the taciturn character, sharecropper Abner Snopes, who is standing trial at the beginning of the story for burning the barn of a local farmer. He is found not guilty due to insufficient evidence, but told not to stay in town, so he moves his family to another town. The reader is presented with a few pieces of Abner’s behavior, a few verbal comments he makes as they move, but for the most part Abner is inscrutable. The focalized character, Sarty Snopes, Abner’s young son, gives interpretations of Abner and his actions, but it’s obvious that because of his age and his relationship to his father, these interpretations are suspect. Thus the reader has to be sensitive to small details. When the family arrives at the next estate on which Abner will sharecrop, the reader observes, with Sarty, Abner approach the house belonging to the landowner, Major de Spain, adopting “the absolutely undeviating course” toward a “pile of fresh droppings where a horse had stood in the drive and in which his father could have avoided by a simple change of stride,” and stepping directly into the dung. After he enters the house,

surprised to find that members of the control group who were exposed to the story containing no mental state descriptions were better able to determine and attribute mental states in a false-belief explanation task (i.e., a task requiring subjects to explain why a figure possessed a false belief about an object) than members of the experimental group, prompting the conclusion that the inferences made by members of the control group “may be all the deeper because the children had to strive to infer meaning.” (p. 266)

against the remonstrations of the servant not to come in, he stands on a bright white rug in the entryway and “scrape[s] his boot clean on the edge of it.” This piece of behavior is crucial, not only to explaining what is to come, but also in trying to understand the trial with which the story begins. It’s clear that Abner deliberately picked up manure on his shoes with the intention of defacing Major de Spain’s home, even before having met the man. To the sensitive reader this suggests a deep resentment and envy of the landed class, which helps explain why Abner ultimately attempts to burn down de Spain’s barn too.

Epistemic opacity is found, to differing degrees, in a large number of narrative artworks. Most of these don’t *exploit* the limitations of the reader/viewer as the previous examples do, however, often offering ways to mitigate the difficulty. As we’ve already seen, even in these cases, narrative statements are often source-tagged and held in suspicion, evident in works like *Pale Fire*, *L’Immoraliste*, *Lolita*, and television shows like *Dexter*. And even if readers and viewers have some reliable access to focalized characters, there are still other characters that are as opaque as people we encounter in real life. Thus there are many opportunities to engage in heuristic mindreading. Let’s look now at ways that this kind of mindreading can be improved.

Aside from the value of sheer practice of mindreading in a context with a wide margin of error and low cost of failure, a large part of the potential for improvement of heuristic mindreading lies in *explaining* failed mindreading attempts. Because of the virtuality of the fictional context, the reader can construct interpretations of characters’ behavior throughout the narrative with little fear of negative consequences. In most instances readers won’t be tempted to try *novel* interpretations, since this is heuristic mindreading, after all, and is performed for the sake of understanding the narrative itself. But she must still decide which cues to focus on, how best to simulate the characters, and what folk psychological principles to apply. Once she does

so, she follows the narrative and gathers further evidence, which serves as feedback for her earlier attempts. Sometimes the later evidence suggests that she was wrong, that her attempt to construct an accurate interpretation of behavior failed, and as a result, the serious reader will take stock, see why she failed, and how she can construct a more accurate interpretation. In other words, she undertakes *reflection* about her heuristic mindreading attempt, made possible by the fictional context. She does this, in part, by asking what explains her failure.

It might have been simply due to *her error*, so she must ask in what ways she might have erred. Did she focus on one particular detail, rather than another more meaningful one? Did she wrongly de-emphasize one, or a set, of details? Did she fail to simulate in the right way? Is there something about her personally, or something about the character, that threatens the analogical relation between her and the character? Did she apply the wrong folk psychological principles? If so, were they inapplicable to just *this* situation or are they simply false? It might have been due to *author error*. Did the author assume false psychological principles? Did he emphasize certain behavioral cues, only to treat them as minor later? Did the later behavior unjustifiably reflect a different set of background beliefs than the earlier behavior? It might have been due to *authorial intention*. Were there omissions, distortions, misdirection? If so, were these judgments made by a particular figure within the narrative, like the actual narrator or a character occupying the role of interpreter of character behavior? Do these judgments seem intended to create an impression that would later be revealed to be in error? If so, were the original judgments legitimate, based on the existing evidence, or was it all due to authorial sleight of hand? Finally, it might have been due to *character error*—that is, error committed by the character whose behavior the reader failed to accurately interpret. Did the character commit errors in *mindwriting*?

Having the time and opportunity to reflect on mindreading failures is an asset presented to the reader by the virtual nature of fictional narrative artworks. More specifically, the value lies in the chance to acquire what was referred to earlier as “metaknowledge”—knowing how successful one is at mindreading and when adjustments are necessary. In real life it’s usually not feasible to direct one’s attention to how one mindreads in everyday relationships. Because the costs of slowing down and engaging in reflection on one’s mindreading attempts are not prohibitively high in aesthetic contexts, and because making sense of the behavior of characters is often necessary before moving ahead in the narrative, readers are able, and often compelled, to take this metacognitive stance toward their previous attempts to mindread. This opportunity rarely exists in real life, but the effects of this reflection can be seen in readers’ eventual relationships with real people. Even if readers don’t walk away from experiences with narrative artworks with the determination to attend to failures in real life, the insights they learn in fictional contexts can cross over. When a reader evaluates her failures and determines that they were due to *her* error, she receives precious information, but even when she determines that they were due to authorial error or intention or to failed mindwriting, she sees what could have made her attempt more likely to be accurate. And these insights can remain and even be applied to her attempts in real life. The value of this opportunity cannot be overstated. At the very least the attentive reader can see that there are gaps in her mindreading attempts. Ideally, the reader discovers what explains these gaps and receives an insight into how to bridge these gaps. Over the course of a life these possibilities begin to mount up. Therefore, the life of the reader stands to provide metaknowledge about mindreading and to improve heuristic mindreading.

Chapter 10

Moved to Mindread: The Role of Critical Equilibrium in Deliberative Mindreading

In the last chapter I began presenting details of my account of how living the life of the reader can improve the mechanisms underlying mindreading. There I focused on the heuristic mode of mindreading and showed how the reader can reflect on her mindreading failures in the fictional context and thus see how she could have done better, but also see what other factors contributed to her failure and how these factors could have been different so that her attempts might have been more successful. In this chapter I complete the account of how engaging with fictional narrative artworks can improve the mechanisms of mindreading. Here I show how such engagement can increase the motivation of the reader to mindread in real life. By witnessing fictional conflict, and the role mindreading failures play in such conflict, the reader realizes the value of accurate mindreading in real life conflicts. In the next chapter we'll see how such witnessing can result in the increase in the capacity to care for others in the real world.

1. Internal and External Perspectives

Before we look at how engagement with fictional narrative artworks can improve deliberative mindreading, we need to consider a distinction discussed by Peter Lamarque¹ and further developed by Gregory Currie:² the distinction between internal and external perspectives. To get an idea of what these two perspectives are, let's return to the passage from Toni Morrison's novel *Sula* discussed in chapter 8. When Chicken Little sinks and Nel and Sula's response is not to save him, but to investigate whether anyone had seen what had happened, the reader can try to make sense of their response in two ways.

The first way involves asking questions concerned with *character*-motivation. Why would the first instincts of Nel and Sula be to remove evidence of their role in the event, rather than to take steps to remedy the event as it was in progress? The reader might draw from what she knows about these characters, where they live, what is expected of females their age and/or race. She might also determine if there is anything peculiar to these two characters, based on previous scenes, that would explain their decision. She might also reason from *this* event to what must be true about the moral characters of Nel and Sula. In short, the relevant factors in this first way of approaching the event would be similar to how we would try to explain their behavior if they were real people.

The second way involves stepping back from the scene, construed as an event occurring in the storyworld, and seeing it as a component of the discourse,³ which prompts the reader to

¹ *Fictional Points of View*, Ithaca, NY: Cornell University Press, 1996.

² *Narratives and Narrators*, Oxford: Oxford University Press, 2010.

³ See chapter 7 for a discussion of the distinction between story and discourse.

ask questions, not about character-, but about *author*-motivation.⁴ The reader then asks what purposes the author is seeking to achieve, and how these purposes are brought to bear on the interpretative problem. Why *this scene here*? What conclusions are we *intended* to draw, based on any textual clues we can gather? Are these conclusions provisional or absolute? Am I *supposed* to feel more or less sympathy for these characters?

These two points of view are referred to by Lamarque respectively as the internal and external perspectives. On the one hand, the internal perspective on fictional works is that of “imaginative involvement;” from this perspective “what we call ‘characters’ exist as ordinary people.”⁵ The external perspective, on the other hand, presumes an “awareness of artifice,” and the possession of “our point of view in the real world” as readers of a text.⁶

When we are working toward making sense of a narrative and finding our footing, we often draw on the interrelations between the internal and external perspectives.⁷ We try to locate possibilities that make sense of what we’ve been presented, but the range of possibilities will, to a great degree, require the assumption of each perspective. Currie⁸ points out that in interpreting narratives, we are after the “salient possibilities,” given certain constraints provided by each perspective. For example, in Mark Twain’s *Huckleberry Finn*, it makes no sense to ask why Huck doesn’t just buy the slave Jim’s freedom, since children were not authorized to make such purchases in the time in which the story was set. Of course, as a way to resolve the tension rising

⁴ A distinction is often made between actual and implied author. The actual author in this case would be Toni Morrison herself, with her own actual mental states, intentions, and biography. The implied author would be a figure constructed strictly from the text, such that authorial intentions are inferred from details in the text, but further details are also interpreted according to the constructed intentions. Thus the implied author is *hypothetical* in nature and open to multiple revisions. In the present context, this distinction isn’t relevant, though at a broader level, I hold with the hypothetical intentionalists, like Jerrold Levinson (1992), that we’re trying to discover what an author *must have* intended, based on the text and details surrounding the text, and not what the actual author *in fact* intended.

⁵ *Fictional Points of View*, p. 53.

⁶ *Ibid*, p. 53.

⁷ This dichotomy of perspectives is also manifest in real life. One example is the religious belief held by some that a deity has certain intentions that are embodied in the occurrence of events. Thus I might not ask why a loved one performed a certain action, but what God intends me to do given the action.

⁸ *Narratives and Narrators*.

in Chapter 32, having this occur would certainly present an out for Twain, but the story itself would lose all verisimilitude—the option is not even on the radar. However, the sudden appearance of Tom Sawyer in Chapter 33 at the Phelps house too seems far-fetched. This is not something that the reader would have expected, and we might even criticize the novel for its use of the *deus ex machina*. We can see, however, why Twain went this way—it provided an exit strategy for Huck at the end of chapter 32, and it's at least internally salient.

For critical purposes too, the perspectives provide constraints on one another. Currie writes

With narrative we are often guided by something else as well; our awareness—perhaps a not very conscious one—of the narrative-maker's own purposes. At that point the external perspective intrudes. Suppose that our story's hero has been shot and seriously wounded. The story, I will assume, is a naturalistic one that encourages us to import into its content the sorts of assumptions about causation we make concerning the real world. Serious bullet wounds are, we assume, life threatening, so the death of the hero ought to be a very salient possibility at this point in the story, by whatever internal standards we choose to apply. But if we are only one-third of the way through the story and the author would be faced with narrative difficulties as well as a rebellious readership if the hero died at this point, we may be confident that the hero is not going to die—though the author might wish us to be more uncertain.⁹

Why don't we include within the range of salient possibilities the death of the hero one-third of the way through the narrative? Because it would make no sense from the author's perspective. Thus our interpretation doesn't take seriously this possibility. Sometimes the hero does die well before we expect it, defying our expectations. For example, in the film *No Country for Old Men*, Llewellyn Moss is presented as the protagonist—the focalized character—which produces in us

⁹ Ibid, pp. 54-55.

expectations that we will accompany Moss for most of the narrative. Then at around two thirds of the way through, he is killed by Anton Sigurh, effectively leaving us with no protagonist. From the external perspective we ask what purpose is served by this event here. It almost makes no sense, however, to ask from the internal perspective why Llewellyn dies when he does. Indeed, we might even ask why it has taken so long, since Sigurh is so adept at killing, and Llewellyn seems so outmatched by Sigurh's skill. It wouldn't surprise us at all to find that Llewellyn died when he did and how he did, if he were a real person being described in a newspaper article or by someone we know.

