Understanding: Moral Evaluation
and the Ethics of Imagining

Christopher Woerner

This thesis is submitted in partial fulfilment for the degree of PhD
at the
University of St Andrews

30/01/13
1. Candidate’s declarations:

I, Christopher Woerner, hereby certify that this thesis, which is approximately 74000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2009 and as a candidate for the degree of PhD in September, 2009; the higher study for which this is a record was carried out in the University of St Andrews between 2009 and 2013.

2. Supervisor’s declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 30/01/2013

signature of supervisor

3. Permission for electronic publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.
The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

(i) Access to printed copy and electronic publication of thesis through the University of St Andrews.

Date 30/01/2013

signature of candidate

signature of supervisor
UNDERSTANDING: MORAL EVALUATION
AND THE ETHICS OF IMAGINING

INTRODUCTION

CHAPTER ONE: THE PRACTICE OF UNDERSTANDING
1.1 Understanding, For and Against 6
1.2 Character Engagement 15
1.3 Simulation 24
1.4 Closing Remarks 34

CHAPTER TWO: UNDERSTANDING AS A PHILOSOPHICAL TRADITION
2.1 Adam Smith and The Theory of Moral Sentiments 35
2.2 Nel Noddings and the Ethics of Care 42
2.3 Richard Hare and Moral Thinking 49
2.4 Closing Remarks 59

CHAPTER THREE: UNDERSTANDING AND ETHICS
3.1 The Prevalence of Understanding 60
3.2 Understanding and Consciousness 65
3.3 Understanding and Non-Human Animals 73
3.4 Understanding, Aggregation and Autonomy 81
3.5 Closing Remarks 85

CHAPTER FOUR: UNDERSTANDING AND AESTHETIC COGNITIVISM
4.1 Understanding and the Cognitivist Gap 87
4.2 Fictive Distortion and Imagined Realism 98
4.3 Closing Remarks 104

CHAPTER FIVE: THE ETHICS OF IMAGINING
5.1 Understanding and Fictive Externalism 106
5.2 Understanding and Interpretation 120
5.3 Closing Remarks 130

CONCLUSION
ABSTRACT

Analytic ethics often neglects the exploration and appreciation of morality as it is actually practised on a day-to-day basis. But by looking at how, in a practical sense, we are able to interact with others in a morally appropriate way we can construct a compelling picture of what some of our most pervasive obligations are. This thesis takes such an approach through the concept of understanding – understanding essentially taken here to involve those processes involved in detecting and correctly responding to beings typically possessing inherent moral significance. In the first two chapters ‘understanding’ and the understanding approach are themselves explicated, and placed in the context of several other related approaches in the English-speaking tradition – Adam Smith’s *Theory of Moral Sentiments*, Nel Noddings’ ethics of care and Richard Hare’s preference utilitarianism. This approach is then used to provide us with an alternative idea about what our moral reasoning suggests to be of fundamental ethical significance, and of what kinds of activity morality recommends to us. The activity explored in most detail here is that of engaging with fiction – or more broadly, fictive imaginings. While understanding shows us that fictional characters and events themselves cannot have an inherent moral valence or significance, it also shows us when and how it is possible and appropriate to ethically assess fictive engagement, be it as creator or consumer. This is seen after exploring how and in what ways our moral understanding can be appropriately applied to and exercised by fictions at all, and why fiction should be of particular interest to the understanding agent, looking at the work of Martha Nussbaum, Jenefer Robinson, Peter Lamarque and others on aesthetic cognitivism. Ultimately this leads us to discern a minimal ethical constraint on our interpretation of fiction and art in general, further proving understanding’s usefulness.
INTRODUCTION

Níor thuig seisean cé mar a bhí an scéal aici. Bhí an ghéarchúis ann agus an mheabhair agus an éirim aigne. Dá mbeadh air margadh a dhéanamh le duine, bí cinnse nach dteipfeadh air. Ba mhaith an fear é ag múinte agus ceist fheithiúcháin, ach níor thuig sé daoine nár múnlaíodh ar a nós garbh féin. Céard ba chiontach leis? An saol bhí aige sna Stáit, an scláthbaíocht, an doichealla, agus an deifir mhór a bhíodh air ag bailiú an airgid? Nó an deamhán a thug sé leis ó dhúchas a bhí ann?

Is minic, nuair a bhíodh a bhean go mishasta agus go cantalach, agus go leigheasfadh an focal ceart feiliúinach í, go n-abraiódh sé focal garbh éigin a chuireadh déisteann uirthi. Thug sé an déisteann geo faoi deara go minic, ach níor thuig sé cén fháth a bhí leis. Bean mar í níor bhuaileadh leis riamh cheana; na mná a casadh leis sna bailte móra san Oileán Úr, ní raibh aon rud is mó a chuireadh aoibhneas orthu nó moladh maith agus bronntanas.

Cheap sé go dtabhairfadh sé bronntanas álainn dá bhean agus go mb'fhéidir go scaoiFeadh sí leis na bealaí aisteacha a bhí léi.

Chuaigh sé isteach go Gaillimh in aon turas leis an mbronntanas a cheannach dí. Chuartaigh sé na siopaí ar fad, agus sa deireadh cheannaigh sé gúna breá den sioda drithleannach a bhí déanta sa bhfaisean ba dhéanáidh. Nó shásodh sin é go gceannódh sé biorán móir óir le cur ina brollach. Bhí sé fial fairsing léi i gcónaí, ach chuir sé beagán thar an iomarca ar na hearráthí seo, ach níor mhiste leis – nach uirthi a hbeadh an taoibhneas na socruithe suas a bhí déanta i gceannád níos chúrsú ná an gúna sin ní fhaca sí riamh.

Agus bhí aoiibhneas uirthi. Gúna nós ól a ná an gúna sin ní fhaca sí riamh. Agus an déanamh a bhí aír? Nuair a rug a fear ar uachtar an ghlúna lena thaispeáint di, cheapfá nach bhfeadfadh aon bhean dá chaoile dá mbeadh sí dál isteach ann, bhí sé chomh cúise sin.

Rug an bhean ar an ngúna go gcuirfeadh sí uirthi féin é.

Táim an-bhui ocht diot, a Shéamais, an-bhui ocht diot ar fad,” ar sise.

Chuir sí uirthi an gúna. Bhí sé an-chúng ar fad dí, agus dúirt seisean é.

“Is amhlaidh is fearr liom é, a Shéamais. Bead sa bhfaisean . . .”

“Ach i gceann cápla mí ní fhéadfadh tú é a chur ort ar chor ar bith . . .”
He could not understand how she had the story. He was astute and intelligent. If he had to do a deal with someone, you could be sure he would not fail. He was a good man for explaining questions of business or politics, but he did not understand people who were not modelled on his own rough ways. Why was he guilty of this? The life he had in the States, his slaving, his depression, and the urgency with which he gathered his money? Or was it a demon he took with him from heredity?

Often, when his wife was unhappy and irritable, and the right word would have cured her suitably, he would utter some rough word or other that aroused her disgust. He noticed this disgust often, but he did not understand the reason for it. He had never met a woman like her before; for the women he had met in the cities of the New World, nothing made them happier than compliments and presents.

He thought he would give his wife a beautiful gift and that maybe this might release her from her strange ways.

He went in to Galway intent on buying this present for her. He visited all the shops, and in the end he bought a lovely silk dress that was made in the latest fashion. This did not satisfy him until he also bought a large gold pin to put on her breast. He was always very generous to her, though he spent a bit more again than was necessary on these items, but it did not matter to him—wouldn’t she be pleased when she saw them?

And she was pleased. She had never seen a more beautiful dress. And the workmanship of it? When her husband grasped the top of the dress to show it to her, you would think that no woman would be able to get into it, it was that narrow. The woman took the dress to put it on.

"I am very grateful to you, Seamus, very grateful indeed," she said.

She put on the dress. It was very narrow indeed for her, and he said so.

"In fact I prefer it that way, Seamus. It’s the fashion..."

"But in a few months you won’t be able to get it on at all..."

He said many other things not to be told here, and not liked by his wife, and when he would not stop his talk the weeping broke upon her again and she went straight back to her room.  

---

1 From “An Bhean a Ciapadh”, by Pádraic Ó Conaire, 1913.
2 Translation my own.
The above is an extract from the short story “The Tormented Woman”, by Galway-born writer Pádraic Ó Conaire. The story tends to be seen specifically as an attack on cleamhnas, the old Irish matchmaking system, but it useful also as an example of profound interpersonal incomprehension more generally. Its beginning sees a middle-aged Irish bachelor named Burke returning home from nearly thirty years spent working in America, and deciding to look for a wife to give him a house full of children. To this end he visits an old friend of his to ask for his daughter’s hand in marriage, which he is granted. But the woman is much younger than Burke, she is from a different time, and not only does she resent being forced into marriage but the two find it impossible to relate to one another. Matters become worse as she loses her first child at birth and is told that to become pregnant again could well kill her. Burke, in his frustration, turns to drink, and on several occasions he rapes his wife. Ultimately she becomes pregnant again and dies in childbirth, leaving her husband to take uncomprehending and unabashed comfort in finally having an heir. The extract given here describes one of his well-meaning but clumsy and shallow attempts at reconciliation; initially he meets with some success, before spoiling it by falling back into his callous habit of fixating on her capacity for breeding before even establishing a proper relationship with her.

This scenario demonstrates an extremely wide-ranging problem, involving people going wrong from a failure to properly understand one another. In the story Burke never develops any real sense of his wife’s perspective, and consequently he never gains the ability to seriously weigh her interests against his own – indeed he cannot even manage to avoid constantly upsetting and disgusting her. Importantly, this is not due to any special maliciousness on his part; it is more a failure of imagination, a cultural and normative narrow-mindedness that is simply unable to process viewpoints significantly different to his own. We can truthfully say of him that he is misogynistic and selfish while at the same time he would no doubt greet such criticism with sheer bafflement, since he lacks the capacity to call his basic assumptions and perspective into question. Contrast the old sayings ‘it’s the thought that counts’ and ‘the road to hell is paved with good intentions’ – clearly it is not just the thought that counts, because we are responsible for checking to see if our intentions are being appropriately manifested. It follows from this that if we want to treat others well we must first know how to treat them.

Consider next this extract from the first act of King Lear, where the king asks his daughter to tell him how much she loves him:

\[
\text{CORDELIA:} \quad \text{Unhappy that I am, I cannot heave} \\
\text{My heart into my mouth: I love your majesty} \\
\text{According to my bond; nor more nor less.}
\]

This title has been variously translated as “The Woman in Torment”, “The Woman Who was Made to Suffer” and “On the Rack”.
KING LEAR:  How, how, Cordelia! mend your speech a little,
Lest it may mar your fortunes.

CORDELIA:  Good my lord,
You have begot me, bred me, loved me: I
Return those duties back as are right fit,
Obey you, love you, and most honour you.
Why have my sisters husbands, if they say
They love you all? Haply, when I shall wed,
That lord whose hand must take my plight shall carry
Half my love with him, half my care and duty:
Sure, I shall never marry like my sisters,
To love my father all.

KING LEAR:  But goes thy heart with this?

CORDELIA:  Ay, good my lord.

KING LEAR:  So young, and so untender?

Again we can see things going wrong for no good reason in what is, sadly, a recognisable fashion. Famously, Lear’s pride leads him to disown his only loyal daughter in favour of the flattery offered by the rest of his scheming progeny, with disastrous consequences. But while his reaction is of course grossly disproportionate, we need not lay the blame entirely at his feet; Cordelia, for her part, seems quite needlessly and excessively restrained, even in the face of her father’s imprudent rage. If only they had understood what they had both meant and felt, one despairs, so much heartache could have so easily been avoided. As these examples suggest, the inadequacy of operating entirely within one’s own perspective is a common and long-running literary theme; unless we try to apprehend the realities of those around us we will misjudge and mistreat them time and again. It is a message that is overlooked at great cost, both by moral philosophers and more broadly by anyone interested in behaving morally.

Nevertheless, it appears to be a lesson persistently overlooked by the majority of ethical discussion in the analytic tradition. We talk of the rational and of preferences as though the multiplicity of human perspectives might easily coalesce into a series of principles or maxims on which we could all agree, or which we can easily weigh against each other, forgetting that fundamental differences between people and cultures might easily lead to irreconcilable differences – the idea that such differences could be simply resolved by preferring the more rational perspective is so crass as to border on the meaningless. This is not to say that we cannot ever think of one position as more rational than another, and it is not to surrender to the unhappy rule of relativism, but still the

---

4 From William Shakespeare’s *King Lear*, first quarto, I.1.93-110.
possibility that perspectives directly conflicting with our own might well be equally justifiable demands that we make an effort to understand others and indeed ourselves. To do this, to coordinate with that which we think wrong or mistaken, and before this to know when another’s apparently wrong perspective is in some sense substantially justified, we must try to hone our ability to know how things seems to others, why they seem that way and what the significance of such seeming is. This is perhaps the most fundamental and important skill a moral agent can have, and it is to be the subject of this thesis.
CHAPTER ONE

1. THE PRACTICE OF UNDERSTANDING

1.1 Understanding, For and Against:

There is a central idea behind that which is to follow in this thesis: that moral agents face an obligation, a moral obligation, to be understanding, where it is my aim to explore what this means and what practical and philosophical implications this has for us. ‘Understanding’ is perhaps not the clearest of concepts at first glance, and to make matters worse I use it here in a slightly non-standard sense. In its common usage, when someone is described as understanding it is generally taken to mean that they, to paraphrase the Oxford English Dictionary, display sympathetic tolerance, or are of a forgiving nature. But this meaning, without further development, is not entirely appropriate to what I have in mind. Take forgivingness, for example. It might well be considered a virtuous trait, but there still seem to be times when forgiveness is not appropriate. People might disagree about the general patterns such cases might take – one might think that some transgressions are so serious they can never be forgiven, while others might believe that all misdeeds, even if terrible, can or even should be forgiven in the event that the transgressor has paid for their wrongdoing, or repented, or both. But of course forgivingness – a forgiving disposition – need not in all cases lead to forgiveness, and so it would be unfair to suggest that forgivingness must recommend forgiveness in what look like inappropriate cases. However, while forgivingness might then be a standing disposition to exercise forgiveness where not inappropriate, understanding as it is meant here is required in all cases, it is never inappropriate, and it by no means implies that understanding must always lead to forgiveness, or even forgivingness. So what is this notion of ‘understanding’ that drives the discussion ahead, and what makes it a suitable term if it is divorced from common usage?

The first thing to say is that there is some significant commonality between these two senses of understanding, its common sense and the philosophical sense which it is given here. ‘Understandingness’ as I mean it can involve forgiveness to the extent that it implies one does not fetishize abstract moral rules – insofar as understanding demands that one pay attention to the personal history, character and psychology of an agent, it might often lead us to forgive in apparent cases of wrongdoing. Being understanding in this philosophical sense implies that one’s eyes are open to the possibility that a given transgression might be trivial, non-culpably unintended, or subject to other extenuating circumstances. It also implies that one is alive to the interests, concerns, and most importantly the perspectives of those involved in a given event or state of affairs, where this includes a fallibilist open-mindedness to the possibility that one’s own perspective could be mistaken or ill-

---

5 Although, conversely, understanding might for this reason also lead us to object to behaviour that seems apparently unobjectionable.
informed. Given that we can never be sure of fully understanding anyone, and more broadly perhaps anything, the most appropriate thing to do is to cultivate an open-mindedness which counts against an unwise tendency to see our own beliefs as better, more accurate or more important than those of others. This all points to perhaps the most significant common ground between philosophical and garden variety understanding; an emphasis on the role of comprehension, as opposed to the more restricted everyday trait of ‘sympathetic tolerance’ which does not imply much more than ‘permissive’ or perhaps ‘conciliatorily permissive’.

Responsible tolerance, like responsible forgivingness, knows that the internal life of a person can sometimes, indeed that it generally does, make a difference to the moral import of their actions, regardless of how far one thinks this difference can go. So proper tolerance should also include my construal of understanding. But what about this idea of understanding as sympathetic tolerance? ‘Sympathy’ is an important concept, an important phenomenon, one which has an interesting history in ethics. In contemporary terms, sympathy is generally given one of two interpretations. As Berys Gaut says, ‘to sympathise with a character is in a broad sense to care for him, to be concerned for him’, but how this concern is cashed out varies. In recent times the prevailing philosophical approach has been that of Gaut, Susan Feagin and Matthew Kieran, who all more or less hold that sympathy as such can be instantiated in a wide variety of emotional states provided they are, in Feagin’s words, ‘in concert with the interests or desires the sympathiser... attributes to the protagonist’. One is sympathetic to y, on this reading, by feeling any of a number of emotions for y’s sake – one can feel sympathetic happiness, sympathetic regret and so on. The other, slightly less popular interpretation is that of people like Peter Goldie, who insists on a more narrow definition of sympathy as being a single, specific emotion of concern for another – the problem being that this seems, as Kieran suggests, to be a conflation of sympathy and compassion.

Sympathy, then, might look like a good candidate if one is searching for a crucial ingredient of optimal moral appraisal. How could one be more motivated to act morally, one might ask, than if one felt the interest of others to be of personal concern to one, as expressed either in one sort of feeling or in many? But as with forgiveness, there are times when sympathy could be inappropriate in a way which is not true of understanding. Imagine you are a reporter, going to interview an unrepentant SS officer who has been given a moderate sentence in a relatively comfortable prison. Now imagine during your visit that this Nazi bemoans his fate pitifully, complaining of how bland the puddings are. For you to feel sympathy or sympathetic sadness in such a case would arguably be an

---

6 Berys Gaut, “Identification and Emotion in Narrative Film”, p. 207.
7 See Berys Gaut, “Identification and Emotion in Narrative Film”, p. 207, Susan Feagin, Reading with Feeling, p. 114, Matthew Kieran, “In Search of a Narrative”, p. 81. The idea here is meant to allow for the fact that one can feel sympathy towards y even if one is mistaken about the interests or desires one attributes to y, or when one feels sympathy on the basis of facts of which y is ignorant (where these conditions are typically seen as defeaters for empathy).
injustice, since it seems inappropriate to feel sympathy for a wrongdoer who is getting very much less than he deserves – more generally, one might say it is inappropriate to feel sympathy for (severely) disproportionate complaints.9 And there is another worry some have expressed over the idea of holding sympathy up as the paragon of moral sentiments: sympathy is partial, and furthermore there does not appear to be anything internal to the concept of sympathy which requires one to correct this partiality. As Goldie says, ‘the partiality of our sympathies... reflects the partiality of our wider dispositions of concern towards our nearest and dearest’.10 Meanwhile, the appropriateness of understanding remains constant across cases – without it we would be unable to determine when sympathy, and in what form, is appropriate, when complaints are or are not disproportionate, and so on, because such things are affected by the characters and experiences of those involved.

So, having discounted the standard definition of understanding, a more explicit statement of my meaning is called for. When I say that moral agents must be understanding, what I mean is that when a moral agent forms a moralised opinion of some person y, or responds in some way to y, they must in quite a broad sense understand insofar as they can what y believes and why y believes it, they must understand what aspects of y’s psychology are in play which y might not be aware of or have control over, and beyond this they must have some reliable sense of why they themselves have the beliefs they do and a standing willingness to be moral – that is, they must be open to the motivational force of reasons that apply to other people. An understanding person understands themselves and others and has some sense of the real moral significance of things. A crucial part of understanding people for the moral agent is also, I want to suggest, the ability to ‘imagine from the inside’, to imaginatively transpose oneself as much as possible into the position of another – not necessarily because it is always a means to better understand people, which in some cases it might not be, but at least because it makes others’ perspectives more alive to us. In this sense such imaginative projection can help us to better understand the significance of what it is we are understanding. It can also in this way act as a strong incentive to follow the Golden Rule – one should not just treat others as one wants to be treated, one should really feel as though in treating y is such a way one is treating oneself (as them, in imagination) in such a way. Let us call this sort of imaginatively rich understanding thick, bearing in mind that understanding need not always involve an extremely vivid imaginative act – indeed this sort of process might often occur subconsciously or without any conscious effort. Let us then call understanding which does not or barely requires such imaginative perspective-taking thin.11

---

9 This more general claim is true of right thinking adults, as it were, but is less accurate when applied to the insane or very young – in other words, when applied to people who cannot be expected to reliably track how proportionate their complaints are.
10 Peter Goldie, The Emotions, p. 216. See also Matthew Kieran, “In Search of a Narrative”, p. 82.
11 This corresponds very roughly with what Alvin Goldman calls ‘high-level’ and ‘low-level’ mind-reading, as shall become apparent later on in this chapter – the primary difference being that thin understanding can also include conscious employment of psychological and behavioural theories in interpreting and assessing behaviour.
A great deal of work in ethics tends to read as though we observe an obvious, objective world of actors and actions, a world in which it is relatively easy for us to point and say ‘this is wrong’, ‘that is unfair’, ‘this is what should have been done’ and so on. But time spent living in the real world should be enough to divest one of this notion. This is not to endorse some kind of broad ethical scepticism or relativism, although they might very well be compatible with something like the position I hold. What I mean, rather, is that one does not have to deny that there is a thing which is unequivocally wrong, or that there is a thing which is unequivocally unfair, to realise and accept that a list of factors, the totality of which when present would, all things being equal, constitute a transaction’s being unfair (say), would not be a failsafe means of telling when to accuse particular people of unfairness. To judge an act by purely external or non-holistic standards is on my reading of affairs to invite misjudgement - the judgement that some person y is doing x, I hold, is significantly incomplete unless one also knows what it is that y believes he is doing, and furthermore what the reasons are for y’s so believing, together with, again, a sense of how one sees things oneself and why. This goes beyond an emphasis on the usual philosophical construal of intention, which does not tend to go past a simple sentential representation of some person’s beliefs and desires, because such a construction is inadequate without at least some sense of the history behind the intention in question and what the relevant terms mean to the subject, both in terms of personal significance and denotation; what is involved here is the understanding of intention in a much more global and imaginatively rich way than analytic philosophers usually discuss.

Despite my dismissal of sympathy as a reliable tool for ethical reasoning, it bears remarking that the form of understanding I wish to promote is in some respects quite similar to another, much earlier conception of sympathy, that set out by Adam Smith in *The Theory of Moral Sentiments*. Smith’s own use of the word “sympathy” was quite broad, ranging from what would today be called *in-his-shoes imagining* to empathy – full sympathy was achieved when, upon imaginatively projecting oneself into the position of another, one felt an emotion which was the same in kind if not in degree as that of the one being sympathised with, but sympathy as a phenomenon also referred to the process of imaginative projection itself for Smith, and whatever emotion was felt as a result of this projection he described as sympathetic. Smith described us as assessing the reasonableness of a given action by comparing it to what we would imagine ourselves doing were we in the same position, with the same personal history. To an extent this is an attitude I want to take from him, the general idea that our ability to imaginatively enter into the position of others is of crucial importance to our knowing how

---

12 Peter Goldie describes empathy as that which occurs when, roughly, one imaginatively assumes the identity of some other person y, feeling as they do as a result of this projection. In-his-shoes imagining, by contrast, is a partial projection of oneself into the other’s position, where one keeps some varying mixture of one’s own values and experiences in the process - I take this to be a slightly misleading division, however, since it implies that a total assumption of another’s perspective is possible at all. I would suggest rather that one’s imaginative projection is always partial, for reasons of economy at the very least. See also Berys Gaut, “Identification and Emotion in Narrative Film”, p. 205.
to be moral. Having said this, I also want to distance myself from the possible implication that our moral evaluation of actors tends to be a straightforward function of our difficulty in understanding them. Smith also put explanatory weight on his division of emotions into the ‘social’ and ‘unsocial’, where these correspond to those emotions which typically are conducive to social cohesion, such as affection, and those not, such as anger. It is therefore tempting to suggest that on Smith’s system, should one be unable to imaginatively enter into the spirit of the one observed, this failure would result in greater approbation if it was a social sentiment (“I cannot believe her generosity!”), greater condemnation if antisocial (“I am dumbfounded by his spite!”). But a random act of petty aggression can at times simply strike one as amusing, even endearing, and a mystifying act of kindness can seem slavish or insincere.  

So this, at any rate, is the crux of my approach. We have a basic obligation to try insofar as we are able to appreciate what is going on around us, for fear of acting unjustly by either having inappropriate reactions or affecting inappropriate responses. The claim here is quite strong – it is not enough simply to refrain from acting wrongfully against someone, or even to know when one’s active help is required – one’s very construal of another’s behaviour, provided at least that it is moralised, does an injustice if it seriously misconstrues his perspective. Since two people, or even the same person on two occasions, can mean something different by what outwardly seems like the same action or statement, since matters of personal history, psychology and culture affect how one constructs or perceives the world, if a moral agent is to live up to such a description they are faced with an obligation to be understanding – an obligation to put effort into trying to appreciate y’s internal landscape, to try to understand as fully as possibly why y did what we judge to be x, because given certain facts about y the idea that he is doing x might be unavailable or highly unlikely to occur to him, and indeed because on fully coming to appreciate his position we might discover it to be more reasonable than our own. Again, this is not to endorse any kind of relativism. Acknowledging that people see things differently, or that different people(s) might have different behavioural norms, is not to suggest that people cannot be deeply mistaken. People can be mistaken about the application of an ethically evaluative term, often in a blameworthy way; indeed this fact justifies the need for understanding. But for an agent to see the world in terms of black-and-white, ‘technicality-based’ moral standards, is to fetishize rules by holding obedience to them over their proper application, like an over-eager census-taker so desirous to leave no head uncounted that he unintentionally includes

---

13 Smith may well have a possible answer for this, but nonetheless it is a naturally perceived problem in his book, and one I wish to avoid. The relation between understanding and Smithian sympathy is examined further in the second chapter.
shop window mannequins. The result is a sham, a crude and empty mockery which defeats the point of the exercise.14

But what might some of the reasons be against lauding this quality of understanding as a central virtue? There are various questions and counterexamples which might well occur to the sceptic here, and we would do well to address them inasmuch as we can before proceeding any further. First, consider the complaint that putting an emphasis on understanding obscures what is important in ethical evaluation, by focusing our attention on the actor as opposed to the act. One might well accept that a person $y$ could himself be acting justly in doing an apparently unjust thing while still believing that what should really matter to our ethical appraisal is that which is being brought about. A consequentialist, for instance, might deem some act $x$ to be good insofar as $x$ typically leads to an improvement in circumstances, while maintaining that $y$’s $x$-ing on a specific occasion was wrong because it led to worse circumstances. Suppose for example that $y$ gives some seafood to a starving man who, unbeknownst to either party, is fatally allergic to such food – we would on this picture be able to direct a moral criticism at someone who, on the face of it, had acted impeccably. What really matters on such a reading are the consequences of an action, where matters of perspective are relevant only in terms of the likelihood a given disposition has to promote the best circumstances. But of course one need not commit to consequentialism in order to have a worry of this kind – without making any significant ethical commitments it seems like a fairly intuitive notion to suggest that, say, if a man is abusing his wife because he genuinely believes he has been wronged and that it is no more than she deserves, the fact that his actions seem reasonable from his perspective should do nothing to alter the fact that hitting one’s wife is wrong and that he is doing this wrong thing.