This tension between perspectives can also present indeterminacy of interpretation, contributing to diverse critical evaluations. Take the final episode of *The Sopranos*. There is very little justification from the internal perspective for the belief that Tony is killed in the end. The hit has been called off, but the scene is set up in ways very similar to those in which Bobby was killed and Silvio was seriously wounded—so similar in fact that we would feel manipulated if it turned out that Tony isn't killed, too. Suppose a "lost scene" were found that followed that infamous black shot at the end of the episode. If this scene revealed that Tony had not been killed, we would have a very hard time figuring out what all the cinematic devices were trying to suggest, since they appear to be foreshadowing Tony's death (external perspective). On the other hand, if the scene reveals that Tony has been killed, we would also have a hard time explaining why it had occurred, since the hit had been called off and there seems to be no reason why someone would still be trying to kill Tony (internal perspective). The only way that the "lost scene" would be satisfactory from both perspectives is if Tony is killed by a patron of the diner and, after it occurs, some indication is given of why the killer is performing the killing—e.g., if

the killer were to say “This is for Phil.” But we don’t get that lost scene, so when we think about the scene from both perspectives the finale is inexplicable.¹⁰

There are then constraints *within* and *between* the two perspectives. Moreover, adopting one or the other can, as Currie argues, aid us in determining salient possibilities in interpretation and also, as I have argued, aid us in constructing criticisms and evaluations. Not only are these aesthetic criticisms—which would be from the external perspective—but could also be moral—when made from the internal perspective. A reader often toggles back and forth between the two perspectives, weighing possible interpretations against the respective constraints. Currie claims that we understand narrative “only by combining the internal and external perspectives;”¹¹ Lamarque says that “there is no reason why they cannot be held simultaneously.”¹² Neither goes so far as to argue that we *do* end up holding both simultaneously. I believe, however, that when trying to adequately explain character behavior, the adept reader does occupy both perspectives—or to be more precise, a perspective having in view constraints provided by both perspectives. And this blended perspective is not so different from the perspective of the author¹³ as he *constructs* the narrative.

¹⁰ For a fuller discussion of the tension between the internal and external perspectives, and the *Sopranos* finale, see my “Narrative Tension: Solving the Puzzle of Anxiety.”

¹¹ *Narratives and Narrators*, p. 52.

¹² *Fictional Points of View*, p. 14.

¹³ I argue in “Non-Branching Moderate Moralism” that the role of narrative constructor is occupied, not by the author, but by the implied narrator, but I leave this complication aside here.

2. The Author as Mindreader

The novelist Charles Johnson once commented to me that his teacher, John Gardner, a novelist himself, often said that he became a morally better person when he engaged in writing fiction. Gardner is well known for his views on the connections between fiction and morality, found in his book *On Moral Fiction*.¹⁴ What is especially pertinent to the current discussion are his statements about the *process* of writing, how writers think about their characters and how this thinking can also be mirrored in how they think about other people. We will see that the method of discovery in the author is very similar to how readers discover facts about characters. Moreover, this method centrally involves deliberative mindreading including *transformation simulation*.¹⁵

Gardner notes that, contrary to appearances, ideas found in fictional works don't spring from the author fully formed, like Athena from the head of Zeus, but emerge out of a process in which a "writer *discovers, works out, and tests* his ideas in the process of writing."¹⁶ At its best, writing fiction is "a way of thinking, a philosophical method."¹⁷ We can see this method in the progression of work found in authors' successive drafts of their novels, like Tolstoy's *Anna Karenina*:

Thus we see Tolstoy beginning with one set of ideas and attitudes in *Two Marriages*, an early draft of *Anna Karenina*—in which Anna, incredible as it seems, marries Vronsky—and gradually *discovering*, draft by draft, deeper and deeper implications in his story, revising his judgments, stumbling upon connections, reaching new insights, until finally he nails down the attitudes and ideas we find

¹⁴ NY: Basic Books, 1978.

¹⁵ See chapter 3 for the differences between transference and transformation simulation.

¹⁶ *On Moral Fiction*, p. 107, my emphasis.

¹⁷ *Ibid*, p. 107.

dramatized, with such finality and conviction that it seems to us unthinkable that they should not have burst full-grown from Tolstoy's head, in the published novel.¹⁸

I've emphasized the words "discover," "works out," "tests" and "discovering" in the quoted material in order to show how writers are often gathering clues about narrative events and character motivation in ways very similar to how readers go about trying to figure characters out and see why they do what they do. Peter Brooks¹⁹ thinks the reader's investigations are akin to those undertaken by detectives in stories like the tales of Sherlock Holmes. There is a determinate course of events that have already occurred in mysteries, and it's up to the detective to gather clues so that he can *re-construct* the events. The reader of a narrative also must gather clues in order to reconstruct events that occur in the storyworld, because it is underdetermined by the discourse. There are gaps, as we've already seen, some of them by virtue of the very nature of narratives and some by authorial intention. Thus, even though the author has already made his discoveries by the time the reader gets her hands on the text, the processes undergone by both the author and the reader are still very similar. So how does the author make his discoveries?

"Much of what a writer learns he learns simply by imitation," says Gardner. He describes in detail how this imitation occurs:

Making up a scene, he asks himself at every step, 'Would she really say that?' or 'Would he really throw the shoe?' He plays the scene through in his imagination, taking all the parts, being absolutely fair to everyone involved [...] and when he finishes the scene he understands by sympathetic imitation what each character has done throughout and why the fight, or accident, or whatever, developed as it did.²⁰

p. 109

¹⁸ Ibid, pp. 108-109, my emphasis.

¹⁹ *Reading for the Plot*, Cambridge, MA: Harvard University Press, 1992, ch. 1.

²⁰ *On Moral Fiction*, p. 109.

The “sympathetic imitation” of which Gardner speaks is a perspective-shift, thus not what Peter Goldie²¹ calls “in-his-shoes imagining,” reflected in asking oneself ““What would I think and feel if I were in his shoes?””²² In fact, it’s a re-centering of perspective of the kind discussed in chapter 3, where we considered transformation simulation defended by Robert Gordon, and chapter 8, where we saw how defamiliarization and foregrounding lead readers to make this kind of shift. The writer’s discoveries are made possible by this shifting of perspectives, producing insights, not only about the characters, but about relating to individuals in the real world as well. “[W]hat one learns if one is oneself a writer who has tested each of his fictional scenes against his experience of how things seem to happen in the world,” Gardner writes, “is that scrutiny of how people act and speak, why people feel precisely the things they do, how weather affects us at particular times, how we respond to some people in ways we would never respond to others, leads to *knowledge, sensitivity, and compassion.*”²³ Through the process of imitating characters, some of whom will differ greatly from the writer, the writer learns about humanity. “In fiction,” he says, “we stand back, weigh things as we do not have time to do in life.” Such weighing allows writers to test their opinions, putting “them under every kind of pressure one can think of” and seeing whether these opinions are justified. Echoing Currie’s formational account seen in chapter 5, Gardner describes the fictional context as “a laboratory experiment too difficult and dangerous to try in the world but safe and important in the mirror image of reality in the writer’s mind”²⁴ in which we can engage in “a simulation of real experience” that is “morally

²¹ “How We Think of Others’ Emotions,” *Mind & Language* Vol. 14 No. 4, 1999; *The Emotions: A Philosophical Exploration*, NY: Oxford University Press, 2000.

²² See “How We Think of Others’ Emotions,” pp. 397-398 for a fuller discussion of this kind of imagining.

²³ *On Moral Fiction*, pp. 113-114, my emphasis.

²⁴ *Ibid*, pp. 115-116.

educational,” though admittedly “unverifiable because it depends on the reader’s sensitivity and clear sense of how things are, a sense for which we have no tests.”²⁵

Through imitating characters, the writer is able to test hypotheses about human behavior, supporting the claim made in the last chapter that folk psychological principles can be tried out in the fictional context, as well as to gain insights into behavior. There are going to be genuine surprises when characters are placed in situations, given certain psychological and moral profiles, and then simulated by the writer. This is as it should be. As Robert Frost once said, “No surprise in the writer, no surprise in the reader.”²⁶ These surprises, the discoveries made by the writer, are where the insights lay.

As I have suggested above, the same kinds of discoveries are made by the reader, when trying to complete the narrative. A major difference between the writer’s discovery process and the reader’s is that the writer toggles back and forth between the internal and external perspectives in order to decide how *things will go*—that is, the nature of the narrative itself is undetermined at this point—while the reader toggles back and forth in order to figure out how *things are*. The writer asks whether the character really would throw the shoe, but also considers whether the story itself is best served by the character acting in this way. The writer could, of course, just include the character’s throwing the shoe and then see how the story fares as a result. The writer’s constraints from the external perspective are not rigid, at least in many cases, because he can always make adjustments later. The reader, by contrast, is strongly constrained by the external perspective, because by the time she is engaged in the discovery process, the writer has already decided how things will go.

²⁵ Ibid, p. 116.

²⁶ “The Figure a Poem Makes,” in *Collected Poems*, NY: Halcyon House, 1939.

There is a different context, not discussed by Gardner, in which the reader may be in a similar position to the writer's position during discovery. When she is trying to reflectively decide what action to perform in the real world, based on how this action will affect others, figuring out what mental states the other person is in and how he will react emotionally, she is in exactly the same position as the writer. She first has to place herself in the situation of the other, based on her inferences about what states the other person is already in, and then simulate the other, adopting *his* perspective, not hers, while also considering how the impact of her action will affect the overall goals she has for herself and others. Possessing an external perspective in real life, barring religious beliefs about a divine plan laid down by the grand "author," involves seeing the narrative of one's life and how one wants this narrative to go.²⁷ Thus a person toggles back and forth between relating to others as a principal player and viewing her own life from the perspective of the constructor of the narrative.

Gardner recognizes that readers are rarely in the same position as the writer in fictional contexts, but there is one kind of fictional work, he says, that does give the reader an experience similar to the writer's:

In this process I describe, the reader is at a disadvantage in that what he has before him is not all the possibilities entertained by the writer and recognized as wrong but only the story the writer eventually came to see as inevitable and right. But the good writer provides his reader, consciously and to some extent mechanically, with a dramatic equivalent of the intellectual process he himself went through. That equivalent is suspense.²⁸

²⁷ See Peter Goldie, *The Mess Inside: Narrative, Emotion, and the Mind*, Oxford: Oxford University Press, 2012 for an account of how we conceive of our selves through narratives.

²⁸ *On Moral Fiction*, p. 114.

There is moral and philosophical value in suspense, he says, because in this kind of work, the writer “presents the moral problem—the character’s admirable or unadmirable intent and the pressures of situation working for and against him [...] and rather than moving at once to the effect, one tortures the reader with alternative possibilities, translating to metaphor the alternatives the writer has himself considered.”²⁹ The resolution is delayed, and in this delay lies philosophical insight, since the behavior of a character “reflects not simply his nature but his nature as the embodiment of some particular theory of reality and the rejection, right or wrong, of other theories.”³⁰ Every character, even the minor ones, he says, has a “philosophical function.”³¹ For Gardner the cognitive value of suspense resides in that period between the set-up and the eventual resolution of the conflict.³² But the writer’s rhetorical purpose isn’t primarily to provide these insights, even if Gardner is correct that they follow from the delays. In suspense authors *must* keep motivations and intentions obscured in characters, in order to sustain the underdetermined nature of the narrative and to allow for misdirection. This is best seen in a particular kind of suspense: detective stories.