In reality, however, these considerations have no force against my account. I do not need to make, as indeed I do not make, any special claim to the moral worth of an actor being of ethical primacy. My aim is simply to provide constructive guidance for moral reasoning in the face of an unavoidable pragmatic constraint. Depending on how $y$ understands the world, $y$ will be accountable for at least slightly different things to at least slightly different degrees. Should there be some definite, externally defined fact of the matter about what the moral state of affairs is in a given situation, the relevant actors might very well be accountable in some way for the extent to which their understanding differs from this, but the more misguided someone’s perspective is, the harder it will be for them to understand the real nature of what they are doing. What they are doing can still be described as wrong, just as wrong perhaps, but our assessment of an actor must nonetheless be constrained by considerations regarding his perspective. This is not at all to say that we must overlook what we believe him to truly be doing. So my claim can comfortably be expanded to suggest that

14 Without wishing to seem dismissive, one might suggest that Kant is guilty of this to some extent, by developing a moral law that applies to all rational beings without respect for whatever other characteristics they might have.
there are at least these two dimensions of some moral state of affairs: the moral character of what is occurring, abstractly considered, and the moral character of each y’s actions in their making it such that z. It is primarily matters pertaining to the second dimension with which I am here concerned, but I do not believe the study of one must encourage overlooking the other.¹⁵

Take the domestic abuse case – imagine two similar instances, where the only difference is in the character of the husband. Suppose that in one case the husband had been subject to a grisly childhood and had never been exposed to any signs of human compassion, suppose that his own mother had been convinced, and convinced him at her knee, that the occasional dust up was totally insignificant, and suppose in the other case that the husband was a card-carrying pacifist feminist, who had enjoyed a loving, liberal upbringing. I wish to make no claim as to whether or not the first husband is doing something more or less wrong than the second, but I am making the claim that the husband in the second case has more to answer for than in the first, since it appears we can expect more from him – put simply, I am suggesting that the extent and nature of someone’s blameworthiness need not be logically inseparable from the wrongness of y.

What about a second concern, relating to problems of demandingness? I have made the quite sweeping claim that moral agents should do their best to know what is going on around them – does this leave everyone having to become investigative journalists, spying on their neighbours and generally living in a state of constant vigilance? Does it leave us with a requirement to study psychology, or to read psychological literature? Or anthropology? Firstly, I would say that in democratic countries, or more generally in any country where it is relatively feasible for its inhabitants to affect its politics and social institutions, moral agents do have a basic responsibility to ensure that these are at least minimally just.¹⁶ In practice this already entails a more extensive set of ethical demands than most people usually imagine. It should also be said, though, that there remains some ethical force behind the idea of privacy, and furthermore that one’s constantly trying to moralise could be very destructive, through causing the duties of a moral agent to seem oppressive, intrusive, dogmatic, judgemental and so on.¹⁷ And we do not face an obligation to study psychology insofar as folk psychology seems generally efficacious – by and large people either are or have the potential to be reasonably good at understanding people, or at least people relatively similar to themselves. Naturally, misunderstandings abound in day-to-day life, but normally in a way which could be resolved by clear and restrained conversation, or by putting more effort and practice into the habit of understanding others; the sort of further insight possessed by an expert in psychology need not be

¹⁵ Indeed the discussion of the third chapter shall go some way toward showing how understanding can in fact lead us to construct a new and plausible approach to moral matters and act objectively considered.
¹⁶ See also Thomas Pogge, World Poverty and Human Rights.
¹⁷ Philip Pettit and David Schweikard make a similar point, suggesting that cooperative ventures tend to be undermined by agents constantly monitoring to ensure they are getting a fair return for their effort. See their “Joint Action and Group Agents”, p. 26.
essential to a morally adequate understanding of someone.\footnote{Having said this, if someone’s behaviour is the result of a serious psychological problem then psychologists might be the only ones able to form an appropriate moral assessment of their character. And we do need \textit{some} people to be psychologists, furthermore, in case folk psychology turns out to be drastically inaccurate.} The case for anthropology is slightly different, however, insofar as the practices of a very different culture can leave one utterly at a loss. Saying that moral agents must study anthropology, though, would of course sound ridiculous; clearly some division of ethical labour is called for. However, if one exists in a society where interaction with members of different cultures is likely, particularly cultures with values or a general outlook which seems strange or different, I would say there genuinely is an ethical imperative to be in some way acquainted with them beyond deferring to what is said by ‘the experts’. One need not do anything so specific as studying anthropology, of course, but one should at least be open to and interested in opportunities to improve one’s understanding. I would say at base that the requirement for understanding applies whenever one interacts with, forms a moral evaluation of, or has a moralised reaction to another person or personal activity, where thick understanding is positively \textit{required} only when one fails to be moved by some \(y\)’s position, but is always a good thing – that is, attempting to honestly reconstruct other people’s perspectives in imagination is never inappropriate, although there are times when doing so might be redundant.\footnote{As already alluded to here, and as we shall discuss later, there are however times when it would be inappropriate to imaginatively recreate someone’s perspective past a certain level of detail, largely owing to contingent (and ultimately defeasible) reasons of privacy.}

By personal activity I naturally don’t mean to say private activity – I mean rather any activity persons engage in, as strong opinions about some activity \(x\) tend to induce in one a predisposition to ascribe \(x\)-ers with related qualities, although of course this can also work the other way around. So if for example one has the strong belief that ‘Formula 1 racing is a vapid, childish sport’, one will be inclined to assume that those who participate in or even enjoy F1 racing are themselves vapid and childish, where the latter belief can also be what gives rise to the former. Indeed the two beliefs might form more or less simultaneously, and in any event they are likely to be mutually supporting, so an obligation to be understanding about persons effectively implies an obligation to be understanding about personal activities. But here one might press the demandingness charge again, since surely we cannot help our emotional responses to things. This might be true in an immediate sense, but it would appear that often we can influence our emotions through practice. Where this is not the case I would propose that an indelibly engrained disposition to some emotional reaction \(x\) is not blameworthy at least as long as it is not acted on or expressed for the reason of its wrongness. There is also the question of how one came to be engrained with such a reaction, though – if it was through choice, in circumstances where one knew or should have known better, then \(x\) \textit{does} entail a blameworthy fact about one, although it is a trickier matter to say if any instance of \(x\) specifically is blameworthy, or if rather the general fact of being disposed to \(x\) is blameworthy, or if merely those decisions leading to such engraining are blameworthy.
A third line against understanding pertains to more fundamental problems of epistemic access. Saying we should be understanding is all well and good, the thought goes, but it seems impossible to completely understand anyone – it seems for instance that in every case we cannot with total accuracy imaginatively create a facsimile of y’s subjective experience.\(^{20}\) So what good is it to speak of an ethical requirement that cannot possibly be realised? There are two answers to this question. The first is that we need not necessarily accept the premise that an impossible standard is thereby a futile one. One can still have less or more understanding reactions, one can be a more or less understanding person; the impossibility of perfection does not entail the impossibility of improvement. The second, alternative answer is to say that realistically speaking, while understanding is always appropriate, morally speaking it need only be so complete. The strategy here is simply to renounce the idea of perfection even as an ideal to aim towards – after all, if ought implies can then we have a reason to operate according to achievable standards. Depending on where one draws the line, this need only amount to quite a trivial concession; not all aspects of y’s experience need be relevant to our attempts at understanding, and beyond this an understanding of y sufficient for proper moral appraisal need not be an understanding capable of producing the exact tenor of y’s perspective. And even then, the problem of epistemic access itself does not disprove the moral relevance of people’s perspectives, it merely suggests that we be humble and cautious in our ascriptions, as I have already said that understanding recommends.

Fourthly and finally, one could also try to attack the claim that understanding is always appropriate. Perhaps some acts are so terrible we would do their gravity an injustice by moderating our reaction to those responsible. And there might well be a class of wrongdoing which is spontaneous; it could be that some wrong acts are done on the spur of the moment, without thinking, reflexively, or in some other manner which obscures the usefulness of understanding the relevant actor’s perspective. What too about unintended wrongs of neglect? A negligent factory owner might not intend the slightest harm by forgetting to inspect his factory’s safety measures, for instance, but in the event of an accident should we entertain his protestations of innocence with understanding? However, I maintain that it simply cannot ever be ethically responsible to ignore the way in which y acts, in a total sense, in doing something. It is important to remember not just that people should be evaluated in light of their understanding of the world, but also that people should be held accountable for the extent to which they failed to ensure their understanding of the world was adequate, particularly in cases which should have seemed important. So a deluded Nazi or a negligent factory owner are still blameworthy even if it made sense from their perspective that their behaviour was

\(^{20}\) Even if it was possible, what is worse, there does not seem to be any way we could know for sure when we had been successful, no way we can hold the two experiences beside each other. Here I think the best we can hope for is that people’s behaviour goes on to be as we would expect given our understanding; this is a worry so fundamental and pervasive, effectively reducing to the problem of other minds, that it need not fall on this account to resolve it.
right. How this cashes out will depend on the seriousness of the case and the difficulty one would have in standing in y’s position and reasoning to the point where one understood oneself to be wrong. In general, then, the less apparent it is that one is in a situation with high moral stakes, and the more misguided is one’s perspective, the less blameworthy one will be in acting wrongly. It should be noted that these two factors could easily feed into one another – theoretically at least, a brain-washed Nazi could be so deluded that it non-culpably fails to occur to him that a camp-bound train of Jews does present a situation with high moral stakes. But practically speaking I think some kinds of evidence are too basic and tangible for one to defensively overlook them. And as far as wrong-doing from spontaneous/reflexive action is concerned, I would say thick understanding could still be appropriate to the extent that it requires us to acknowledge that it might not make sense to describe y as being responsible for the resulting x – indeed it might even require us to pity them for being the unhappy instigator of x. Even then, thin understanding is clearly also required to make sense of how people can do certain things without really meaning to or thinking about it.

So, we have established what understanding is and that it is a crucial characteristic and a crucial skill for the moral agent. Where do we go from here? In the remainder of this chapter we shall look at two concepts closely related to and very important for understanding, the concept of identification as discussed in the philosophy of art and the concept of simulation as discussed in the philosophy of mind. Both of these phenomena involve something like in-his-shoes imagining, an essentially transpositive or projectional form of understanding whereby one imagines oneself, consciously or otherwise, as being another person to one extent or another. This will flesh out the imaginative and particularly motivational or affecting element of understanding, providing us with a picture of the psychological machinery at work as well as laying the groundwork for the fourth and fifth chapters in particular, in which we shall examine the importance of non-linguistic understanding and the usefulness of fiction for the exercising of imaginative capacities typically central to one’s ability to competently understand and care about others.

1.2 Character Engagement:
Identification has been a much disputed concept. Some people, for instance Noël Carroll and Gregory Currie, hold the notion of identification to be mistaken, confused and misleading. Others who defend identification, such as Alex Neill, Berys Gaut and to an extent Murray Smith, disagree or part ways in how they describe or categorise what sub-processes are involved. Nonetheless it seems something very much like it forms an important part of understanding. In the present section we shall examine

---

21 This is perhaps better put by suggesting that one imagines being in the other person’s position, where this includes having some or all of their properties, rather than imagining that one is the other person. This is the sort of move Berys Gaut makes, for example, in the face of Wollheim’s complaint that the symmetrical nature of an identity relation would leave one imagining one was character x and that x was, counterintuitively, oneself. See Richard Wollheim, *The Thread of Life*, p. 75 and Berys Gaut, *A Philosophy of Cinematic Art*, p. 256.
some of the debate surrounding the concept of identification, in order to discern what exactly it is thought to involve and to weigh what reasons there are for and against believing in or appealing to it. While most of the discussion ahead will refer to films as examples, it should not be forgotten that identification is generally taken, at least by those who believe in it, to occur both in engaging with all forms of fiction (or all fictions containing characters, at any rate) and to social interaction in real life, although this of course is not to deny that one might fail to identify with any of the characters in a given story or scene.