²⁹ Ibid, pp. 114-115.

³⁰ Ibid, p. 115.

³¹ Ibid, p. 115.

³² Apparently, there is also cognitive value found in narratives that present *no* final resolution to conflict. See Maja Djicik, Keith Oatley, & Mihnea C. Moldoveanu, “Opening the Closed Mind: The Effect of Exposure to Literature on the Need for Closure,” *Creativity Research Journal* Vol. 25, No. 2, 2013, pp. 149-154. In this study evidence was found suggesting that “reading a literary short story led to a significant decrease in participants’ self-reported need for cognitive closure [i.e., a need to reach a quick conclusion in decision-making and an aversion to ambiguity and confusion]” which results from the fact that when “one reads fictional literature, one is encouraged to simulate other minds, and is thereby released from concerns for urgency and permanence.” (p. 153) The researchers recognize that this might be merely a short-term change and that further studies need to be run to determine how likely long-term effects are to occur, but they speculate that “[i]t is likely that only when experiences of this kind accumulate to reach some critical mass would they lead to long-term changes of meta-cognitive habits,” (p. 153) which is consistent with the argument I have made about how living the life of the reader is more apt to *transform* audience members in the way necessary to address the moral problem discussed in Part I than the processes described by particularist theories.

3. Whodunit?

In detective stories, like Doyle's Holmes stories and Agatha Christie's tales of Miss Marple and Inspector Poirot, the (implied) narrator is always suspect. In some instances the reader is along for the ride with the detective, so she sees what the detective sees and nothing more. But in other cases, the reader is offered glimpses of certain events, suggesting that the narrator knows who committed the crime under investigation or at least knows more than he is letting on. If I were conversing with a real person who is telling me that he had witnessed a murder, I would become very impatient with him if he related the account in such a way that I'm never sure who the killer is. Suppose he gives me hints that it was perhaps John, the insurance agent, but it could also have been Benny the butler, but don't rule out Chrysta, the jilted daughter-in-law. At some point, I am likely to exclaim in exasperation, "For God's sake, just tell me who killed the old man!" But in detective stories, even though the reader is manipulated in the same way, she doesn't respond in the same way. The narrator describes or depicts a hand holding a revolver, aimed at a character who has stumbled on the truth—the reader doesn't scream at the book or the screen, "For God's sake, narrator, just look up at the face so that we know who it is!"

The reader doesn't always give the narrator a pass like this, however, since sometimes she feels unduly manipulated. Wayne Booth³³ discusses the possibility of the reader's *losing faith* in the narrator when this kind of thing happens, because she feels like she's being tricked. "If Jane Austen can tell us [in *Emma*] what Mrs. Weston is thinking," Booth asks, "why not what Frank Churchill and Jane Fairfax are thinking?" When we ask this question, though, we have

³³ *The Rhetoric of Fiction*, Chicago: University of Chicago Press, 1983.

shifted from the internal perspective, treating the account as a non-fictional relation of actual events, to the external perspective, where the answer to the question is “because she chooses to build a mystery, and to do so she must refuse, arbitrarily and obtrusively, to grant the privilege of an inside view to characters whose minds would reveal too much.”³⁴ Indeed, Austen must keep the reader in doubt about some things for a long while, though she also wants the reader to have some knowledge that Emma herself doesn’t have—creating a tension between “maintaining some sense of mystery as long as she can” and working to “heighten [...] dramatic irony, usually in the form of a contrast between what Emma knows and what the reader knows.”³⁵ Booth suggests that, when the mystery is no longer relevant, in subsequent readings, “we discover new intensities of dramatic irony resulting from the complete loss of mystery” because we know the “abysses of error Emma is preparing for herself.”³⁶

This interplay of suspense and mystery and dramatic irony suggests that readers relate to narratives in two very different ways. On the one hand, readers of mysteries and suspense narratives are granted incomplete access to characters for the sake of the mystery. On the other hand, readers are often granted a wider access than the detective and the other characters, for the sake of suspense—i.e., the affective state of being uncertain of what is happening and/or who is doing it. Consider how different viewing from the internal perspective the fictional scenes in suspense and mystery is from viewing situations in real life. In general the reader enjoys a vantage point in many fictional scenes that would rarely be afforded in real life, but in suspense, the vantage points provided by the narrator would *never* be possible in real life. We allow this feature to pass generally unnoticed because we accept the terms of the work based on the external perspective. The internal perspective we adopt in these works is continually tempered by

³⁴ Ibid, p. 254.

³⁵ Ibid, p. 255.

³⁶ Ibid, p. 255.

facts that we accept from within the external perspective. This is less a matter of toggling back and forth between perspectives and more like an equilibrium reached between the two. From this point balanced between the two different perspectives, the reader can make evaluations of the decisions and judgments made by the characters, while also taking due account of the author's decisions and judgments. When the reader makes these evaluations from this blended point of view, we can call her perspective *critical equilibrium*.

Critical equilibrium allows the reader to see the fictional situation *at once* as having some of the same features as real life situations and as reflecting authorial goals and intentions. It's a quite powerful resource to have available, because when she maintains critical equilibrium, the reader will be able to grasp the predicaments of characters in a deeper way, and better understand the nature of the fictional *conflict* as it unfolds on the page or screen. Conflict sits at the center of most narratives, motivates the plot, gives characters opportunities to act and react to one another, and moves through a natural cycle of birth, climax, and resolution. Longer narratives may feature multiple conflicts, such as war narratives, like the novel *All Quiet on the Western Front* and the film *Full Metal Jacket*, or one sustained grand conflict accompanied by several subsidiary conflicts, as in the novel *The Magic Mountain* or the television series *MASH*. Conflict doesn't have to be *negative*, taking the form of combat or disagreement. In general terms it manifests itself through antagonistic forces marshaled against the protagonist(s), presenting resistance, which could exist in situations marked by excitement or happiness, such as planning a vacation. Challenges will accompany this activity, but these challenges won't necessarily be negative and the planner won't necessarily fail to meet them. This would still count as conflict.

More often than not, however, fictional conflict is negative. Characters are forced to deal with tragedy, betrayal, oppression, deceit, or overwhelmingly long odds of succeeding. And

since it's situations like these that are most commonly the same kind found in deliberative and moral reasoning, I will focus on fictional narrative artworks featuring negative conflict.

In mysteries and suspense the conflict is almost always negative. Moreover, the reader is led to feel a kind of anxiety, a necessary component to enjoying the work *qua* suspense. One could, as Booth suggests, re-read or re-watch a work of suspense, and appreciate some of the other elements or take a meta-critical stance and see how it operates as a suspense, all the while not feeling any negative affect as a result, but then one doesn't really experience it *as* suspense. Thus one condition for effective fictional conflict—in suspense in particular, but generally in all fictional works—is plausibility from the internal perspective. The reader must believe, or make-believe, that the characters really are facing the depicted challenges in order to feel the associated state of anxiety. If the conflict seems manufactured, the reader can certainly recognize why the conflict is deemed necessary, from the external perspective, but from the internal perspective, the purported conflict simply won't be thought of as genuine conflict by the reader. A second condition is that the conflict be substantial. While it's possible to motivate a narrative with conflict that might not appear substantial on first glance—such as planning a party, in Virginia Woolf's *Mrs. Dalloway*—most of the time even these instances turn out to be more than minor. This isn't to say that there cannot be narratives that have no, or very thin, conflict. These are works that are less plot-driven and more focused on character interactions, the paradigm example being the film *My Dinner with Andre*, in which Wally Shawn has dinner with Andre Gregory and the two engage in a lengthy conversation. But when a narrative clearly places conflict in a major role, the conflict must be recognized by the reader as substantial enough; otherwise, the role given to the conflict will seem unjustified. Finally, for conflict to matter to the reader, it must involve at least one character, and usually more than one, for whom the reader deeply cares. One

reason for this is that the reader won't have any affective response to a putative conflict that involves no sympathetic characters. The typical response to situations like this will not be emotional at all, but a cold recognition that certain events are shaping up.³⁷ A second reason is that it's difficult to recognize conflict *as such* in fictional situations that don't involve characters for whom one cares. If a character is faced with the possibility of losing his job, and he sets out to achieve several accomplishments in order to guard against his termination, the reader may fail to see this as conflict, if the character has not earned some of the reader's concern. Again, it will seem less like conflict and more like simple narrative events. Of course, the reader could, from the external perspective, recognize what is *supposed to be* conflict, and perhaps even follow the narrative understanding what is intended, but the affective response won't be there, and thus the purported conflict doesn't secure what genuine conflict would.

These general points apply across the board in narratives motivated by conflict. Detective stories and mysteries feature a kind of conflict that draws most strongly on the reader's ability to mindread. When the reader herself is involved in trying to gather clues that would resolve the conflict, she is a player, though in an indirect way. For example, she may set herself the task of finding out whodunit long before the solution is revealed. Or she may decide to focus closely on every clue, every piece of evidence, and determine whether the author has engaged in illegitimate misdirection, such as having a trustworthy character claim that he had dusted for fingerprints when he really hadn't.³⁸ In examples of mysteries, therefore, the reader engages in

³⁷ This is not to say that readers must positively regard the characters for whom they care. They can warmly respond to a conflict involving a character for whom they have negative feelings, but in this case, they may hope that the character emerges from the conflict having received his comeuppance. This is still deeply caring for the character, since they are invested emotionally in what happens. While all of this is true, it's still the case that in *most* cases, readers will hold the character in a positive regard.

³⁸ One real world example of this occurred as audiences viewed Alfred Hitchcock's *Stagefright* for the first time. Viewers naturally assumed that flashbacks are *factive*—what is represented in these scenes is true, since the filmmaker, or implied narrator, is presumed to be relaying narrative content to us in the flashback. In Hitchcock's film, however, a flashback turns out to be false, even intentionally deceptive. The suspense of the narrative relies on

mindreading in order to do her part to resolve the conflict. As such the mindreading here is deliberative, rather than heuristic.

Lisa Zunshine³⁹ argues that mysteries and detective stories give the reader a “work out” in mindreading. Because these stories call on our “ability to store representations under advisement and to reevaluate their truth-value once more information comes in,” they push mindreading to “its furthest limits [...] by explicitly requiring us to store a lot of information under a very strong advisement—that is, to ‘suspect *everybody*’—for as long as we can possibly take it” and finally “to readjust drastically much of what we have been surmising in the process of reading it.”⁴⁰ This is true as far as it goes, but we should also recognize, as Zunshine doesn’t, that *any* narrative with the same *kind* of conflict gives a similar workout to our mindreading. Granted, we aren’t engaged in mindreading every character in stories other than mysteries, but we do often mindread as part of understanding the conflict and *explaining why* the conflict exists to begin with. Moreover, since the reader is in a virtual space, as discussed before, and because she occupies the point of critical equilibrium, she also has pieces of knowledge—not only about the outer and inner lives of the characters, but also about some of the author’s purposes—that she can draw from in constructing interpretations of character behavior, thus getting some clarity about what must occur to resolve the conflict. Indeed, it’s being in this position that allows the reader to aesthetically *appreciate* the nature of the fictional conflict. Thus, as the reader engages with fictional conflict in many kinds of narratives, she is thereby engaging in deliberative

the viewer’s assumption that the flashback is factive, when it’s not. Audiences felt that their trust had been violated because of the inversion of convention. But then too Hitchcock might have responded that the assumption about flashbacks wasn’t justified and that his use of it showed as much.

³⁹ *Why We Read Fiction: Theory of Mind and the Novel*, Columbus, OH: Ohio State University Press, 2006.