Let us begin by looking at some reasons against believing in identification. Perhaps the most sustained criticism of the concept, at least in the field of aesthetics, has been offered by Carroll, and it is to his early objections that we shall now turn. In *The Philosophy of Horror*, he suggests that identification refers to the duplication in someone of the emotional state of another, since identification must describe a symmetrical relation between the viewer and viewed, or there is no useful reason to refer to whatever process as ‘identification’. It cannot involve more than the duplication of emotional states specifically, he says, because a complete duplication of another’s inner states (and thus an illusion that one is oneself the character one is watching) would leave mysterious the fact that horror fans do not sharpen stakes while watching Dracula, or run screaming from cinemas. He suggests for similar reasons that identification cannot involve a duplication of the observed character’s beliefs. But even this apparently moderate form of identification does not pass muster in Carroll’s eyes, since the relation between the feelings of the audience and the feelings of the fictional characters seems at all times to be an asymmetrical one. There are basically two ways in which this is the case, for Carroll. Firstly, he suggests that the audience’s emotions tend to be different in kind from those of the characters they observe, because the audience and fictional characters are so differently situated in relation to the fictional events the story involves. We often have different information about a fictional situation than the characters contained therein do, as when we fear for the enraptured yet doomed swimmer at the beginning of *Jaws*. Similarly, he claims, we feel suspense as we watch someone being attacked by a monster, even though the character would not have the time or freedom to indulge such an emotion. Secondly, Carroll contends that the audience’s feelings are always different in direction from those of the fictional characters. As we watch someone stumble blindly through a dark and dangerous scene, his claim goes, the fear we feel is purely other-directed, rather than self-directed as the fictional character’s emotions presumably are.

His alternative explanation of what happens when we engage with fiction he calls ‘assimilation’. The idea here is that one has some ‘sense of the character’s internal understanding of

---

22 Noël Carroll offers some further objections in his *The Philosophy of Motion Pictures*, but as these are addressed more or less explicitly at simulation theory and fiction we shall examine these complaints in the next section.


24 Ibid., p. 90.

25 Ibid., pp. 90-91.
the situation’ without having to duplicate their minds in oneself, while at the same time one views the action externally, as ‘a situation involving a protagonist who has the viewpoint she has’, where to gain an internal understanding of the character ‘we need only have a sense of why the protagonist’s response is appropriate or intelligible to the situation’ (ibid., p. 95).\(^\text{26}\)

But Carroll’s criticism is far from satisfactory – he attacks identification by means of defeating a series of straw-men, before claiming for himself a slightly murky concept which seems to involve something suspiciously similar to what identification theorists generally talk about. The first thing to note is that while, etymologically speaking, the term ‘identification’ might be seen to imply a symmetrical, double relation, this is not to prove that such is the meaning actual English speakers intend when using the word. As Gaut points out, this line of attack is rather like arguing that televisions do not exist because they do not really involve seeing things at a distance.\(^\text{27}\) This is an important defence, as criticisms of identification usually or always turn on some such overly simplistic illusion/duplication interpretation of the concept. For this reason, indeed, the best defence of identification is to describe the idea clearly and fully.

Identification is not a joining of minds but a case of imaginative projection, it does not signify some kind of identity mistake on the part of the audience-member (though the audience may sometimes, if rarely, become ‘lost’ in a character) but an attempt, sometimes conscious and sometimes not, by a spectator to enter into and thereby better understand and judge the feelings or perspective of another. It is, to use the old metaphor, to put oneself in another’s shoes, to imagine how one would relate to a situation had one led the life of the one observed. In this sense identification does not entail complete emotional resonance, it does not demand that we feel in imagination as the observed other seems to, although the ability to identify with someone does imply that one finds their perspective to be at least reasonably intelligible.\(^\text{28}\) To find it impossible to identify with someone, at least where this is not the result of insurmountable differences in either person’s experiences, is normally to judge that their perspective is anomalous to one extent or another, in that their feelings seem unwarranted or irrational relative to the relevant person’s character and circumstances – although it might also be because one simply finds a character too uninteresting or their situation too bizarre to allow the imaginative/emotional effort.\(^\text{29}\) It should also be noted that one’s capacity for

\(^{26}\) Ibid., p. 95.

\(^{27}\) Berys Gaut, “Identification and Emotion in Narrative Film”, p. 202. Presumably one might not approve of this counterexample were one to believe in the transparency of film, but we need not go into this here. See for example Kendall Walton, “Transparent Pictures”.

\(^{28}\) This differentiates identification from the related notion of empathy somewhat – depending on whom one reads, this necessarily involves experiencing either the same or a congruent emotion to the one observed. There is some disagreement as to whether or not empathy need involve identification, but the prevailing opinion is at least that whatever empathetic emotion one feels, in order for it to count as empathetic it must be in some way directed by the emotion of the other.

\(^{29}\) It should be noted here that certain irrational feelings can be a deeply-engrained part of someone’s character, and so that we could find some irrational behaviour to be reasonable; if someone has an irrational phobia toward rabbits, for instance, we could still think of their fear-behaviour (within limits) as not being unreasonable. The
identification is limited by one’s natural ability, and so that one might fail to identify with others simply because one is not imaginatively equipped for the work required. To the extent that this might be true of someone, it would on our picture represent a serious handicap to their abilities as a fully-fledged moral agent, which will become more and more apparent as we go on.

Identification, then, allows us to form a strong emotional reaction to another’s situation, based on their place in it but not wholly determined by what their behaviour suggests is their own reaction, since imagining ourselves to be in a given situation helps to ‘bring home’ and make vivid its significance. Let us return now to Carroll’s complaints. His implication that we feel suspense rather than fear when watching someone being attacked by a monster simply seems false from experience. His argument that the fear we feel while watching a horror film is entirely other-directed rather than self-directed is circular in simply assuming that identification does not occur, and furthermore conflicts again with the phenomenological evidence. His claim that we as spectators react to a total situation, containing characters whose perspective we have ‘some sense of’, rather than to a situation in which we imagine ourselves from the characters’ perspectives, seems as hasty and unfair a way to deny identification as declaring that a chocolate cake cannot contain flour since the cake is moist and flour is not – it is not the contention of identification theorists that a fictional situation is only understood through the perspectives of its characters, and it would take an implausibly compartmentalised picture of human thought to believe that what emotion or information one gleans from imaginative projection is not affected by or integrated into a broader analysis of the observed situation in which it occurs.

What, then, are some of the reasons for believing in identification? It seems as though there are a number of prima facie illustrations of identification, or something very like it, doing good work. It is a commonplace, for instance, to observe that much critical talk about fiction involves the notion of identification. Not only do people talk about characters in terms of the degree to which they can identify with them, however, introspection should also reveal that when one is trying to understand some case of puzzling behaviour one naturally tries to imagine how things must seem for the one whose behaviour is under examination. Of course, it is not the case that one must always engage in imaginative projection in order to understand the behaviour of others – especially, perhaps, when such behaviour closely conforms to the social norms and expectations one comes to possess over time. This points to two further, similar roles for imaginative projection, namely those of reading ambiguous behaviour and predicting behaviour in others. Neill further points out that empathic identification with another can ‘play an important part in the education of emotion’; if we are moved to feel the same way about some set of events as another does, even though we ourselves would have felt differently in

fear itself in such an instance would not be an object of our estimation so much as the person’s capacity to deal with this fear.
their place, this can expand our horizons and help us develop our own perspective.\(^{30}\) As Feagin remarks, it takes special skill on the part of the author to do this, but despite her relative pessimism about the edifying potential of literature we should remember this as an important effect fiction can have, while still bearing in mind that it also has the potential to make us \textit{worse} by reinforcing stereotypes and so on.\(^{31}\)

So, overall, identification, like sympathy in the previous section, looks like a promising ingredient of optimal moral appraisal. But it, too, suffers from something of an in-built tendency to prejudice, and beyond this identification is generally construed as imaginative, where the discussion thus far has suggested that there might be other important, non-imaginative elements involved in understanding. It is because of this fact, that identification implies imaginative projection while it is not the case that our entire understanding of fictional characters (or real people) is gained from such imaginative projection, that we should move the focus of our study from identification alone to the broader notion of \textit{character engagement}. A good place to start is in describing Murray Smith’s work in \textit{Engaging Characters}, as it is perhaps the most thorough and empirically informed study of character engagement available – thus it is both useful for a further filling out of how we understand others and as a bridge to our coming examination of simulation theory in the next section, the most empirically conscientious element of understanding’s support here.

Smith ascribes more or less the same meaning to ‘empathy’ as we in the above have ascribed to ‘identification’, and this he divides into two parts: involuntary mimicry and voluntary simulation, where the latter is properly the imaginative process we have been discussing, and the former is a set of non-imaginative factors which can contribute to the accuracy of imaginative projection.\(^{32}\) He holds that the consumer of (film) fiction is active; in engaging with fiction one constructs an interpretation of it by means of whatever ideas and norms one possesses or is familiar with. That is to say one is not, he believes, deterministically ‘positioned’ by the text. At the same time, Smith acknowledges that the text does \textit{constrain} one in one’s interpretation, further adding that one need not be ‘entirely caught within [one’s] cultural assumptions’.\(^{33}\) This last thought makes way for him to employ the concept of schemata, an explanation of the cognition of perception in terms of continually adjusting patterns with which the mind categorises, and forms hypotheses about, the sensory data to which it is exposed. Many of the schemata one employs, consciously or otherwise, are culturally specific, and to this extent at least one’s interpretation is influenced by external social factors. Indeed many schemata may be ‘automatized’ through repeated use to the point where they become so engrained as to be

---

\(^{30}\) Alex Neill, “Empathy and (Film) Fiction”, pp. 179-180.


But as he explains, it is still in the very nature of schemata to be subject to revision; as one encounters different scenarios and experiences new things one is stimulated to revise mistaken or incomplete models. One might on the strength of past experience expect the children in one’s shop to buy sweets, but once some children also start to buy tea one must revise one’s understanding of the situation in order to better predict or map such scenarios. Of course, a given schema can also entail ‘hierarchically ordered alternatives’ as to the manner in which something might unfold.\(^\text{35}\) ‘Assimilation’ is when new information is absorbed and processed by pre-existing schemata. ‘Accommodation’ is when schemata are revised in the light of new or unexpected information. Rejection, though, can also occur, when a schema or set of schemata are abandoned entirely. Smith also notes that information can also either be discarded or go entirely unnoticed because it does not fit with strongly held schemata. In terms of interpersonal interpretation, these observations could be subsumed within the kind of theory theory account which will be discussed in the next section.

For Smith, when engaging with narrative fiction one performs, broadly speaking, two kinds of imagining: central and acentral. Central imagining is that where one imagines a scene from the inner perspective of a character. Acentral imagining, by contrast, is the imagining of a scene from a character-neutral, ‘external’ standpoint within the scene. It is Smith’s position that the understanding of narratives involves both kinds of imagining, and to either one he ascribes a respective set of processes, ‘empathy’ and what he calls ‘the structure of sympathy’.\(^\text{36}\) Empathy, for Smith, involves three processes: emotional simulation, motor and affective mimicry, and autonomic reaction. Only the first of these three is voluntary, and involves an imaginative assumption of the other’s position, sometimes including a sort of ‘affective trial and error’ by which we decipher behaviour by rehearsing various possible emotions and then selecting the one we find most appropriate.\(^\text{37}\) The other, involuntary components of empathy (automatically mimicking gross emotional states or physical movements, as well as making the same reflexive responses as some character to something in the fiction) are not examples of central imagining as such, but they are important contributing factors to one’s ability to empathise with characters.\(^\text{38}\)

Smith’s structure of sympathy, meanwhile, is comprised of three elements – recognition, alignment and allegiance. In recognition, one picks out and/or separates continuous (rather than transparent.\(^\text{34}\) But as he explains, it is still in the very nature of schemata to be subject to revision; as one encounters different scenarios and experiences new things one is stimulated to revise mistaken or incomplete models. One might on the strength of past experience expect the children in one’s shop to buy sweets, but once some children also start to buy tea one must revise one’s understanding of the situation in order to better predict or map such scenarios. Of course, a given schema can also entail ‘hierarchically ordered alternatives’ as to the manner in which something might unfold.\(^\text{35}\) ‘Assimilation’ is when new information is absorbed and processed by pre-existing schemata. ‘Accommodation’ is when schemata are revised in the light of new or unexpected information. Rejection, though, can also occur, when a schema or set of schemata are abandoned entirely. Smith also notes that information can also either be discarded or go entirely unnoticed because it does not fit with strongly held schemata. In terms of interpersonal interpretation, these observations could be subsumed within the kind of theory theory account which will be discussed in the next section.

For Smith, when engaging with narrative fiction one performs, broadly speaking, two kinds of imagining: central and acentral. Central imagining is that where one imagines a scene from the inner perspective of a character. Acentral imagining, by contrast, is the imagining of a scene from a character-neutral, ‘external’ standpoint within the scene. It is Smith’s position that the understanding of narratives involves both kinds of imagining, and to either one he ascribes a respective set of processes, ‘empathy’ and what he calls ‘the structure of sympathy’.\(^\text{36}\) Empathy, for Smith, involves three processes: emotional simulation, motor and affective mimicry, and autonomic reaction. Only the first of these three is voluntary, and involves an imaginative assumption of the other’s position, sometimes including a sort of ‘affective trial and error’ by which we decipher behaviour by rehearsing various possible emotions and then selecting the one we find most appropriate.\(^\text{37}\) The other, involuntary components of empathy (automatically mimicking gross emotional states or physical movements, as well as making the same reflexive responses as some character to something in the fiction) are not examples of central imagining as such, but they are important contributing factors to one’s ability to empathise with characters.\(^\text{38}\)

Smith’s structure of sympathy, meanwhile, is comprised of three elements – recognition, alignment and allegiance. In recognition, one picks out and/or separates continuous (rather than

---

\(^{34}\) Smith refers to an example of transparency as being when for instance the custom of shaking hands becomes so familiar that one no longer thinks anything of performing it, or indeed that one comes to notice its omission.

\(^{35}\) Murray Smith, Engaging Characters, p. 51.

\(^{36}\) While in Engaging Characters Smith stresses that central imagining plays a less prominent, though by no means insignificant, role, in “Imagining from the Inside” he seeks to amend his position by, among other things, saying that central imagining allows for ‘a qualitatively different [set] of emotion[s]’ (1997, p. 426). The upshot of this is that central imagining is recognised to have a unique ability to allow us to (try to) comprehend different ideological or experiential perspectives.

\(^{37}\) This idea is taken from Robert Gordon, one of the founders of simulation theory.

\(^{38}\) This is a clarification Smith makes in “Imagining from the Inside” (ibid., p. 416-417).
coherent) characters. In alignment, the two factors involved are spatio-temporal alignment (where the camera, say, follows a character, or when a character gets screen time) and subjective access (where we learn more about the inner perspective and mental states of characters, in whatever way). In allegiance, one constructs a hierarchy of preference for the characters contained in a narrative, typically determined by the extent to which one either approves of a character or approves of a character in comparison to others in the same narrative. It is worth noting that though these processes are described as acentral, they affect and are affected by one’s central imagining – Smith of course stresses that all of these phenomena are interdependent.

Although Smith’s account is both convincing and comprehensive, it does suffer from one glaring problem which prevents us from accepting it wholesale; he cashes out his concept of allegiance in overly moral terms. Somewhat puzzlingly, Smith insists on the psychological claim that unless one is deliberately trying to be perverse, one will always be most in favour of those characters in a fiction that one thinks are morally best, where this can mean not only morally good in an ultimate sense but also morally good in comparison to other characters in the same story. This is quite a strong claim to make, and furthermore is one which seems to conflict with experience. In a footnote he claims that he was motivated to adopt this stance because ‘once we accept the possibility of perverse allegiance, it becomes much more difficult to prise apart alignment and allegiance, because every alignment can be said to produce allegiance’. But again this does not seem true.

Alignment is merely that process by which one is granted greater access to the subjective position of a character, while allegiance is more to do with the feeling of sentimental loyalty one has for a character (not meaning the word ‘sentimental’ in a pejorative sense). Allowing for perverse allegiances does not appear to do anything to muddy the distinction between the concepts of alignment and allegiance, it only makes it more difficult to predict which characters in a story one will most favour. But this seems a harmless result; in fact it could even be suggested that this is an improvement on matters, since a theory which, by implication at least, purports to know in advance which characters everyone will find most sympathetic is probably an inadvisably over-confident one.

Be that as it may, Smith’s 1999 article “Apparently Perverse Allegiances” is an attempt to shed light on this issue, which he avoided in Engaging Characters. As the title suggests, Smith does not change his position on what determines allegiance so much as he claims that what apparent experience one might think counts against him is mistaken – essentially, that ‘we feel an allegiance

---

39 A character who loses her personality, say, is not coherent to the extent that her qualities change drastically throughout a narrative, but there is still some continuous character to whom all of a set of shifting qualities belong.

40 A mistake which is arguably shared by Noël Carroll – see The Philosophy of Motion Pictures, p. 181. Carroll goes from the plausible claim that movies tend to have generic or dumbed-down morals for the sake of wider appeal, to the regrettably strong claim that protagonists are portrayed as ‘morally righteous’.

with... a character [not] because of the perverse act that they engage in [but] in spite of that act. He gives five factors which contribute to what we suppose to be perverse allegiance. First, he says immoral characters with whom we align ourselves are typically ‘diamonds in the rough’, that we enjoy a safe thrill directing our emotional support or attachment to characters whom we imagine to be bad but know secretly to be good when it really matters. Second, he suggests that bad characters do not prompt feelings of allegiance because of their negative qualities, but because of other positive qualities that they display, as with our liking Dr. Hannibal Lecter for his wit and charm yet deploring him for his murderous cannibalism. Third, he says that many films which encourage one to sympathise with amoral or immoral characters, such as Pulp Fiction, are in fact taking an ultimately moral stance by making some comment on the moral establishment, as it were. Fourth, he claims that people have an amoral fascination in the strange, and the familiar in the strange, and so that people are interested by evil characters either on the basis either of how extremely alien or unsettlingly familiar they are. Fifthly and finally, Smith makes a vague gesture at the end of his article to the idea that fiction can function as a safe outlet for urges and interests which society proscribes, as well as thereby reminding one of what it is that is good about acting morally.

Claims one and three seem fair enough, and can be granted without remark. Claims four and five seem to be true, yet they do not seem to do anything to defend Murray’s position, indeed if anything they point to the fact that a system as moralising as Smith’s is not a sustainable one. It is his second claim, about liking evil characters not for their evil or bad qualities but for their good ones, which is most interesting. In the course of making it, he also appeals to the concept of ‘star charisma’, saying that often we sympathise with evil characters in films because they are played by famous actors, where this either heightens our awareness that we are watching a fiction, and so that normal behavioural standards are not in effect, or ‘dilutes’ the negativity of the character by invoking ‘a fund of positive associations’. One should immediately notice here that this is a poor defensive strategy, insofar as it is left completely mysterious how we might have reason to sympathise with evil or bad characters in literature as opposed to in film – an example perhaps being something like Robert Lovelace in Clarissa, or Shylock in The Merchant of Venice. It is also problematic inasmuch as one normally assumes famous actors are famous because they act well, in which case one is left with the

---

42 It should be said that Smith admits genuinely ‘perverse’ allegiances – that is, structures of allegiance which do not match the ‘moral structure’ of the story – can sometimes be formed, but he still maintains this is almost never what is really going on. Ibid., p. 222.
43 Ibid., pp. 223-225.
46 Ibid., p. 234.
48 Ibid., p. 227.
49 The presumption here being that one reads The Merchant of Venice – although even then this is perhaps not a good example insofar as Shylock’s machinations are to some extent justified or at least intelligible.
odd idea that convincing performances make for unconvincing performances.\footnote{This is not, admittedly, an entirely fair point, since conceptually at least it seems possible for a performance to at once be emotionally convincing or affecting and strikingly fictive.} Furthermore, actors are frequently typecast, and to this extent a familiar actor in an evil role could just as easily carry a fund of negative associations – even then, it is far from clear that likeable evil characters are always or even usually played by well-known actors; Smith in this regard has at best found quite a minor contributing factor.

Despite this, the idea that evil characters are only likeable for their relatively good qualities remains an interesting one, especially when the good qualities concerned are tied up closely with the bad – imagine an extremely witty villain who only made cruel jokes while harming the innocent, where his jokes always centred on his immoral acts. The idea of a cruel joke being bad to the extent that it is cruel and good to the extent that it is witty can be puzzling since, if the wit depends crucially on the very cruelty of the remark, it can seem all but impossible to separate the two.\footnote{See also Berys Gaut, “Just Joking: The Ethics and Aesthetics of Humour”.

\footnote{Indeed, this offers something of an answer to Currie’s problem of personality, an observation regarding the fact that we often sympathise or like characters in fiction whom we very much would not in real life. See Gregory Currie, “The Paradox of Caring”, p. 65.}} We can set this problem aside for now, however, in that Smith’s claim still seems to conflict with experience; sometimes, in fiction at least, it very much is the negative qualities of a character which attract one to him or her. Take for instance the television series Battlestar Galactica: arguably its most memorable and likeable character is Dr. Gaius Baltar, an utterly reprehensible coward who repeatedly stoops to new lows in an effort to keep himself safe and comfortable. Through his negligence and greed the human race is nearly wiped out at the beginning of the show, and as time goes on he continues to do everything imaginable to disgrace himself further. There is then at least one phenomenon which Smith has overlooked, as this example will suggest to science fiction fans. A character can stand out from its neighbours in a fictional landscape, and earn our allegiance, purely because it is the most interesting or insightfully-rendered character on offer.\footnote{One could in fact concede this point even while rejecting cognitivism, or indeed moralism.} The importance of this fact will readily strike those who support cognitivism as an explanation of fiction’s value. To the extent that people pursue some kind of emotional or experiential learning when engaging with fiction, they will be attracted to characters from whom they can learn, and it is obvious that we can often learn more from examining people’s flaws and mistakes than their virtues and triumphs. The strikingness with which a character is rendered has little or nothing to do with their moral worth.\footnote{One could in fact concede this point even while rejecting cognitivism, or indeed moralism.} Essentially, art attracts interest, and interest tends to foster allegiance. But we shall more closely examine the differences in our fiction-responses and nonfiction-responses, together with their possible ethical implications, in the fifth chapter. For now, though, we shall next explore how something like identification has been defended under the banner of ‘simulation’.

\footnote{See also Berys Gaut, “Just Joking: The Ethics and Aesthetics of Humour”.

\footnote{Indeed, this offers something of an answer to Currie’s problem of personality, an observation regarding the fact that we often sympathise or like characters in fiction whom we very much would not in real life. See Gregory Currie, “The Paradox of Caring”, p. 65.

\footnote{One could in fact concede this point even while rejecting cognitivism, or indeed moralism.}
1.3 Simulation Theory:
For much of the twentieth century, philosophers of mind and cognitive scientists saw people’s everyday understanding of one another as operating according to a *theory of mind*. The precise manner in which such a stance – generally referred to as ‘theory theory’ – is to be spelled out varies between different theorists, but it is possible to give a description that while general enough to apply to the overwhelming majority of cases remains reasonably informative. For the most part, we can say that such an approach portrays people as forming determinations about, for example, the inner states or future behaviour of others on the basis of a non- or semi-conscious, systematic knowledge structure of inter-related, law-like generalisations which are often described as being sentential in nature and as defining the terms they introduce (such as ‘belief’, ‘desire’ and so forth – this is a nod to the functionalist idea that to understand any term in a theory one must know about the ideas with which it connects or amongst which it is situated). Although the analogy is not universally endorsed, it is also commonly suggested that such a theory of mind bears significant structural similarities to scientific theories, as might perhaps seem natural from the description just given. Furthermore, this knowledge structure, though it is presumably affected to some extent by one’s experiences, is most often supposed to be essentially innate in much the same way that Chomsky argues certain fundamental rules of language are innate.

In his 1986 article “Folk Psychology as Simulation”, Robert Gordon argued that a different sort of process was involved in people’s understanding of one another. His suggestion was that in attempting to properly interpret someone else one does not employ or consult some series of nomological generalisations about psychology but instead that one, in imagination, assumes their position. This sort of position has some precursors, Gordon himself mentioning R.G. Collingwood, the *Verstehen* approach to the social sciences, and elsewhere David Hume and Adam Smith.\(^{54}\) Just as with theory theory, proponents of this alternative approach – which has since gained the general classification of ‘simulation theory’ – vary in the terms by which they cash out the nature and implications of the process involved. For a time it was often assumed (primarily by theory theorists, it must be said) that the two theories, henceforth TT and ST respectively, were mutually exclusive, but in the last decade or so most commentators have come to believe that some combination of the two is possible, if not probable, as different situations seem more appropriate to one or the other, and because the two can be said to regulate or otherwise feed into one another.\(^{55}\) Gordon takes a somewhat stronger position, though, to the extent that he believes all of people’s intrapersonal interpretation is essentially simulative, where he does not see this as precluding the idea that people

---


\(^{55}\) One noteworthy turn around, for instance, being Stich and Nichols, initially some of ST’s most bitter opponents, who later constructed a hybrid account in *Mindreading*, 2003.
might form generalisations as rules-of-thumb on the basis of accrued simulative experience. The majority of early debate revolved around empirical findings in developmental psychology, where theorists were eager to show that either ST or TT was more apt to explain findings than the other, and hence to prove at least which was more fundamental to human psychology. For present purposes however we need not examine the debate closely here, inasmuch as we need merely to establish the possibility and likely workings of simulation; where psychological generalisations are concerned it does not seriously affect the thesis here presented if the sorts of rules people consciously or otherwise employ are more properly thought of as comprising a theory or merely a set of simulation-derived rules-of-thumb.

It is worth noting one area of scientific research, though, before discussing ST directly, which is the work on mirror neurons done by neuroscientist Vittorio Gallese and others. These are neurons in human and primate brains, among others, which fire in a pattern corresponding to that in which an observed member of an anatomically similar species would in performing actions of certain kinds. In effect, it has been suggested to amount to physiological evidence of the brain mimicking observed behaviour and thus, it is supposed, evidence supporting ST. Similarly to the cases arising from developmental psychology, though, these findings are in some dispute, and beyond this give rise to the need for some further clarification of the concept of ST. For now, though, we shall turn to an exploration of some of the earlier arguments and examples which have been given for and against ST, in order to elucidate how feasible it is as a general explanatory strategy for the workings of intrapersonal, and indeed interpersonal, understanding. Once its basic appeal has been subjected to some scrutiny we shall return to consider what motor neurons and other recent research mean for ST.

Gordon begins “Folk Psychology as Simulation” by considering how reliable our self-predictions are, making the oddly unsupported claim that such reliability could not be provided by a system of psychological laws. He initially talks about predicting one’s relatively immediate actions once one has ‘made up one’s mind’, which he suggests are predictions resulting from practical reasoning, which in turn he suggests provides a bridge to the idea that simulated practical reasoning can be used as a predictive device. He illustrates this by giving the subsequently oft-cited example of

56 See for example Robert Gordon, “Folk Psychology as Simulation”, pp. 65-66 and “Sympathy, Simulation, and the Impartial Spectator”, p. 736. Alvin Goldman’s own stance is in this respect very similar, although he is comfortable in granting the theory theorist that such generalisations might sometimes fail to be the result of simulation – see for instance “Interpretation Psychologised”, p. 88.