⁴⁰ *Ibid.*, pp. 123-124.

mindreading. And as Gardner points out, one of the things that is there to be realized by someone undergoing this process is that doing so “leads to *knowledge, sensitivity, and compassion.*”⁴¹

While recognizing that reading mysteries and detective stories gives our mindreading a workout, Zunshine further claims that this workout has very few practical advantages. She analogizes the workout readers get in detective stories to bodybuilding, and, “just as not everybody is an avid bodybuilder,” she says, “so also not everybody is an avid detective-novel reader or is even remotely interested in detective narratives.” But take comfort, she suggests, because just as bodybuilding is not necessary to “keep our muscles from atrophying,” since we “still get enough indirect exercise from our everyday activities,” people don’t *have to* read detective stories—“or even much of any fiction”—because they “still get plenty of interaction with our environment to keep our [mindreading] capacity ‘in shape.’”⁴² And even if there are benefits for our mindreading capabilities resulting from reading detective stories, again just like bodybuilding, the benefits are superfluous. She writes

Just as overdeveloping one’s triceps, biceps, and trapezoids generally does not give the bodybuilder any particular advantage in her everyday activities—it certainly does not make one more adept at handling such crucial items as a pen, a laptop, a phone, and a fork—so keeping on a steady diet of detective stories does not make one a particularly discerning social player. It does not help me see through somebody’s lies and it does not help me to know which ‘clues’ to pay attention to in order to get to the truth of a given matter. [...] [Detective stories] stimulate [our mindreading capability] without providing the kind of ‘educational’ benefit that we still implicitly look for in what we read. Delight they do, but instruct they don’t, or at least not in the traditional sense of the word *instruction.*⁴³

⁴¹ *On Moral Fiction*, pp. 113-114, my emphasis.

⁴² *Why We Read Fiction*, pp. 124-125.

⁴³ *Ibid.*, p. 125, emphasis in original.

There are several questionable points in Zunshine's argument, however. As we saw in chapter 3, our mindreading capacity is very much not "in shape." The psychological studies showing how mindreading attempts are inaccurate more often than they're accurate should persuade us that her claim lacks empirical foundation. But since this is the case, bodybuilding is not an apt metaphor. Zunshine's presumption when she uses this metaphor is that, just as people are perfectly able to make their ways in the world given their natural muscle mass and only choose to engage in bodybuilding in order to develop more muscle for non-practical purposes, people are also already perfectly able to gain accurate determinations of the thoughts and feelings of others. Thus, when they read fiction, any improvement that occurs as a result is superfluous. While it may be true that in general people "get by" using their flawed mindreading capabilities, it's not true that when it comes to deliberative situations, they are able to get by in the same way. This is why bodybuilding is not the best metaphor, while *weightlifting* and/or *athletic training* would be better. Could the Olympic athlete get up in the morning, drive to work, tote his groceries from the store to his car, cut his lawn, without undergoing any of the training for the Olympic games? Sure. Could this athlete be competitive at the Olympics without training? Absolutely not. Thus, in the context of deliberative mindreading, when the soundness of one's decision relies crucially on getting the mental states of others correct—that is, contexts like moral reasoning—improving our naturally flawed capacity is as imperative to the moral agent as training is to the Olympian. Zunshine's argument here is therefore incomplete and overlooks a vital way that mindreading in fictional contexts can benefit readers in their everyday lives.

Based on the discussion in the last two sections we can conclude that the writer and the reader who confronts fictional conflict *as* conflict, while occupying the perspective of critical equilibrium, both stand to become better deliberative mindreaders, especially as these contexts

require the use of transformation simulation, the de-centering on one's own perspective and the re-centering on the perspective of characters involved in the conflict. We now must see, in addition to such improvement, how the reader may be more *motivated* to engage in deliberative mindreading in real situations. In order to get started thinking about this, let's consider an example.

4. A Notorious Example of Failed Mindreading: *Notorious*

Alfred Hitchcock's films have been hailed for decades as paradigms of suspense, but there are at least two different *kinds* of suspense in his films that accomplish two very different goals. The first kind of suspense places the viewer in the scene with one of the characters and effects a kind of viewer identification with the character. In this *identificatory suspense* the viewer experiences affective states on the part of the character, either concomitantly or vicariously. In concomitant identificatory suspense the viewer is granted the epistemic perspective of the identified character, and as the character experiences anxiety, worry, and dread, so does the viewer and for roughly the same reasons. An example of this kind of suspense is *The Birds*, in which Melanie (Tippi Hedren) is completely unable to explain the phenomenon of the aggressive nature of the birds attacking the residents of the town and does what she can to escape the vicious attacks; as she does so, the viewer is generally in the same epistemic state as Melanie's. In vicarious identificatory suspense the viewer has more knowledge than the identified character and the knowledge that she has creates the anxiety, worry, and dread that the character *would* feel if he had the knowledge. Examples of this are *Psycho*, in which Marion

(Janet Leigh) is murdered in the famous shower scene, by Norman Bates (Anthony Perkins) who approaches the shower without Marion's knowing it; and *Rear Window*, in which the character Lisa is sneaking into suspected murderer Lars Thorwald's (Raymond Burr) apartment, while he is supposed to be out, and Thorwald returns early. In the latter the viewer sees Thorwald return unexpectedly early alongside the character Jeff (Jimmy Stewart), who watches from his apartment across the courtyard while bound to a wheelchair. The viewer experiences vicarious *and* concomitant worry—vicarious, because it's on behalf of Lisa, and concomitant, because it's the same worry that Jeff feels on Lisa's behalf. Identificatory suspense is located primarily within the internal perspective, since toggling back to the external perspective would render the worry inert—no real person is in any mortal danger.

A second kind of suspense is tied to the dramatic irony mentioned by Booth earlier. This *detached suspense* occurs when the viewer has knowledge that certain characters don't, as in vicarious identificatory suspense, but the knowledge, if it weren't lacking, wouldn't necessarily create anxiety in the characters. In this kind of suspense the viewer experiences uncertainty *because* the knowledge is missing and not by virtue of the *content* of the missing knowledge. In order to see the difference between identificatory and detached suspense, consider the film *Rope*.

In *Rope* both kinds of suspense are at play, so it's crucial to tread carefully. The first scene shows David being strangled by Phillip (Farley Granger) and Brandon (John Dall). After this initial scene Phillip and Brandon hide the body in an antique chest, after which they host a dinner party, attended by David's father; his aunt; his fiancée, Janet; Janet's ex-boyfriend, Kenneth; and Rupert (Jimmy Stewart), the old high-school teacher of Phillip, Brandon, and David. The viewer watches as David's friends and family serve themselves from the chest in which the body is hidden, even as the guests wonder wherever David could be, since it's not like

him to be late without calling. Phillip *enjoys* the rising tension, as occasionally it looks like one of the unknowing friends or Phillip's maid, Mrs. Wilson (Edith Evanson), is going to open the chest for necessities for the party, while Brandon *obsesses over* and *dreads* the looming consequences of getting discovered. Thus Phillip and Brandon have knowledge the rest of the guests lack, and each has a different affective response to having the knowledge. The viewer undergoes two sets of affective responses to having the knowledge as well. First, she experiences vicarious dread on the part of David's loved ones. It would be a truly awful way to find out that one's son or fiancé has been murdered by discovering the body as one is eating food served from the very place the body has been hidden. This is vicarious identificatory suspense. But, second, the viewer has a kind of detached response, rooted primarily in the external perspective, in which she tends to enjoy, with Phillip, the close calls, but enjoys the experience in part because she feels the obsessive dread of Brandon as well. What she's enjoying is seeing how the scene plays out, as certain characters lack knowledge that she herself has. It's this second kind of suspense that I am calling detached suspense.

Detached suspense can result in insight, evident in cases in which a viewer realizes just what significance the missing knowledge bears in the scene, and further how crucial having the knowledge is to the nature of characters' actions. While it's questionable whether the knowledge lacked by David's loved ones is really valuable—would they be better off learning that David has been murdered and his body has been literally lying right under the food they are eating?—in other narratives we might see the value of missing knowledge by observing how dire the consequences are of lacking the knowledge. This occurs, I will argue, in Hitchcock's *Notorious*.

The plot of the film is as follows: American intelligence agent Devlin (Cary Grant) approaches the daughter of a German nationalist recently convicted of treason against the U.S.

and asks for her help in infiltrating a circle of her late father's friends in Brazil, whom the U.S. suspects of engaging in activities intended to avenge Germany's loss in World War II. The daughter, Alicia (Ingrid Bergman), is well-known for being a woman of loose morals, heavy drinking, and sleeping around, so Devlin has no compunction about suggesting to her that she might have to cross some moral lines in completing the job. She agrees to do the job, without having heard the specific details, but on their trip down to Brazil together, the two fall in love. Devlin seems willing to believe her when she says that her newfound love for him has changed her, but this belief is put to the test when Devlin discovers what her task in Brazil is going to be: getting an old friend of her father's, Alexander Sebastian, to fall in love with her in order to gain his trust and find out what the German conspirators are up to. Devlin's skepticism about Alicia's transformation gets the best of him, and, despite his defending her honor with his superiors when she is not present, as he lays out the details of the job for Alicia, he tries to conceal his true emotions and the fact that he defended her earlier, leading her to believe that he has no faith in her declarations of love. She thus takes the job, thinking that Devlin never really loved her, and ends up marrying Sebastian. She then finds out that Sebastian is part of a program to manufacture nuclear weapons to be used against the U.S. Along the way, Sebastian discovers that she is working for the U.S. and, with his mother's help, slowly poisons Alicia, until Devlin realizes that he's been a fool and comes to her rescue, getting her away from the house and to a hospital.

The viewer feels both identificatory and detached suspense in this film. The identificatory suspense occurs when Alicia's role is almost discovered by Sebastian, and then after the discovery. At first, the viewer knows, when Alicia doesn't, that she is being poisoned—vicarious identificatory suspense—and later, Alicia learns this, too, but by this time, she is too weak to do

anything about it, so she lies in bed and slowly approaches death—concomitant identificatory suspense. But the viewer also gets to experience detached suspense, since she knows, while Devlin and Alicia don't, that each of them wishes the other to step up and demand that Alicia reject the job. What the viewer gets to observe is utter failures of mindreading, which failures cause the *entirety* of the fictional conflict.

To see how this occurs, let's begin with the pivotal scene in which Devlin and Alicia suffer the mindreading failures that eventuate in Alicia's taking the job. Prior to this scene, Devlin and Alicia had declared their love to one another on the balcony of Alicia's hotel room, before finding out the details of the job. She tells him she's a changed woman and he lightheartedly replies with "Actions speak louder than words." He leaves her in the hotel to prepare dinner while he goes out to meet up with their boss, Paul Prescott, to learn the specifics of the assignment. He learns from Prescott that Alicia is to get Sebastian to fall in love with her, the implication being that she will have to become a sexual target for Sebastian in the process. Devlin objects when he learns of this, suggesting that Alicia may not be comfortable with such a task. Prescott laughs it off, saying that a woman like Alicia should have no trouble doing what is necessary. Devlin is clearly rattled by both Prescott's insinuation and the possibility that Prescott is right.

When he returns to the room, Devlin is different, acting coldly toward Alicia, but also showing worry and concern that the woman to whom he earlier declared his love might eagerly take up the assignment. Alicia is still enjoying the afterglow of their earlier conversation as he relates to her the object of their assignment, trying to hide his real feelings, already revealed to the viewer in the scene with Prescott. Alicia responds by saying that she knows Sebastian, but was never interested in him in that way. The following exchange then occurs:

Alicia: I suppose you knew about this pretty little job of mine all the time.

Devlin: No, I only just found out about it.

A: Did you say anything? I mean, that maybe I wasn't the girl for such shenanigans.

D; I figured that was up to you, if you'd care to back out. [With a facial expression of hope, which Alicia misinterprets as skepticism]

A: I suppose you told them Alicia Huberman will have this Sebastian eating out of her hand in a couple of weeks. She's good at that. Always was.

D: I didn't say anything.

A: Not a word for that little lovesick lady you left an hour ago?