57 This phrasing might be considered a little misleading – it is not the pattern in the observed subject’s own firing of neurons which is being mimicked, but rather the observer’s neurons fire in much the same way they would were the observer to reproduce the action they observe. Hence the qualifier ‘anatomically similar’; such basic mimicry may occur when one observes a non-human animal engaging in an action the human body can reproduce, but not with actions it cannot. For instance, one might on a neural level mimic the observed action of a dog or monkey biting, but one does not do so in the case of a dog wagging its tail. See for instance Alvin Goldman, Simulating Minds, pp. 122-123.

one trying to work out what one’s response would be to the sound of footsteps coming from one’s basement. The idea is that such a question need not or is not to be considered in terms of abstracted rules, but instead that one need merely create the situation in one’s imagination and, in the context of one’s imagined scenario, decide what to do. To the extent that one’s simulated situation captures the relevant details (we shall return to this stipulation, which Gordon only hints at in the article presently under discussion, a little further on), one’s decision should be the same in the actual event – indeed, one can even learn surprising things about oneself in the context of such imaginings, as Gordon points out. He then extends this idea to the prediction of others as such, citing the case of chess players predicting their opponent’s moves by imaginatively assuming their position and considering what strategies they would formulate in such a position. He calls such reasoning ‘hypothetico-practical’, going on to further extend it to interpretations of others’ present behaviour, too.

As an additional observation we might offer, in order to make the transition from imagining oneself in counterfactual situations to imagining oneself to be someone else seem more intuitive, it is worth pointing out that one can, and reasonably often one implicitly does, imagine oneself as having different qualities. Imagining oneself to hear the sound of footsteps is one thing, but consider imagining, in one’s youth, what one will do when one retires or has children, for example, at some point in the relatively distant future. Such an imaginative project requires that one take into account the fact that certain aspects of one’s character may become different over time. Suppose one is at the beginning of one’s academic career – writing is a joyful form of self-expression, and one feels eager to help one’s students with their studies. At the same time it seems prudent for one to save money for one’s retirement, but what sort of thing will one be interested in pursuing when the time comes? Such a question is worth asking to the extent that it will for instance have some bearing on the amount of money one should set aside. So one might wonder if one is likely to continue giving lessons in a part-time or unofficial capacity after one’s retirement, which could provide an additional source of income. Having spent a lifetime writing and teaching, however (in imagination or otherwise), one might find oneself to in fact be quite bored and disgusted by the academic scene. In such a scenario one is not only imagining oneself in a different state, as when one imagines being confronted by different and frightening practical concerns in the footsteps case, but as actually being quite a different person.

---

59 Robert Gordon, “Folk Psychology as Simulation”, pp. 61-63. It should be noted that describing this process in terms of the use of one’s imagination is due to Gregory Currie (see his “Imagination and Simulation: Aesthetics Meet Cognitive Science”), but I cannot see that Gordon would have an objection to the use of the term per se, or at the very least as it is employed here.

60 Ibid., p. 63. Much the same claim is later made by Kendall Walton; given the right anxieties, imagining spelunking can induce a revealing level of panic. See “Simulation, Spelunking and Slime”, pp. 38-9. However, it should be borne in mind that there could be a significant difference between one’s expected and one’s actual reaction to some course of events. Consider how attempted or conspiracy to murder is punished less harshly than actual murder; part of the rationale for such a difference is surely that one cannot know one will in the event be able to perform certain acts. The simulation theorist can absorb such a consideration, however, by suggesting that such unreliability applies primarily to highly unusual circumstances.

61 Ibid.
Similarly, one might daydream about being a veteran of a third World War, or an interstellar diplomat thirty years from now, in a radically different future. So really there should be nothing strange about the idea of imagining oneself as an observed other person, as potentially having entirely or fundamentally different qualities.

At this point one might raise the objection Goldman and Gordon respectively mention, that, while it might seem reasonable to suppose we sometimes attempt to ‘put ourselves in someone else’s shoes’, it does not from introspection appear as though simulation is the standard means by which people interpret one another. The initial response both Goldman and Gordon give is that any account of interpretation, be it ST, TT or otherwise, must appeal to the idea of processing going on beneath one’s conscious awareness, and as Goldman remarks, ‘it is a psychological commonplace that highly developed skills become automatised’. Their responses differ, however, in that Goldman goes on to allow that many cases of seemingly instant interpretation may in fact be the result of one employing some already possessed schema, which he then suggests would normally have been constructed or ‘historically derived’ from past simulations. Gordon, on the other hand, suggests that in fact what one might think of as simulation proper does not in fact occur most of the time. Rather, he claims that ordinarily one performs a more economical sort of total projection, where one simulates the other person’s perspective but without adjusting for any differences in their character.

Stich and Nichols have argued against the possibility of ST by offering examples of simulation failing to predict behaviour in a manner which, they suggest, it should not. If simulation is ‘process-driven’, as Goldman describes it, then one’s simulations tend to be accurate because the systems they emulate bear important similarities to one’s own internal systems. That is, one does not need a theory about the minds of others in order to construct a working simulation of another’s perspective, one need merely run a simulation on oneself and take note of the outcome – one does not need a sophisticated understanding of psychology to notice how one feels when imaginatively assuming someone else’s position. Stich and Nichols refer to this, the idea that simulation need not or does not involve an understanding of how minds generally work, by saying that ST is not ‘cognitively penetrable’. TT on the other hand, even if thought to be largely innate, obviously must to some extent be cognitively penetrable; as Stich and Nichols say, ‘if there are some unexpected aspects of the system’s behaviour that are not captured by our principles, then our predictions... should be less accurate’. Thus they take it as an argument against ST and for TT if they can show that running

---

64 Ibid.
65 Ibid., p. 85.
66 Stich and Nichols, “Folk Psychology: Simulation or Tacit Theory?”, p. 150. They admit in a footnote that this is not a standard use of the term.
67 Ibid.
simulations yields false predictions of behaviour resulting from a failure to include some universal aspect of human psychology; a theory may be incomplete, the thought goes, but a process-driven simulation should have basic psychological tendencies in-built. They offer three such examples of (supposed) false prediction, listed below in condensed form.\textsuperscript{68}

CLOTHES: You are in a shopping centre, and are asked to choose between several (unbeknownst to you) identical products. One expects, one imagines oneself, to select randomly, but in fact statistics show people tend to choose objects on the right; one’s simulation is inaccurate.

LOTTERY: A man conducts an office lottery, with some people simply being handed a ticket upon purchase and others being allowed to select their ticket from a set. Both tickets cost the same. On the morning of the lottery, days later, the man tries to buy back the tickets he has sold. Will there be a disparity between the amount of money either sort of ticket holder will ask for? One expects not, but in fact those offered one ticket from a selection insist on several times the average amount asked for by their colleagues.

APTITUDE: If one is led to form a belief, for example that one has a talent for telling if someone is lying, but then the relevant supporting evidence in question turns out to be false or misleading, would one continue to believe one had the talent in question? One expects not, but tests show otherwise.

One need not go along with all of Stich and Nichol’s expected reactions to such cases (I for instance was not surprised by the results of either the lottery or belief perseverance examples), but there is certainly something worth discussing here; it seems safe to say that at least one of these examples will seem surprising to everyone who is new to them. However, the simulation theorist is not helpless against their attack. Gordon argues that Stich and Nichols’ examples suffer from problems of methodology and what Gordon calls ‘imaginative impenetrability’.\textsuperscript{69} In methodological terms, it interferes with the application of these simulation tests to overtly build in important influencing factors; to specify that all the products are identical in the first case, for instance, or in the third case that the belief one is led to acquire is false. If these factors are supposed to be hidden from someone undergoing such tests \textit{in reality}, it is clearly inappropriate to introduce them to the simulator at the outset of the example. As far as imaginative impenetrability is concerned, Gordon’s point is that ‘even if visual imagery is a product of the visual-perceptive system running off-line, it is surely going to

\textsuperscript{68} \textit{Ibid.}, pp. 150-152.
\textsuperscript{69} Robert Gordon, “Reply to Stich and Nichols”, pp. 175-177.
bypass some stages of visual processing, especially early stages – and the [phenomenon concerned] may originate at one of these stages’. That is, simulation need not guarantee accuracy when elements of the relevant situation involve internal processes which are not employed in the imaginative rendering of such a situation. As Gordon says, ‘to imagine drinking six martinis doesn’t make one drunk, not even off-line… that’s a shortcoming of simulation, but it isn’t a shortcoming of the simulation theory’. This returns us to the qualification mentioned earlier in this section, that one’s simulated scenarios should prove accurate to the extent that they capture the relevant information. Of course, one might also note that one’s simulations might be inaccurate simply because one does not have the sensitivity or imagination to properly adjust to the differences in someone else’s character, or even to imagine themselves in alternative circumstances with sufficient vividness.

Thus far we have set out, roughly, how ST works, we have given some reasons to find it intelligible, perhaps even attractive, and have defended it against a criticism which in fact ended up helping us to constructively delineate the limits of simulation and ST. But there remains an aspect of ST which has great significance for the current project. Gordon talks briefly about the phenomenon of ‘unsolicited emotional accompaniments’, that is, he makes the plausible assumption that should one tell groups of children a story about a girl, Emily, losing her doll, looking for it and eventually finding it, if one then asked the children if Emily was happy to find her doll, a possibly small but consistently occurring number of children would respond not only by saying ‘yes’, but by saying so in a very glad tone of voice, with a happy facial expression and so on. Given that this seems like a natural reaction for a child to have in such a situation, one might well wonder why they would be so moved by the imagined scenario if they worked out the answer purely by means of a ‘cool calculation’ involving psychological laws. Gordon’s suggestion is that the children would themselves feel gladdened by Emily’s discovery of her doll because they are simulating her ordeal – and that when children are not so moved, it is probably because they are sufficiently sophisticated to suppress what emotions result from their simulation. Goldman, meanwhile, mentions a study in which people were in different ways instructed to observe a subject supposedly suffering from his hand, inside some kind of box, being exposed to terrible heat – some were told to pay close attention to the subject’s behaviour, some to imagine how the subject felt, and some to imagine what it would be like to be the subject, with their own hand in the box. Those in the third category apparently suffered the most discomfort, as observed both through verbal reports and physiological evidence. This all amounts to a strong suggestion that

---

70 Ibid., p. 176, my italics.
71 Ibid. Alvin Goldman gives similar objections to Stich and Nichols, emphasising the disparity between being immediately presented with the features and outcome of an experiment on the one hand and undergoing an experiment extended in time with an eventually revealed nature, with all the differences this would make to one’s psychology, on the other. Alvin Goldman, “In Defence of the Simulation Theory”, pp. 202-203.
73 Alvin Goldman, “In Defence of the Simulation Theory”, p. 204.
imaginative projection is apt to cause us to feel moved by the plight of others – in other words, it supports the claim made earlier in this chapter that the imaginative element of understanding has the power to make the concerns of others more alive to us.

There have been some further attacks on ST made more recently, however, in The Philosophy of Motion Pictures by Carroll, who is extremely sceptical about the usefulness of ST in understanding our engagement with fiction. His objections are several, and can be broken down as follows: 1. Simulation cannot be an exhaustive account of the way in which we engage with fiction. 2. We are not allowed enough time in movies to simulate characters. 3. People do not generally understand their own feelings, and so cannot be expected to simulate usefully. 4. Fictional characters behave in ways we would not, so simulation could not usefully predict their actions. 5. Movies are generally aimed at ‘untutored audiences’ who are ill-equipped to perform such a sophisticated means of understanding characters. 6. Simulation in movies is redundant because characters and plots tend to follow simple and familiar stereotypes which can be instantly understood, so positing simulation in such instances goes against the principle of parsimony (i.e. Ockham’s Razor). The answers to Carroll’s objections should be apparent from the preceding discussion – indeed, it seems as though Carroll’s criticisms are largely predicated on the simple failure to grasp ST.

1. No-one claims that ST provides a total explanation of our engagement with fiction or fictional characters. 2. One of ST’s original claims to fame was that it required less time and effort than TT, and in any case Carroll here drastically underestimates the power of the human mind. 3. It has been argued time and again that ST does not require that one understand one’s own feelings in a particularly in-depth manner. 4. ST does not, and never has, suggested that we only use our own characteristics in predicting others’ behaviour, even if some have suggested that we standardly do this until something goes wrong. 5. As with 3, the point of ST is that it does not require a vast body of technical knowledge, even if such knowledge can supplement one’s simulations – and it certainly is not taught to some privileged group of experts. 6. This is Carroll’s most interesting objection, although it is made more persuasively by Kieran. The first thing to say is that if simulation is (in one sense) a pervasive, semi-conscious affair, then our doing it naturally need not be undercut by a higher-level understanding of the story and characters – indeed it is not clear that introspection can prove ST is not involved in the understanding of stereotypes; for instance these could be default sets of simulative guidelines which one does not bother to properly confirm as being accurate. This connects with the second response, made by Goldman to Kieran: ‘The question of simulation versus inference

74 Noël Carroll, The Philosophy of Motion Pictures. Objection 1 is found on p. 172, 2 through 5 on p. 173, and 6 on pages 168 and 174. Matthew Kieran makes similar points in his earlier “In Search of a Narrative” – for Alvin Goldman’s refutation of Kieran’s article, see Simulating Minds, pp. 288-289.
is the nub of the ST-TT dispute, *a matter not readily settled in the armchair*. So [he] is not entitled to conclude that simulation by the reader isn’t needed, or isn’t used.\(^75\)

In *Simulating Minds*, Alvin Goldman examines a great deal of recent research on the recognition and transmission of emotions. This leads him to amend his account of ST in certain ways, not all of which are relevant to matters at hand. Most fundamental perhaps is the distinction he brings in between high- and low-level mindreading. He defines the former thus:

“High-level” mindreading [has]… one or more of the following features: (a) it targets states of a relatively complex nature, such as propositional attitudes; (b) some components of the mindreading process are subject to voluntary control; and (c) the process has some degree of accessibility to consciousness.\(^76\)

It should not come as a surprise that high-level mindreading is, of the two, closer to the phenomenon we have just set out as that originally described by ST, and by extension to identification. Low-level mindreading, by contrast, consists of reflexive and unconscious processes such as motor mimicry and the sort of apparently immediate ‘emotional contagion’ that is in play when we catch a smile or when a baby starts to cry at the sound of weeping. Essentially, Goldman sees two different prototypes as applying to either kind of mindreading respectively – mirroring in low-level cases, and what he calls enactment-imagination or E-imagination for high-level mindreading.\(^77\) This corresponds with the division others draw between more and less conscious processes of emotion recognition and/or transferral.\(^78\) As high-level mindreading essentially consists in those processes already discussed in much of this chapter, it is on low-level mindreading that we shall now focus. First we must acknowledge some of the research around which Goldman builds his theory, before then going on to ask whether Goldman’s treatment of the available facts, at least insofar as they bear on the issue of simulation as it concerns us, is satisfactory.

Earlier in this section, passing mention of mirror neurons was made, and it is now that we return to examine what is involved and what effect they have or are said to have on ST. Initial research into mirror neurons in macaque monkeys in the mid-nineties showed that neurons involved in the planning and execution of some action would fire not only when a subject would perform some action, but also that the same neurons would fire when a subject merely observed some reproducible action on the part of another subject. Other, roughly contemporaneous research shows that humans observing the actions of others in so doing stimulate the same muscle groups as those under

\(^{75}\) Alvin Goldman, *Simulating Minds*, p. 289, emphasis added.
\(^{76}\) Alvin Goldman, *Simulating Minds*, p. 147.
\(^{77}\) Ibid.
\(^{78}\) For others who make similar distinctions see for example Murray Smith, *Engaging Characters*, or Carl Plantinga, *Moving Viewers*. 

31
observation, but that in the observer these signals are inhibited. Other research Goldman mentions to bear this out was done in the mid-eighties, where ‘a group of patients with prefrontal lesions compulsively imitated gestures or even complex actions performed in front of them by an experimenter’. This was explained in the study as resulting from ‘an impairment of the inhibitory control that normally governs motor schemas, or plans’. Two further instances of research show on the one hand that one mirrors an entire set of actions bodily even when its concluding stage takes place outside of one’s sight, and on the other that ‘a monkey’s premotor cortex discharge[s] both when the animal performs a specific action and when it hears a sound related to that type of action’. Goldman also mentions research showing that our bodies react as though they had been physically touched in much the same way as we see someone else to be touched, and that many of the same brain regions are affected both when one is hurt and when one sees a loved one hurt.79

While this is doubtless very interesting, one might wonder if such findings really do anything to support ST as such – certainly, as Goldman himself admits, none of this lends decisive proof to simulational mindreading, where this implies the imputation of mental states to others through simulation. One might even worry, despite Goldman’s assurances to the contrary, that it is inappropriate to even refer to such mirroring as simulation at all. Let us answer this more fundamental worry first.

Simulation as a concept seems to entail some degree of intention. That is, no amount of similarity between two objects or events can guarantee that one is a simulation of the other – rather, one thing is a simulation of another if the former is intended to resemble the latter in important respects. Note that even here total or near total similarity is not an essential factor; a computer simulation of an earthquake, for example, will in most respects be totally dissimilar to an actual earthquake, and yet the former can reliably provide useful information about the latter by virtue of the fact that the thing simulated and the thing simulating share certain important characteristics, in this case common causal laws and information about a particular area’s population, geography and so on. What the important characteristics of any given simulation are, naturally, will vary according to the purpose for which the simulation is undertaken. But the kind of simulation presently under discussion, low-level simulation, is automatic, and not from force of habit; these are hardwired quirks of the human brain which seemingly at no point stem from conscious control. Surely, then, if such ‘simulation’ is described as having an intended aim or purpose, this can only be done in the spirit of metaphor. In the same way we might say that the purpose of photosynthesis is to produce nutrition for plants, without meaning that photosynthesis is something really done for the sake of nutrition. But such an objection would be rash. Goldman suggests, instead, that one should characterise a simulation as any process whose function it is ‘to duplicate or resemble the other… in some significant

respects’. The crucial difference here is that while ‘purpose’ implies agency, ‘function’ is sufficiently neutral that we can comfortably point to evolutionarily developed species traits as having functions, without thereby implying that such traits are deliberately instilled. Insofar as it seems appropriate to describe mirroring as a primitive biological means of coordination and/or communication, a means whose function it is to produce one state which in turn reproduces another, it is appropriate to describe mirroring as simulation.

But what about mindreading? The definition of simulation just given, it must be acknowledged, does not entail the active imputation of mental states to others, and Goldman himself explicitly declares that mirroring, as he construes it, does not entail mindreading. Indeed, as Goldman also points out, initial research into mirror neurons was done on non-human animals, specifically monkeys, to whom many people will not want to impute the capacity for mindreading.81 In the absence of scientific studies exploring this issue specifically, Goldman, hoping to show that mindreading can be based on mirroring, refers to several studies of only slightly tenuous relevance, but while interesting the results are not enough to convince one of anything particularly far-reaching.82 However, the claim Goldman wishes here to defend seems sufficiently weak and intuitively appealing to grant without overwhelming proof. If it is definitely established, as indeed it seems to be, that people automatically mimic each others’ actions and feelings to some extent, on some level, then if such mimicry is sometimes introspectively detectable (as with pain responses, for instance), and if the case for high-level simulation made in the preceding discussion was compelling, it seems more than likely, to say the least, that the ‘information’ we generate through mimicry should sometimes inform our efforts at high-level mindreading. Indeed it seems likely that the results of mimicry often demand to be taken into account, as when we cannot help but writhe in discomfort while watching a man we consciously suppose to be comfortably hammering nails up his nostrils.

As it stands, we can see that both high-level mindreading and low-level mimicry contribute significantly to the understanding of others, and that mimicry can appropriately be described as simulation. Whether or not low-level simulation itself involves mental state ascription is not so vital to our present concerns that we must wait for conclusive proof one way or the other. If there is a general lesson to be learned from studies into mimicry and motor neurons it is not that they provide evidence for ST properly speaking, but rather that the human mind is geared toward the internalisation of other people’s states. This is not ST, but it certainly makes the world more ST-friendly.

80 Ibid., p. 37.
81 Ibid., pp. 136-137.
82 Ibid., pp. 137-140.
1.4 Closing Remarks:

There are two reasons for the preceding discussion of ST. First, I take it as obvious that arguments for simulation are also, by extension, arguments for identification, and by the same token arguments for thick understanding. Second, I wish to support my prior suggestion that of the two broad means we have of understanding people, simulative and theoretical, the former is far more likely to produce in us a moving sense of y’s perspective, motivation and interests. This is not to say that we are guaranteed to be so moved — one might for instance find upon putting oneself in y’s position that one’s censure intensifies, because one supposes their actions to be particularly uncalled for given their perspective. Indeed this possibility forms the basis of Adam Smith’s picture of how we morally assess one another. And even if one were to feel the same as they, even if one came to empathise with the other, as Peter Goldie remarks this does not ensure that we will care about our so empathising, it does not even ensure our possible antipathy will be lessened — although at the same time I believe there is more of a connection between empathy and moral motivation than Goldie perhaps allows. My aim in this thesis is not so much to demonstrate that understanding or even empathy must be the genesis of our individual moral concerns, that they are psychologically inseparable, though I am tempted to believe that for our species generally empathy might have been an evolutionary progenitor of moral awareness. My aim, as suggested at the beginning of this chapter, is to explore what understanding requires of us and what ways there are to make us better at understanding, not because understanding necessarily makes morality but because understanding is vital to properly deploying moral concepts, at least as applied to actors if not necessarily actions abstractly considered. Now that we are armed with a sufficiently robust account of understanding and the imaginative and psychological processes it involves, we can proceed in the following chapters to see what understanding has to tell us about various ethical issues and to explore how usefully understanding can be brought to the philosophy of art. But first we shall in the next chapter set my concept of understanding in what we might loosely think of as a tradition of similar philosophical theories, all of which employ the idea of something like thick understanding, identification or simulation, to illustrate the history and growth of the concept, starting with the work of Adam Smith and then looking to more recent examples.

CHAPTER TWO

2. UNDERSTANDING AS A PHILOSOPHICAL TRADITION

2.1 Adam Smith and The Theory of Moral Sentiments:

Adam Smith is arguably, if perhaps not unquestionably, the greatest proponent of something like the understanding approach mapped out in this thesis. His Theory of Moral Sentiments (TMS), at the time its six editions were published over the latter half of the eighteenth century, was perhaps the most thorough and insightful philosophical study of moral and social psychology ever written, and adopted a humane and realistic approach to understanding morality as a human practice. Furthermore, his work can easily be seen as a precursor to simulation theory; indeed he is explicitly discussed by at least one of its first proponents.84 TMS might in fact be more comfortably situated in the history of psychology than philosophy; while it does contain some normative claims, which very much come out of and are interwoven with the psychological picture Smith constructs, such normative claims themselves are relatively sparse, and further are not themselves given any systematic or specific justification. TMS is perhaps most appropriately seen, then, as a descriptive text, one whose purpose it is to describe the actual workings of moral and social psychology – but it is nonetheless one with some ambitions toward ethical theory. Indeed it is this very approach of discussing morality as it is constrained by imagination and psychology that is essential to the understanding approach. If anyone could be said to have started the philosophical tradition of understanding, therefore, or at least a tradition recognisable as such, it would be Smith. This is true even if, as we shall see, his work was not without fault, and his position was not entirely consistent with the one here advanced. But let us first set out his psychological picture in brief, before going on to consider what problems we might find with it, or with the normative claims Smith makes in the course of TMS.

Though our brother is on the rack... our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations.85

TMS is about sympathy – and not in the modern sense of compassion or pity. For Smith, sympathy is manifested not by a specific emotion but by any emotion we feel for someone else on mentally putting ourselves in his position or, most especially, any emotion we feel in kind with someone on mentally putting ourselves in his position; as mentioned in the first chapter, any emotion felt as a result of this

84 See Robert Gordon, “Sympathy, Simulation, and the Impartial Spectator”. Note, though, that this article did occur relatively late in the development of simulation theory.
process was described as sympathetic. Smith’s idea of sympathy as such refers to this psychological mechanism, call it imaginative projection, which essentially has a threefold function: it enables us to understand other people, to form broadly moral assessments of others and ourselves, and to develop something approaching an independent moral standard. The first part of Smithian moral psychology relates to basic assessments of ‘propriety’. Here the idea is that x forms an impression of y’s internal life through imagining how x would feel in y’s position – and by comparing this simulated reaction to what overt behaviour suggests is y’s actual reaction, x forms an assessment of how reasonable y’s reaction is. As far as this goes, it sounds very much like simulation theory. But the story becomes more complicated, and indeed more unclear, when it comes to the matters of conscience and moral growth.

First, it must be said that Smith does add many provisos onto what I have called the basic picture here. Insofar as these tend to relate to the inherent imperfection of imaginative projection, and of people more generally, these are also very much in line with the understanding approach. For the sake of ease, we can summarise these qualifications as relating to roughly two interrelated factors: strength of imagining and partiality. From the start of TMS Smith establishes that, though it is a universal capacity, some are more inclined toward or affected by sympathy than others, and sympathetic emotions, even when of the same kind, tend to be different at least in degree. But further than this he makes repeated reference to the fact that humans have an egoistic streak, and a natural inclination to have far greater interest in the people close to or similar to themselves. He also notes that we can have sympathetic reactions that are impossible for those with whom we are sympathising, although strangely he does not seem to be interested in what complications this might imply for his theory of assessment. As it stands, realistic qualifications though these are, they do make the picture look grim if one hopes for universal peace and understanding (just as it helps to explain why we still do not enjoy such today). More philosophically speaking, though, this might give one cause to worry that Smith condemns us all to cultural relativism. And it is this concern which demands that we examine the second part of Smithian moral psychology, ‘the man within the breast’.