D: I told you, that's the assignment.

A: Well, now, don't get sore, Dev. I'm only fishing for a little birdcall from my dream man. One little remark such as, 'How dare you gentlemen suggest that Alicia Huberman, the *new* Miss Huberman, be submitted to so ugly a fate?'

D: That's not funny.

A: Do you want me to take the job?

D: You're answering for yourself.

A: I'm asking you.

D: It's up to you.

A: Not a peep? Darling, what you didn't tell them, tell me. That you believe I'm nice, and that I love you, and I'll never change back.

D: I'm waiting for your answer.

A [turning away, with a sad smile]: "What a little pal you are. Never believing me. Not a word of faith. Just down the drain with Alicia. That's where she belongs. ... Oh, Dev. Dev ... " [gets a drink] ... "When do I go to work for Uncle Sam?

D: Tomorrow morning.

It's clear to the viewer, of course, that Devlin is concealing his true emotions, since she sees evidence of his emotions in his apparent ambivalence in the scene with Prescott, but what the viewer knows is that it isn't ambivalence, but deep worry, because of what was also revealed in the balcony scene. Alicia doesn't have the same information, though she is practically begging him to tell her that this is how he responded when he first heard of the assignment. Still, Devlin's facial expressions and way of speaking clearly indicate that, while he says that it's her choice, he desperately hopes that she will turn the job down, as it may be appropriate for the woman she used to be, but the *new* Alicia has no business with such things. Alicia doesn't read this in his behavior, however, interpreting his reserved manner as a complete lack of faith in her transformation. She responds to her interpretation of his behavior with mock amusement, which, juxtaposed with her sincere pleas that he tell her not to take the job, leads Devlin to interpret her as being rather comfortable with the assignment. Her interpretation of Devlin's behavior persuades her to believe he doesn't really love her, while his interpretation of her persuades him to believe that she doesn't really love him. These beliefs then work to feed the respective interpretations until each of them has wildly inaccurate judgments of the other. These inaccurate judgments culminate in her taking the job and committing herself to fully convincing Sebastian that she is romantically interested in him, ultimately resulting in their getting married.

The viewer has a close up view of this portrait of failed mindreading and the dire consequences that follow. Other similar scenes follow. One scene, in which Devlin and Alicia meet for her to tell him what she has learned, occurs at the racetrack, as Sebastian and his mother watch the horses from their luxury box, and Devlin and Alicia find one another in the crowd below. She tells him what she has learned, then adds a further detail: "You can add Sebastian's name to my list of playmates." Devlin responds, "Pretty fast work, wasn't it?" to which she

sharply retorts, “That’s what you wanted, wasn’t it?” It’s here where the viewer gets to see how the two have had time to reflect on that earlier pivotal scene, and also where they reveal some of the content of their mental states in that earlier scene, content to which each of them had inadequate access at the time. As the following exchange shows, too, some bridges might have already been burned, despite what they learn about each other:

Devlin: I can't help recalling some of your remarks. About being a new woman. Daisies and buttercups, wasn't it?

Alicia: You idiot. What are you sore about? You knew very well what I was doing.

D: Did I?

A: You could have stopped me with one word. But no, you wouldn't. You threw me at him.

D: I threw you at nobody.

A: Didn't you tell me to go ahead?

D: A man doesn't tell a woman what to do. She tells herself. You almost had me believing in that little hokey-pokey miracle of yours ... that a woman like you could ever change her spots.

A: Oh, you're rotten.

D: That's why I didn't stop you. The answer had to come from you.

A: I see. Some kind of love test.

D: That's right.

A: You never believed in me anyway, so what's the difference?

D: Lucky for both of us I didn't. It wouldn't have been pretty, if I'd believed in you. If I'd figured, “She'd never be able to go through with this. She's been made over by love.”

A: If you only once had said that you loved me. Oh, Dev.

D: Listen. You chalked up another boyfriend, that's all. No harm done.

A: I hate you.

D: There's no occasion to. You're doing good work. ... Number ten's out in front. Looks as if Sebastian knows how to pick 'em.

A: Is that all you have to say to me?

D: Dry your eyes, baby. It's out of character. ... Except, keep on your toes. It's a tough job we're on.

She reveals to him that earlier she just needed him to tell her that he believed in her, but this might have been revealed to her had she been more attuned to the subtleties of his behavior, for he certainly did believe that “she’d never be able to go through with this.” He reveals to her that he wanted her to reject the job, not because *he* wanted her to, but because *she* was no longer the type of woman who would take that kind of job. In the earlier scene, however, he wasn’t able to communicate that to her, because he had allowed his worry that she wouldn’t reject the job to prevent his seeing that she had absolutely no desire in taking it. Now, however, it may be too late, since she has forged ahead and has (presumably) slept with Sebastian.

There is one last opportunity for either of them to call a stop to the process, in a scene where Alicia tells Devlin and Prescott that Sebastian has proposed marriage. Prior to Alicia’s entering the room, Devlin responds to another man’s insinuations about Alicia’s character by defending her, showing he is still in love with her and still has faith that she is not relishing the assignment. When she enters the room, telling them of the proposal, he tempers his behavior, again suggesting to Alicia that he really is not affected by what she is doing. It’s clear to the viewer, however, that by Alicia’s telling them all this, she intends to elicit from Devlin a final refusal that she be on this assignment. After she informs them of the proposal, Devlin too tries to provide Alicia an out, by suggesting that marrying Sebastian will present too long a delay for their plans, with the implication that the *honeymoon* will keep her away from the house for too long. She doesn’t read this as a way out, however, interpreting Devlin as merely raising a tactical worry. Since it’s just a tactical worry, she responds by saying that she can get back to the house fairly soon after the honeymoon, the mere mention of the word causing Devlin to leave the room.

Again, she fails to make the appropriate inference based on Devlin's behavior, interpreting him to have been satisfied by her solution. She thus goes on to marry Sebastian.

The viewer is never sure whether or when Devlin and Alicia will realize that they have completely misinterpreted the behavior of the other. The viewer hopes that they will make the realization before Alicia takes the job, but it doesn't happen; she hopes it happens before Alicia sleeps with Sebastian, but again it doesn't happen; she hopes it happens prior to the marriage, and it doesn't happen. In the end Alicia's life depends on this realization and just when it looks like it's not going to happen, it does. What the sensitive viewer finally realizes is that the *entire* fictional conflict emerges from *utter mindreading failures*, due in part to bad *mindreading*, in part to bad *mindwriting*. What she also realizes is how better mindreading and/or mindwriting could have forestalled the looming catastrophe resulting in Alicia's nearly dying. Admittedly, this is an extreme case, but the extreme case is often necessary to show even the sensitive viewer something that less drastic cases cannot show: the value of mindreading in preventing or eliminating conflict altogether.

5. Moved to Mindread

Gardner claims that there can be philosophical insight in suspense due to the possibility of having laid out before the reader/viewer several different characters each of whom holds "some particular theory of reality" and then testing whether any or all of these theories should be rejected or retained. I have suggested a different way that suspense can provide such insight—by showing how valuable mindreading is to the prevention and resolution of conflict, in part, by

showing how mindreading *failures* can be directly responsible for the emergence of such conflicts to begin with. Just as the insight Gardner describes can be extended from the fictional context to the real world, the insight I describe gained from a work like *Notorious* can be so extended as well. Once a viewer of this film, or many other similar films, or a reader of similar literary narratives discovers the practical and moral value of mindreading in fictional situations, she can start to see how mindreading plays a similar role in real situations.

First, consider the motivations of a viewer of a film like *Notorious* who enjoys the detached suspense resulting from observing dramatic irony in the relationship between Devlin and Alicia. As I said before, at one level, from the internal perspective, the viewer wants each of them to learn of their mistake as soon as possible—either before she takes the job, or before she sleeps with Sebastian, or before she marries him. The viewer cares about the two characters and doesn't want their relationship to crumble. But at a different level, from the external perspective, the viewer says, "I want them to realize their mistake . . . eventually, but not yet. I'm enjoying this too much for it to end so soon." This is, of course, an egoistic way of seeing the conflict, but from the external perspective, there are no moral implications, since the characters aren't real people and the pain they feel in the fictional conflict isn't real either. If, however, the viewer knew two real people who were in the same kind of situation, she would be more likely to desire that the conflict be resolved *as soon as possible*; and if she *herself* were in the situation, she would almost surely desire this. Thus seeing the role of mindreading in the emergence of many fictional conflicts and also seeing its value in preventing and resolving such conflict can result in a viewer's being more motivated to mindread—and mindread *accurately*—in her own life. Of course, we are often guilty of inaccurate mindreading in our own lives, while not being aware of it, but because the reader/viewer has both gained metaknowledge of her mindreading abilities, as

discussed in chapter 9, and because she is comfortable occupying the position of critical equilibrium, she will also be able to imaginatively step outside the real conflict of which she is a part and reflect on the details much as she would from the external perspective in fictional conflicts. Not only is she more motivated to mindread accurately, she is also better prepared to know when she has mindread inaccurately and she knows how best to address the underlying phenomena that cause her to do so, both essential components to becoming a *better* mindreader as discussed in chapter 3. Therefore, along with the reader's increased motivation to mindread accurately, if she is an aesthetic adult and has made engagement with fictional narrative artworks a part of her life, she also has had the relevant experience making her more likely to *be* an accurate mindreader.

Chapter 11

Seduced into Caring: Virtue and the Life of the Reader

This dissertation might be analogized to a romantic film of 1940's Hollywood. In Part I we encountered a moral problem: people suffer under a severe limitation to care for others. The opening credits and first few scenes therefore give us a picture of a moral agent ill-equipped to be good—a lonely character, in need of a companion. I gave an explanation of this moral problem, based in part on the kind of motivation virtuous agents should possess—compassion, understood as empathetic concern—and in part on psychological evidence showing that most of us suffer from low empathic accuracy. This has the significant implication that many of us are unable to identify what is good for those around us, and because of this inability, we are also unable to properly weight others' interests. An argument was provided at the end of Part I, in chapter 3, however, that the problem might be solved, since those working on the issue of empathic accuracy also give indications of how we can improve our empathic accuracy, thus making us better empathizers and therefore more likely to possess the right kind of motive. I suggested there that aesthetic experience stands to improve us in the necessary ways to improve our empathic accuracy. There's hope for the disconsolate character, but we need to see, from the cast of characters who line up like potential suitors, which one, if any, represents the path to goodness.

In Part II we saw the basic conceptual framework surrounding the question of whether art and ethics overlap in a meaningful way. This framework was offered so that it could be seen that many of the most prominent theories of how aesthetic experience can improve us morally are unable to provide a solution to the problem in Part I, because they are particularistic in nature,

and also, due to their particularist nature, that most of these theories show only that aesthetic experience can have *pre-ethical*, rather than *ethical*, effects. I concluded at the end of Part II that these theories should be rejected, at least for our purposes. None of the would-be suitors for the lovelorn character succeeds in leading the character to the right kind of salvation.¹

In Part III a different suitor appears, one who adopts a different approach from the immediate and direct ones found in the particularist theories. In chapter 7 I gave descriptions of the aesthetic adult and a life this individual might live called the *life of the reader*. My decision to begin my account in this way was a result of our seeing how particularist views, focusing on individual works and experiences, lacked the proper scope, which should be expanded beyond individual experiences in order to reflect what kinds of effects are produced by aesthetic experiences found within the context of life as lived, allowing for aggregative effects that happen over time and as a result of multiple aesthetic experiences, as well as interactions between these experiences and those found in other parts of our lives. When we expanded the scope in this way, we were able to see in chapters 8, 9, and 10 how improvement in those key areas discussed at the end of chapter 3 are initiated and maintained as a result of living the life of the reader. By the end of chapter 10 the part of my account showing how living such a life improves our ability to *empathize* was complete.

This is only part of the solution, however. Our sad character is intrigued by the newly arrived suitor, who has been able to make significant progress already, but the suitor has to make one last pitch to convince our character that genuine romance is within reach. It's in this last scene, presented in this last chapter, that we can see the completion of what has turned out to be a *seduction*. When we remind ourselves of the primary problem—serious limitations in

¹ I say “the right kind of salvation” so that it's clear that I have not argued that we *cannot* become better moral agents along the lines argued by those theorists discussed in chapters 5 and 6.

sympathy and caring—we can see that an improvement in empathizing is only part of the solution. Insofar as the deficit in caring is caused by low levels of empathic accuracy, much of the problem is addressed; nevertheless, it's still possible for a moral agent to improve empathic accuracy, yet still fail to care for those around her. In short, she might possess the empathy-component of compassion, but not the concern-component, in which case, my account, as those discussed in chapter 5 and 6, would also show merely that aesthetic experience can produce pre-ethical effects. In this last chapter I intend to demonstrate that living the life of the reader can increase one's desire to show concern for others. Given the effects discussed in chapters 8, 9, and 10, the reader extends not just concern, but empathetic concern—that is, concern backed by a warranted belief about what is good for those around her. The emergence of empathetic concern constitutes a genuine ethical effect. We will see that in the end the reader is seduced, just as the lonely character in the old romances of the 1940's and 1950's is. But she is seduced not into romance, but into caring for others, by being drawn into a complex of cognitive, affective, and moral adjustments, not for the sake of any further goals, like becoming a better person, but because doing so is necessary for aesthetic enjoyment. And aesthetic enjoyment is its own reward, even if there are further rewards to be gained. Let's see how this last scene unfolds.

1. Moved to Care: Philosophical Evidence

There are two ways that it might be shown that we are moved to care for real people around us as a result of living the life of the reader. The philosophical approach would give a description of how we relate to fictional narrative artworks and the ways this relation might

change us in significant ways after a period of time in order to increase our capacity and desire to have concern for others. The psychological approach would provide evidence that people who engage with artworks do as a matter of fact show more concern for real people as a result of the aesthetic engagement. In this section I will provide my own philosophical argument for why aesthetic adults are moved to care by living the life of the reader, while in the next section I will present some of the psychological evidence suggesting that audience members are in fact moved to care by aesthetic experience.

Let's think again about the film *Notorious* and how viewers might be affected by the depicted relationship. In the last chapter we saw how the sensitive viewer might pick up on the fact that Devlin and Alicia are in the situation they are in because they have failed to mindread accurately—in part, because they allow their own insecurities to obscure their mindreading attempts, in part, because these insecurities affect their ability to mindwrite. As noted before the viewer can experience these failures from both the internal and external perspectives, thus saying on the one hand, “Oh, poor Devlin and Alicia ... someone must step in and tell them they are being fools!” while, on the other hand, saying, “But this is too good to end now! Let these characters continue to toil in ignorance of the other's feelings!”

We'll see in the next section that there is evidence that when audience members are drawn into narratives in the right way, they end up caring for fictional characters, but we've already been presented with evidence in chapter 9 that audience members often cognitively and affectively process the actions of fictional characters in the same ways they process the actions of real people. With a film like *Notorious* the viewer is drawn in, which thus prompts him to care for the characters, and, if the evidence in chapter 9 is correct, to observe the scenes with Devlin and Alicia in ways very similar to how he would react if he was watching the same events

occurring in the real world. Let's assume that the viewer is in what I called in chapter 10 critical equilibrium—that is, he is able to see the film from *both* the internal and external perspectives at once. When he makes the above comment from the internal perspective, he genuinely cares for the characters, because it would be difficult to adopt this perspective without caring for these characters. When he makes the above comment from the external perspective, he is thinking about how pleasurable the experience is and how the fictional characters need to continue to suffer for his enjoyment. This is the case because, from the external perspective, he can see that the fictional characters are *not real*, which means that there is no reason to feel any responsibility to address their suffering, which itself is not real. When he approaches Devlin and Alicia's relationship from the position of critical equilibrium, however, he both *cares* for the characters and judges that this caring doesn't impose any moral obligations on him. Why? Because *as he cares* for these characters, he recognizes that those for whom he cares are not real people and have no claims upon him. In other words, in the fictional context, caring for the characters is the path toward determining that the characters represent no real moral obligations. He *acknowledges* the suffering, in the way that John Gibson describes acknowledgement, but in acknowledging their suffering, he recognizes that the suffering occurs in a world—the story—of which he is not a part.

As I argued in chapter 10 the reader can occupy the position of critical equilibrium in real life. She is able to approach real events occurring around her both from the internal perspective—indeed, she is a player in these situations—and from the external perspective, because she can step back, take an *impersonal*² view of the situation, see the dynamic relations between those involved in the situation, and make a reasoned judgment of what should occur, in a way very similar to how John Gardner described the process of the working writer in chapter

² See chapter 1, section 3 for a discussion of the distinction between “impersonal” and “impartial.”

10. Once the reader has adopted critical equilibrium in both fictional and real contexts as a matter of course, this stance toward moral situations becomes habit. She just does it whenever the situation calls for it.

Part of occupying critical equilibrium *just is* determining the interests of those around us and taking a stake in whether those interests are served. Again, as I have said before, this isn't equivalent to altruism, since in some situations, the morally appropriate action will be to act, not *out of* the interests of others, but *against* them.³ This is very different, however, from acting *in ignorance* of the interests of others. Being able to approach a situation of which one is a part from the internal and external perspectives at once gets us toward acting from Murdochian love, as described in chapter 1, where I argued that Murdochian love is best understood as an attitude toward oneself and one's own interests—*viz.*, that one's own interests shouldn't be prioritized simply because they are one's own. We now have a way that moral agents can actualize the state of Murdochian love—by occupying the position of critical equilibrium in real life, just as the reader does in aesthetic contexts.

It might seem that I face the same problem facing informational particularist theories discussed in chapter 5—the problem of the missing desire. Since I am claiming here that the reader can be moved to care from aesthetic experience, because the possibility of a change in motivation centrally involves a change in desires, I too would have to confront the same problem.

It's true that the same challenge exists, but it's no longer properly described as a problem. We should remember that the reason the informational particularist theories were presented with

³ These will often be examples of what Rosalind Hursthouse (*On Virtue Ethics*, NY: Oxford University Press, 1999, chapter 3) calls tragic dilemmas, in which, even after an agent has done the good or right thing, there is a “remainder” about which the agent may feel remorse or regret.

this problem was that they were purely informational and that they were particularist. It's difficult to see how the learning of new information, in one experience, can produce a new desire that isn't itself the only ethical effect. We aren't committed here, however, either to the claim that the change is purely informational or to a particularist focus. Because we have expanded the scope, we no longer have to limit ourselves to informational or formational changes. Indeed, most of the changes we've been discussing occur over a period of time, through the process of habituation, to such an extent that it might be impossible to isolate the individual changes that are responsible for the overall transformation, even if I have spoken in Part III as if we can make these fine-grained determinations. Nevertheless, when the reader has undergone the changes described in chapters 8, 9, and 10, and is able to occupy critical equilibrium, she now has the capacity to care for those around her in ways that others might not have. New desires can be produced by aesthetic experience, but pinpointing just when these desires are born isn't an obligation for a non-particularist account. All that is necessary is to show that the desires eventually come into being as a result of living the life of which aesthetic experience is a crucial part.

Another worry someone might have with the present account is that there are persuasive counterexamples showing that living a life in which aesthetic experience plays a crucial part demonstrably does *not* move a person to care. Two examples are often given: First, many of the upper echelon officers in the Nazi government—figures directly responsible for devising and implementing the efforts to commit genocide against the Jewish people—possessed a passion for the arts, yet had no qualms in ordering soldiers to march millions of human beings to concentration camps and gas chambers. Second, there is an oft-repeated account in which 19th century playgoers, after watching tragedies depicting the woes of the poor and underprivileged

and shedding copious tears for the plights of the characters, leave the theater and literally step over the poverty-stricken beggars in the streets outside the theater, not even seeing the real poor before them, much less shedding a tear. While these examples are extreme, it isn't rare also to find people who are devout fans of novels or films who, when confronted with an opportunity to address the suffering of a real person, turn a blind eye to the suffering. In short, the objection here is that the account I have provided still doesn't show how the Simpleton described in chapter 5—the figure who sees suffering, but doesn't acknowledge it—can be transformed by aesthetic experience.

But we must keep in mind that the “life of the reader” is a term that appears deceptively simple, yet describes a rather complex set of conditions. Even if the Nazi officer and 19th century theatergoer are able to adopt critical equilibrium within the aesthetic context, what their behavior outside the aesthetic context demonstrates is that either they haven't been able to adopt the same perspective in real life or that, if they are so able, they are *so* morally corrupt that art cannot save them. I have never claimed that the morally evil person can be turned good. It is challenging enough, as we've seen, to show how the *average* person—one who has moral blindspots but isn't corrupt—can be transformed into a more caring moral agent. Regarding the voracious reader and moviegoer, we have no evidence that he is able to adopt critical equilibrium even in the aesthetic context, much less in the real world. More than likely, these individuals come to art as a means to escape the world in which they live, engaging in what Iris Murdoch calls “personal fantasy”—“the tissue of self-aggrandizing and consoling wishes and dreams which prevents one from seeing what is there outside one.”⁴ We're not concerned with whether these individuals undergo any transformations.

⁴ *The Sovereignty of Good*, p. 59.

A third worry about claims made here that living the life of the reader can move someone to care more as a result of living this life is that we might be mistaking a common cause of both living this kind of life and caring for others for a causal relationship between the former and the latter. If it turned out that people choose to live this kind of life because of a natural propensity that also is responsible for their caring for those around them, this would tell us nothing about whether the life of the reader plays a causal role in an increase in caring. We would have, in effect, two causal chains with no intersecting links at the crucial point under discussion.

This is a legitimate worry, but responding to it will take us away from the philosophical approach to the issue of whether living the life of the reader moves one to care and toward the psychological approach, which is discussed in the next section.

2. Moved to Care: Psychological Evidence

The worry, described in a slightly different way, is that the reason we might find people who live the life of the reader more disposed to show concern for others is not what I have claimed—*viz.*, that there is a causal relation between living this kind of life and being a person who shows concern for others—but that those who are *already* disposed to show concern for others are drawn to fictional narrative artworks. If this is the case, then there isn't *one* causal relation between living the life of the reader and showing concern for others, but *two* causal relations: one between living the life of the reader and some unknown entity and a second one between showing concern for others and some unknown entity—and there is no guarantee that it's the same unknown entity in each case. In order to show that this causal relation genuinely

exists we would need evidence showing that, prior to exposure to the appropriate fictional narrative artworks, a person lacks some degree of caring, but after such exposure, they demonstrably acquire it.

There is a body of psychological research investigating whether and to what degree readers of fictional narrative take on a more empathetic and caring attitude toward others.⁵ Many of the studies explore to what extent readers of fictional narrative are “transported” into the storyworlds of the narrative and whether a reader’s being transported into the storyworld has any discernible effects on the reader after having read the narrative and, if there are such effects, whether they are lasting or short-lived. Evidence suggests that readers who regularly consume fictional narratives experience improvement of social and interpersonal understanding rooted in empathy, whereas readers who limit themselves to non-fictional, expository materials “may accrue a deficit in social skills as a result of removing themselves from the actual social world.”⁶ There is also evidence that when readers report being transported into a narrative as a result of emotionally identifying with one of the characters and one of the narrative’s main characters models prosocial helping behavior, the reader is more likely to demonstrate prosocial helping behavior afterward, prompting the researcher to conclude that “it is possible that this modeling also had an effect on helping behavior, independent of affective empathy.”⁷

A recently published paper suggests that being transported into fictional narratives can have lasting effects on readers’ attitudes toward others, especially evident in subject reports of having a more solid connection to members of groups the readers previously held in disfavor.

⁵ For a comprehensive review of this body of literature, see Raymond A. Mar & Keith Oatley, “The Function of Fiction is the Abstraction and Simulation of Social Experience.” *Perspectives on Psychological Science* Vol. 3, No. 3, 2008, pp. 173-192.

⁶ Raymond A. Mar, Keith Oatley, Jacob Hirsh, Jennifer dela Paz, Jordan B. Peterson, “Bookworms Versus Nerds: Exposure to Fiction Versus non-Fiction, Divergent Associations with Social Ability, and the Simulation of Fictional Social Worlds,” *Journal of Research in Personality* Vol. 40, 2006, p. 705.

⁷ Dan R. Johnson, “Transportation into a Story Increases Empathy, Prosocial Behavior, and Perceptual Bias Toward Fearful Expressions,” *Personality and Individual Differences* Vol. 52, 2012, p. 152.

The authors of this study⁸ found that readers of fictional works engage in a phenomenon called “experience-taking,” a means to being transported into the storyworld, and that, as a result of experience-taking, readers can undergo changes in attitude and behavior toward real persons bearing resemblances to fictional characters with whom they identify in the works. These changes in attitude and behavior, the researchers suggest, can be for the better or worse, depending on the nature of the experience shared by the reader; but the general point is that when readers are encouraged to identify with characters, they feel concern for these characters, and this concern is thus extended to real people resembling the characters.

The paper describes six studies. The first three were intended to test the hypothesis that “when experience-taking occurs, readers simulate the events of a narrative as though they were a particular character in the story world, adopting the character’s mindset and perspective as the story progresses rather than orienting themselves as an observer or evaluator of the character.”⁹ The second three were intended to determine whether this kind of experience-taking can result in durable changes in behavior and attitudes. The term “experience-taking,” defined as when “readers let go of key components of their own identity—such as their beliefs, memories, personality traits, and ingroup affiliations—and instead assume the identity of a protagonist, accepting the character’s decisions, outcomes, and reactions as their own,”¹⁰ is contrasted with (a) feeling as though one is merely observing the character’s actions, like a fly on the wall, and with (b) what they call “perspective-taking,” defined as “first anchoring on one’s own perceptions or judgments and adjusting away from the self to surmise the other’s experience.”¹¹

⁸ G. F. Kaufman & L. K. Libby, “Changing Beliefs and Behavior Through Experience-Taking,” *Journal of Personality and Social Psychology* Vol. 103, No. 1, 2012, pp. 1-19.

⁹ Ibid, p. 2.

¹⁰ Ibid, p. 2.

¹¹ Ibid, pp. 2-3.

What they call perspective-taking is what we earlier referred to as “in-his-shoes imagining,”¹² a process similar to transference simulation. Experience-taking, on the other hand, is much more like what Gardner calls “sympathetic imitation,” and what we’ve been calling transformation simulation.

The first set of studies is intended to answer the question whether readers actually engage in experience-taking while reading fictional narratives. The researchers hypothesized that three factors would be relevant to the possibility and depth of experience-taking: self-concept accessibility, narrative voice, and shared group membership between reader and character. The first factor, self-concept accessibility, “the extent to which one’s personal identity is salient when reading a work of fiction,”¹³ would be found in the reader herself. They predicted that being in “a state of reduced self-concept accessibility should promote higher levels of experience-taking by making it easier for readers to ‘forget’ themselves and simulate the experience of a character.”¹⁴ The second factor, narrative voice, would be found in the narrative itself, and they predicted that stories told in the first person voice would be more likely to effect experience-taking than those told in the third person voice. The third factor, shared group membership between reader and character, would result from the meshing of features of the reader and the narrative. They predicted that the chances of experience-taking are higher when the reader and character belong to the same relevant group, such as race, gender, geographical location, culture, than when the character is affiliated with an outgroup relative to the reader. The highest likelihood of experience-taking would be found in stories told from the first personal perspective of a

¹² See chapter 2, section 1.

¹³ “Changing Beliefs and Behavior Through Experience-Taking,” p. 3.

¹⁴ *Ibid.*, p. 3.

character who belongs to the same group as the reader, who herself has diminished the conscious awareness of her own self.¹⁵

Results from the first three studies conformed to their predictions. The first study suggested that narrative voice and group membership do affect the likelihood of the reader's being able to experience-take, and that readers of first person narratives with characters belonging to the same group as the reader report that they simulate the character's experience. The second two studies showed that self-concept accessibility is directly related to the likelihood of experience-taking. The conclusion of the first set of studies is that when conditions are right, readers engage in experience-taking, a deeper and more profoundly interesting phenomenon than relating to characters as a mere observer or engaging in in-his-shoes imagining.¹⁶

The second set of studies—the fourth, fifth, and sixth studies outlined in the paper—were run in order to determine whether the experience-taking undergone by readers has any behavioral or attitudinal effects. In the fourth study the variables of narrative voice and group membership were manipulated—subjects read stories told from either the first or third person, including characters from either the reader's ingroup or outgroup. All of the stories featured a character voting on Election Day. The researchers wanted to see “whether the behavioral effects of experience-taking were powerful enough to emerge days after the initial laboratory session when readers first encountered the character,”¹⁷ so they asked subjects on the day of the experiment, prior to reading the story, to report how likely there were to vote in the 2008 Presidential election

¹⁵ This diminishing of awareness of the self is similar to the process of decentering discussed in chapter 8.

¹⁶ What the authors don't address, but would seem to be an implication of their findings, is that, in order to experience-take, readers would have to assume the internal perspective only, at least while they are engaged in experience-taking. Toggling back and forth between perspectives is almost surely an obstacle to such imagining. This point might appear to be trouble for my claim that the adept reader, to which my account most obviously applies, occupies the point of critical equilibrium, a perspective intermediate between the internal and external points of view. But we should recall that the adept reader not only has the ability to occupy this point, but the wisdom and the necessary experience to know when to avoid one or the other perspective. Thus an adept reader caught up in a story will very often disregard the external perspective, for the sake of the identificatory experience.

¹⁷ “Changing Beliefs and Behavior Through Experience-Taking,” p. 9.

the following week, and followed up with subjects the week after the election to find out whether they had voted. The results showed that those readers who had identified strongly with the voting character and had reported a low likelihood of voting ended up voting in significantly more cases than would have been predicted based on the earlier reports. These results lead the researchers to conclude that

[S]haring a group membership with a character from a story told in first-person voice promoted an enhanced level of experience-taking, and the impact was still evident in participants' behavior several days afterward. These findings strikingly demonstrate that through their choices in crafting the language and the content of their stories, writers can heighten the likelihood of readers' taking the character's subjective experiences as their own and, thus, emerging from the story with their identities, mindsets, and actions transformed.¹⁸

The fact that the behavioral effects occurred several days later also suggests, the researchers say, “that experience-taking has the potential to create *durable* changes in behavior.”¹⁹ The importance of the writers' “choices in crafting the language and the content of their stories” is highlighted by the researchers, since getting a reader to identify with a character is an *achievement* largely brought about by the writer's skill. One of these choices has to do with how to present the character to the reader. The fifth and sixth studies were run to determine whether any difference is made when an outgroup character is immediately presented as such or delayed until later in the narrative. The fifth study involved heterosexual subjects who read a story featuring a central gay character and the sixth involved Caucasian subjects reading a story featuring a central African American character.

¹⁸ Ibid, p. 10.

¹⁹ Ibid, p. 16, my emphasis.

The fifth study outlined in the paper had male heterosexual subjects read stories featuring either a gay or straight male main character and, in the stories in which the character was gay, the sexual orientation was revealed either early or late in the narrative. Prior to reading the stories, subjects were administered a questionnaire intended to determine their attitudes toward homosexual people. The same questionnaire was administered immediately after subjects had read the story. As they read the story, subjects were asked to rate characters on two scales: feminine/masculine and calm/emotional. The point of this was to determine to what extent readers were applying stereotypes associated with gay males to the character, as the characters were created without any unusually feminine or emotional characteristics.

What the researchers found was that, in the group of subjects who read the story in which the character was revealed as gay early in the narrative, readers didn't report experience-taking and their attitudes toward homosexual people remained basically the same as before and in the group of subjects who read the story with a straight character, no differences of attitude were reported there either. But in the group of subjects who read the story in which the character was revealed to be gay later in the narrative, after the reader had already spent some time getting to know the character, subjects did report experience-taking. Moreover, stereotyping was less present in this group than in the other groups. And perhaps most notable of all, there was a significant difference between the pretest reports of attitudes toward homosexual people and those given right after reading the narrative, showing that "that readers of the gay-late narrative reported a significantly more favorable attitude toward homosexuals than did readers of both the gay-early narrative [...] and the straight narrative."²⁰

The sixth study involving race produced similar results: when the race of the character was revealed later in the narrative, subjects reported experience-taking. And when they reported

²⁰ Ibid, p. 11.

experience-taking, they also showed evidence of attitudinal changes toward African American people. These results lead the researchers to conclude thus:

The final two studies showed that revealing a character's stigmatized group identity—specifically, his identity as a homosexual or as an African American—later versus earlier in a story was an effective technique for overcoming the barrier to experience-taking that nonshared group membership typically creates. When an outgroup member's identity was revealed later rather than earlier, readers were just as likely to simulate his experience as if he had been an ingroup member. Further, as a result of this experience-taking with the outgroup member, readers of the late revelation narrative were less likely to judge the character stereotypically, and these readers expressed more favorable attitudes toward the character's group. As a whole, the results from these six studies validate our conceptualization of experience-taking as an immersive, simulative experience with the power to change readers' self-concepts, behaviors, and attitudes.²¹

This study gives strong evidence that readers can be moved to care for real individuals after reading fictional narratives, and, most important for the present chapter, the change occurs *as a result of* reading the narratives. In other words, what we have here is not mere correlation, since the self-reports provided prior to reading the narratives reflect a generally negative attitude toward members of the outgroup.

Another recent paper provides even more persuasive evidence that what we're seeing isn't mere correlation, but causation. This paper²² summarizes two studies. The first study gives results similar to those we have just discussed: when subjects read a fictional narrative involving a high degree of emotional identification with a central character and were thus transported into the narrative, subjects reported having a higher degree of empathizing, while subjects who read

²¹ Ibid, p. 15.

²² P. M. Bal & M. Veltkamp, "How Does Fiction Reading Influence Empathy? An Experimental Investigation on the Role of Emotional Transportation," *PLoS ONE* Vol. 8, No.1, 2013, pp. 1-12.

non-fictional accounts did not report the same effects.²³ The second study represents strong evidence against the mere correlation hypothesis, however, since it involved subjects reading a fictional narrative that didn't involve a high degree of emotional identification and into which subjects weren't transported. If the relation in the first study were mere correlation, there shouldn't be any change in reports of degree of empathizing, but subjects actually reported a lower degree of empathizing. The researchers conclude that there is a causal relation between being transported into the fictional narrative: a higher degree of transportation led to higher degree of empathizing, while "low transportation led to lower empathy over time."²⁴ They provide no definitive explanations for why this occurs, but they speculate that "when a reader is not able to identify with a text and does not become transported, this might lead to disengagement, with the reader being distracted and frustrated" and that when "readers disengage from what they read, they possibly become more self-centered and selfish in order to protect the sense of self in relation to others."²⁵ If this explanation is anywhere near correct, the claim that what we are seeing is mere correlation is highly doubtful. Both of these studies also show the importance of a steady diet of fictional narratives, since no one knows whether the narrative one is about to read is going to be one with a high degree of transporting or not and it's inevitable that one will often take up narratives with a low degree of transporting. In the long run it's better to have taken up as many narratives as possible in order to have experiences of narratives with a high degree of transporting.

One more relevant factor is discussed in this paper: the researchers say that the "effects of fiction experience on empathy are guided by an absolute sleeper effect," which is defined as

²³ Interestingly, when subjects reported being transported into the non-fictional accounts—newspaper reports of events of riots in Libya and the nuclear meltdown in Japan—they also reported becoming less empathic.

²⁴ "How Does Fiction Reading Influence Empathy?" p. 8.

²⁵ *Ibid*, p. 8.

“when the effects of a manipulation do not present themselves immediately, but manifest themselves over time.”²⁶ They write that,

[F]or sleeper effects to occur, an incubation period is needed, in which people can rethink and relive that what has been read. Research on incubation has shown that spending some time on unrelated activities may enhance the effects of resolving problems, because an individual unconsciously connects the information from fictional narratives (e.g., people facing problems in their lives) with daily encounters, and consequently find new solutions through perspective taking and showing sympathy for other people.²⁷

The importance of “incubation” and “spending some time on unrelated activities” brings back into view the reasons why we have focused on a life in which aesthetic experience plays a central role, rather than isolated experiences. Being moved to care is part of the transformation, but becoming a person who cares, a person possessing the virtue of considerateness, is a slow and gradual process that must happen over time. We will now be able to see how the transformation takes place.

8. Seduced into Caring

We should take note of the fact that the psychological studies discussed in the previous section all presumed a kind of “transporting” of the reader into the narrative. For this transporting to occur, the reader for her part must be willing to let go some of her egocentric

²⁶ Ibid, p. 3.

²⁷ Ibid, pp. 3-4.

tendencies and the author for his part must do what he can to lure her into the storyworld and get her interested in the plight of the characters. Often this transporting results in the reader's being *led to care* for some or all of the characters, a feat requiring great artistic and rhetorical skill. Art has tremendous power to transform audiences, but it doesn't just happen—such transformations happen, when they do, as a result of talented individuals making the right choices based on a kind of practical wisdom gained by the experience of making art. Of course, an author may decide to forestall reader identification, thus subverting reader expectations, and present the character in ways intended to thwart experience-taking. Nothing aesthetically evaluative necessarily follows from the absence of transporting; but for the average reader, it's probably true that transformations of behavior and attitudes won't follow from reading narratives featuring characters for whom the reader doesn't have any concern. We saw that an effective means of bringing the reader into this relation with a central character is the adoption of the first person voice combined with a delay in the exposing of character's membership in an outgroup relative to the reader; but this isn't a necessary means. All that is necessary is that the story be told in the right way, so that the reader is *drawn in*. Whether there are other means is a matter for the psychologists.

In addition to the story's being told in the right way in order to draw in the reader, the effects of time were highlighted in the point about sleeper effects. There need to be spaces in between individual experiences of fictional narratives and opportunities for reflection on the narrative events as well as the blending of the experience with daily life. Peter Kivy²⁸ argues that some of the most profound effects of art—literature in particular—often happen in “the gaps” between individual experiences. The most basic instance of this is in the time spent between sessions of reading—such as when one reaches a chapter break and stops reading for awhile.

²⁸ *Once-Told Tales: An Essay in Literary Aesthetics*, Malden, MA: Wiley-Blackwell, 2011.

When this happens, Kivy says, we have “matter we are meant to think about, during the gaps, of a philosophical, or psychological, or other nature, depending of course on the work in question.”²⁹ Some works of fiction read by a serious reader aim at “enlightenment, both moral and intellectual” and in the gaps between reading, “the reader is meant to think about this material, *as part of the literary experience*.”³⁰ When you think in this way, “you are usually [...] thinking about the content of what you have read, not the form and structure. You are thinking about the story, the characters, their motivations, what meaning the author might intend to convey with her story.”³¹ Thus the insights you gain, according to Kivy, come not from seeing the story as a mere vehicle for insight. You must approach the work on its own terms, be affected by it in the ways that art affects us, and only *then* will you gain insight. Approaching a work of art with only non-aesthetic aims—like wanting to learn moral truths or how to be a better moral agent—is more often than not going to be self-defeating. This doesn’t mean, however, that one cannot experience these benefits from engaging with art—it just means that the work should be treated first and foremost as art. This is why the worries raised by the radical autonomists described in chapter 4 fail to gain traction. The claim that it’s illegitimate to value a work of art for its cognitive or moral, because doing so fails to treat the work *qua* art, assumes a *value-monism*, but art as an institution has the power it has because of its *multiple* dimensions of value. We can value it for things other than aesthetic experience, but in most cases, valuing the work in any other way demands first that we value it aesthetically.

Iris Murdoch says the reward of attention in the aesthetic context is beauty, but more broadly, the reward is aesthetic satisfaction. We enjoy the fictional narrative artworks that have aesthetic value and it’s because we *enjoy* art that we keep coming back to it, time and again. As

²⁹ Ibid, p. 81.

³⁰ Ibid, p. 81, emphasis in original.

³¹ Ibid, pp. 81-82. Notice the similarity of this description to my discussion of critical equilibrium.

we do so, if we're seeking those artworks that really challenge not only our cognitive capacities, but our worldviews, we ourselves change—for better or worse. The reason we cannot say that all change is for the better is that, in some sense, the reader comes under the spell of the artist—at least the artist possessing the artistic and rhetorical skill we have been discussing. When the reader is under the artist's spell, she is led to enter cognitive and affective states virtually without knowing that it's happening. Why? Because of the promise of the aesthetic enjoyment and beauty that lies there waiting to be discovered. But then the reader is led to do what she already wants to do.³²

The moral payoff for living the life of the reader is the amelioratory effects on mindreading discussed in chapters 8, 9, and 10, as well as the increased capacity to have empathetic concern for others discussed in this chapter. Recall the claims from Warnock we amended in chapter 1:

- (1) the deficit in caring is supremely important (no other deficit is greater) when we are trying to solve the human predicament
- (2) it's the object of moral philosophy (a) to expand our sympathies, if possible; (b) to mitigate the damage done by our limited sympathies, if (a) is not possible

If it's the object of moral philosophy to expand our sympathies, then presumably the way a moral philosopher would approach this task is to offer propositional knowledge that has the

³² See Wayne Booth, *The Company We Keep: An Ethics of Fiction*, Berkeley, CA: University of California Press, 1988, where it's argued that we relate to authors of fiction as if they are our friends. Once the promise has been made, we take the author at his word, but at the same time, once we've agreed to trust the author, we now have the obligation to follow the directives necessary for the fulfillment of the promise.

desired effect.³³ In this dissertation I have argued that art is an effective means of expanding our sympathies, not by virtue of introducing new propositions, but rather by bringing the person who lives the life of the reader to make certain cognitive and affective adjustments in order to make sense of fictional narrative artworks so that these artworks can be enjoyed. The conclusion drawn here is that there is a *substantive, isomorphic* relation between a person's becoming and remaining an aesthetic adult, thus living the life of the reader, and a person's possessing the right kind of motive necessary for being a virtuous agent. But as we've seen, a particularist approach is too limited and too limiting, since the long-term aggregative effects are not brought into view and the obvious fact that aesthetic experience happens within the messiness of life is intentionally ignored. The effects of art, I contend, emerge slowly and gradually, and require the butting of aesthetic experiences against real world issues. The vision of audience members retreating from the vagaries and infelicities of life in order to have an aesthetic experience isolated from the experiences of life is both idealistic and obsolete. People usually experience art from *within* the chaos of life. And the aesthetic adult, I have maintained, draws no distinction between aesthetic and real life experience—for this person, *art is a (significant) part of life*.

Art might be a more effective means of expanding our sympathies than moral philosophy, because we *don't* approach it in order to become better persons. Murdoch says that we become better able to see the world as it really is, through the capacity for *attention*, as a result of engaging with art. In the case of “art and nature such attention is immediately rewarded by the enjoyment of beauty,” but in “the case of morality, although there are sometimes rewards, the idea of reward is out of place. Freedom is not strictly the exercise of the will, but rather the

³³ Although we should keep in mind Nussbaum's claims discussed in chapter 6 that novels themselves might be considered works of moral philosophy.

experience of accurate vision which, when this becomes appropriate, occasions action.”³⁴ This is why the best metaphor I can offer for illustrating how art transforms us is *seduction*. If we adopt the black and white Hollywood version of seduction, the target of the seduction is a person who really does want to be romanced, but expresses resistance for any of a number of innocuous reasons—e.g., convention, low self-esteem, uncertainty, false impressions. The seducer sees that the target wants to be romanced and commences to try to get the target to accede in a number of ways: attractive presentation, luring, garnering sympathy, faux-rejection, teasing, suggestion—really any method short of force.³⁵ The lovelorn character we are presented with in the beginning of the film has now found a suitor who can provide salvation. But it takes the seducer an extended period of time and a subtle approach to succeed.

Art takes those persons who want to be good, who wish to be caring and considerate individuals, and seduces them with aesthetic enjoyment and beauty into becoming the persons they want to be. Just as a seduction won’t work with a target thoroughly uninterested in romance, art cannot make a bad person good. It can, however, make a person who desires to be good *better*. “That the highest love is in some sense impersonal is something which we can indeed see in art,” says Murdoch, “but which I think we cannot see clearly, except in a very piecemeal manner, in the relationships of human beings.”³⁶ Even more than seeing love in art, the reader must be in a state of love—under the Murdochian conception of love, which is not the standard conception—as she engages in art. The artist must “silence and expel self, to contemplate and delineate nature with a clear eye, [and] is, in respect of his work, a good man, and, in the true

³⁴ *The Sovereignty of Good*, p. 67.

³⁵ Unfortunately, in the old movies, force is sometimes used because the one character believes that the other really wants to be kissed, in spite of the vituperative protests, and nothing short of force will accomplish this. Obviously, this kind of behavior is not being endorsed in my account.

³⁶ *The Sovereignty of Good*, p. 75.

sense, a free man.”³⁷ Similarly, the “*consumer* of art has an analogous task to its producer: to be disciplined enough to see as much reality in the work as the artist has succeeded in putting into it, and not to ‘use it as magic’.”³⁸ Seeing “beauty in art or nature is not only (for all its difficulties) the easiest available spiritual exercise; it is also a completely adequate entry into (and not just analogy of) the good life, since it *is* the checking of selfishness in the interest of seeing the real.”³⁹ It’s what Murdoch calls “unselfing” that is, by my lights, the most valuable process that an aesthetic and moral adult undergoes as a result of the seduction into caring. Unselfing—“the checking of selfishness in the interest of seeing the real”—is more than preparatory to acting and reasoning morally. It’s at the core of what constitutes being a good person.

The film now ends with our view trained on a final Murdochian notion—humility. Murdoch suggests that humility is the “central of all virtues.” Considerateness depends on unselfing, and unselfing requires the attenuation of the regard for one’s self. “The good man is humble,” Murdoch writes, and “[h]umility is a rare virtue [...] unfashionable [...] and often hard to discern. Only rarely does one meet somebody in whom it positively shines, in whom one apprehends with amazement the absence of the anxious avaricious tentacles of the self.”⁴⁰ Humility may very well be the mother virtue, since it’s difficult to see how any of the other virtues can exist without it. The film began by our posing the question of how to expand our sympathies and the answer we end up with is that we should look *away* from our selves and toward what is good for others. As the closing credits roll, let Murdoch’s words serve as the soundtrack:

³⁷ Ibid, p. 64.

³⁸ Ibid, p. 64, my emphasis.

³⁹ Ibid, pp. 64-65.

⁴⁰ Ibid, p. 103.

In fact any other name for Good must be a partial name; but names of virtues suggest directions of thought, and this direction seems to me a better one than that suggested by more popular concepts such as freedom and courage. The humble man, because he sees himself as nothing, can see other things as they are. He sees the pointlessness of virtue and its unique value and the endless extent of its demand. Weil tells us that the exposure of the soul to God condemns the selfish part of it not to suffering but to death. The humble man perceives the distance between suffering and death. And although he is not by definition the good man perhaps he is the kind of man who is most likely of all to become good.⁴¹

⁴¹ Ibid, pp. 103-104.