Smith had several terms for this ‘man within the breast’, calling it also a ‘demi-god’ and, perhaps most instructively, ‘the impartial spectator’. The idea, essentially, is that as we go through life, interacting with others, we develop a socially acquired set of standards, by means of observing what people’s reactions tend to be. This general process, he says, is what allows us to have any second-order feelings at all, providing an illustrative passage considering the mental life of a person who had never experienced social contact; ‘he could no more think of his own character... than the

86 Ibid., see also pp. 21-2.
87 See for example ibid., pp. 135-7, also his Lectures on Jurisprudence, p. 184; ‘those persons most excite our compassion... who most resemble ourselves, and the greater the distance the less we are effected by them’.
88 Ibid., p. 12. Smith takes this as a sign that ‘sympathy... does not arise so much from the view of a passion, as from that of the situation that excites it’ (ibid.), seemingly unconcerned by the fact his examples would seem to show our sympathetic reactions can in this sense be radically inaccurate and seemingly irrational.
beauty or deformity of his own face'. 89 As we develop, we acquire an internalised voice or perspective which stands to some imperfect extent outside of our prejudicial concerns and drives us to, for example, pursue the praise-worthy as opposed to praise – insofar as we might garner praise to which we are not for some reason entitled, while no-one external might realise this, we cannot escape the eyes of our simulated spectator who naturally has access to all the same information that we do. 90 At all times seeking to be realistic, though, Smith of course does not think that the voice of conscience controls everyone, and indeed he talks about how, for reasons of fashion and conformism, people sometimes emulate or venerate qualities which they know to be wrong, and of how strong emotions can feel so self-justifying as to temporarily skew our perspective entirely. 91 His overall position is well-captured when he says that our moral judgements ‘must always bear some secret reference, either to what are, or to what, upon a certain condition, would be, or to what, we imagine, ought to be the judgement of others’ – so while he believes that our moral perspective is socially derived, and indeed socially constrained, he does not think it is socially determined, he allows for a crucial gap between what we think is right and what our society thinks is right. 92

However some might still feel that this does not go far enough to guarantee freedom from cultural relativism. Fonna Forman-Barzilai, for instance, is extremely sceptical of the potential for cosmopolitan appropriation of TMS – that is, she feels Smith to be a ‘localist’ who is primarily concerned with the social dynamics of people who actually mix in a world without much travel or integration of social strata, and who fails to provide an adequate foothold for objectivism or universalism. She says this specifically in contradiction to the claims of those such as Amartya Sen, Charles Griswold and Jennifer Pitts, who each have a more liberal, inclusive and objectivist reading of Smith’s thought. One particularly telling point of disagreement in Forman-Barzilai’s book comes when she mentions Pitts’ and Sen’s separately suggesting that TMS encourages ‘openness toward unfamiliar values and practices’ and that the impartial spectator can ‘draw on the understanding of people who are far as well as those who are near’. 93 Forman-Barzilai, for her part, thinks of this as a noble but mistaken claim which reads more in Smith than is actually there – and she has good reason to, given that, as she herself mentions later on, Smith does make some ‘woefully insular’ remarks relating to travel and education; ‘he was suspicious of sending children to foreign schools, or supplementing university education with travel, since these practices tended to ‘hurt most essentially the domestic morals’’. 94

89 Ibid., p. 110. Smith does not himself use the term ‘second-order’, of course. For the purposes of this quoted passage one should probably assume a dearth of reflective surfaces in nature.
90 Ibid., pp. 130-1.
91 Ibid., p. 64, p. 201.
92 Ibid., p. 110, italics added.
93 Fonna Forman-Barzilai, Adam Smith and the Circles of Sympathy, pp. 178-180.
Regardless of whether Smith himself thought experience of other cultures improved one’s general moral perspective, though, it seems safe to say that he did not think morality as such was purely a culturally relative phenomenon. The most commonly cited passage of his in this regard is V.2.15, relating to the practice of infanticide in Ancient Greece, which he declares to be wrong in no uncertain terms.\textsuperscript{95} It is earlier in the very same chapter that Smith writes:

The different situations of different ages and countries are apt... to give different characters to the generality of those who live in them, and their sentiments concerning the particular degree of each quality, that is either blamable or praise-worthy, vary, according to that degree which is usual in their own country, and in their own times.\textsuperscript{96}

His explanation of Greek infanticide is to suggest that killing children is at least less surprising in a culture where anyone’s hardship is such that ‘it is frequently impossible for him to support both himself and his child’, where this practice remained even in later, better circumstances because ‘uninterrupted custom... so thoroughly authorised this practice’ that the Ancient Greeks lost ‘all sense of its dreadful enormity’.\textsuperscript{97} So from this we can see first that Smith does make unqualified moral assertions, second that he is comfortable with describing at least some entrenched cultural practices as wrong, and third that he has some deflationary story ready for why moral opinion should vary between times and cultures. Insofar as this relates to the more modern idea of ‘multiculturalism’, it is open to Smith to defend his arguably insular perspective by appeal to the fact that different nations face different circumstances, and have different histories (and thus different hang-overs); nations $x$ and $y$ could have differing moral perspectives even where both are epistemically or ethically on a par, yet greater experience of $x$ might provide one with no greater insight into the workings of $y$ since both perspectives are context-dependant. This could, after all, result in an atypical and thus antisocial set of idiosyncratic norms; expatriates can seem alien to everyone. Now, it is possible for the objectivist or universalist to insist that the two moral perspectives, if they are different yet equally valid, operate according to the same underlying principles\textsuperscript{98} – and it seems reasonably plausible to suggest that, should there be a universal morality which all cultures in some way respond to or pursue, seeing how (perceived) morality changes according to context could offer opportunities for enlightenment beyond those available within a relatively static system. And if \textit{this} goes through then I think it would be fair

\textsuperscript{96} \textit{Ibid.}, 204.
\textsuperscript{97} \textit{Ibid.}, p 210, p. 201.
\textsuperscript{98} Tim Scanlon calls this ‘parametric universalism’. See \textit{What We Owe Each Other}, p. 329.
to say that *TMS* does leave us, *by implication*, with something of a directive to experience other cultures, even if Forman-Barzilai is right in thinking Smith did not himself see it like this.\(^99\)

However, there does not seem to be any special reason why we should assume all cultures’ moral beliefs to be on a par. This might in fact, on the Smithian picture, give us a reason to *avoid* contact with other cultures, since Smith does believe that, though we can question public opinion, we have a deep-seated tendency to be effected by any set of standards with which we regularly deal; he writes that we have:

[A] natural disposition to accommodate and to assimilate, as much as we can, our own sentiments, principles, and feelings, to those which we see fixed and rooted in the persons whom we are obliged to live and converse a great deal with, [which] is the cause of the contagious effects of both good and bad company.\(^100\)

Insofar as we have at least prima facie reasons to think less of *some* cultures’ moral perspectives, because they engage in practices which we think of as unmistakably wrong (infanticide, for example), Smith might very probably have us steer clear of them for fear of actually becoming worse people. This sort of worry is compounded by the fact that, as Forman-Barzilai complains, ‘Smith asserted the distinction between praise and praise-worthiness without saying a word about how ordinary people... differentiate these standards’.\(^101\) Ultimately it might come down to how good people are at reasoning beyond their own social contexts, which for his part Smith certainly does not think is easy for most people. So while he seems to allow good reasons for liberal risk-takers to look through the window of their semi-independent perspectives in an effort to better see their strange neighbours, he also provides pessimistic conservatives with reasons to draw their curtains against the outside world.

This discussion, though, does not indicate a problem with *TMS* as such, nor does Forman-Barzilai think it does — it is important to remember, again, that *TMS* is primarily a descriptive work, attempting to empirically investigate morality as a social phenomenon, and the practical and epistemic problem of how to reconcile conflicting but deeply held cultural values is a very real, and very difficult one. Consequently *TMS* should, in a sense, leave us with this problem; at least it certainly is not obliged to solve it. This does, however, suggest why understanding must build on or move forward from Smithian sympathy, since it is an overall aim here to consider what genuine normative demands specifically are tied up in the phenomenon of interpersonal understanding. Indeed there are

\(^99\) It should be noted that Smith thought general moral rules were derived from experience of what conduct seemed right or wrong (see for example p. 159), and as such might have balked at this talk of moral principles underlying moral or cultural perspectives. However, I think it is fair to distinguish here between general rules for moral conduct, which Smith has in mind, and whatever underlying principles or story that might justify them on a theoretical level, which Smith does not consider.


at least two further reasons for us to reject, or rather to avoid fully embracing, Smith’s views. The first and most trivial reason is that, having been written in the eighteenth century, *TMS* naturally contains some empirical claims and observations which no longer seem accurate. D.D. Raphael for instance is scornful of Smith’s suggesting that each bodily organ or sense is the only means by which people can judge that same sense in others; ‘I judge your sight by my sight... I neither have, nor can have, any other way of judging about them’. So too we might do well to overlook Smith’s armchair anthropology, for instance his pronouncement that ‘among rude and barbarous nations’, as opposed to civilised ones, ‘the virtues of self-denial are more cultivated than those of humanity’. Smith might also be accused of a certain degree of complacent confidence in the wisdom of nature, regardless of whether or not one chooses to interpret his talk of unintentional order theistically. In his discussion of what we might today call moral luck, for instance, he glibly concludes that people’s common, irrational intuitions in this regard are to be valued as ultimately for the best – because, he assumes, we can’t know what the attempted criminal is really capable of, incomplete gratitude makes people try harder to help successfully and undeserved censure makes people more careful to behave well!

Worse still is his dismissal of the fact that

If [a man] was to lose his finger tomorrow, he would not sleep to-night; but, provided he never saw them, he will snore with the most profound security over the ruin of a hundred million of his brethren... Smith writes this in the context of imagining a European hearing about a terrible disaster in China, as a means to illustrate man’s grossly disproportionate degree of concern for himself and those close to him in various ways. His conclusion is to suggest that we are all better off being untroubled by the fate of those to whom we have no connection, and so cannot help, that this fact ‘seems wisely ordered by Nature’, going so far as to suggest that ‘if it were possible to alter in this respect the original constitution of our frame, we could yet gain nothing by the change’. One hopes that in the modern day, with greater democratic control of government policy and a profusion of international aid organisations, Smith would see fit to modify his ethically laissez-faire attitude here. Nonetheless, even if in context his argument seems inoffensive it is still surprisingly lacking in moral vision.

---

102 This is not to suggest there are no other grounds for complaint in *TMS* – his suggestion that failing to benefit does ‘no real positive evil’, for instance, is an arguable failing that we need not see as simply endemic to his time (ibid., p. 78) – but they are not sufficiently relevant to warrant discussion here.
104 Ibid., p. 205.
105 Ibid., pp. 92-105.
106 Ibid., p. 136.
107 Ibid., p. 140.
The second reason we should not be satisfied with Smith’s theory has more to do with its adequacy even within its historical context, or more specifically with the fact that some of the fundamental assumptions it makes are unwarrantedly bold. There are at least two good examples of this. One was highlighted by David Hume, who objected to Smith’s belief that ‘nothing pleases us more than to observe in other men a fellow-feeling with all the emotions of our own breast’.\(^\text{108}\) If this were truly the case, Hume suggested, a ‘hospital would be a more entertaining place than a ball’, and yet it seems plausible to suppose that people sometimes, if not frequently, are quite annoyed to have their good cheer sympathetically lowered by the arrival of unhappy company – that is, sentimental agreement broadly is not always enjoyable.\(^\text{109}\) Smith’s own reply to this problem was to suggest that on such occasions the feeling of the unpleasant passion outweighs and obscures the basic pleasure of coming to resonate emotionally with another person, but it might have been more prudent to simply weaken his claim; all he needs for his depiction of social interaction to work is the claim that sympathy usually produces pleasure, rather than the strong claim that it always does. There simply does not seem to be any need, when positing an innate human drive, to suggest that its satisfaction must always be pleasurable – indeed it is hard to think of any human appetite which cannot distort into a joyless compulsion.

The second, and perhaps more serious, case of Smith’s theoretical overreaching is his insisting that all interpersonal understanding involves sympathy or imaginative transposition. ‘The passions, upon some occasions, may seem to be transfused from one man to another, instantaneously, and antecedent to any knowledge of what excited them...’ Smith admits, yet maintains that such is the case simply because some emotions are sufficient to ‘suggest to us the general idea of some good or bad fortune that has befallen the person in whom we observe them’.\(^\text{110}\) As the discussion of the first chapter here indicated, however, not all interpersonal understanding can be characterised as imaginative; indeed Stephen Darwall notes specifically in relation to Smith that newborn babies for instance engage in mimicry on a psychologically-bare, reflexive level, and beyond this there do not seem to be any compelling reasons to believe that interpersonal understanding can never be theoretical in something like the way suggested by theory theory.\(^\text{111}\) While it is true that the approach to understanding presented in this thesis places more emphasis on, and gives more attention to, the imaginative, it is still a stated belief here, in line with the ST/TT debate, that simulation is not the be-all and end-all of moral or social psychology.\(^\text{112}\) Insofar as the imaginatively rich aspect of

\(^{108}\) Ibid., p. 13, see also pp. 15-16.

\(^{109}\) Ibid., p. 46.

\(^{110}\) Ibid., p. 11.

\(^{111}\) Stephen Darwall, ‘Empathy, Sympathy, Care’, pp. 264-6. We can describe babies’ crying in such cases as thin understanding to the extent that thin understanding overlaps with low-level simulation, as discussed in the first chapter.

\(^{112}\) Indeed understanding could in some instances compel us to a purely theoretical understanding of some creatures, in particular those with whom we have no common biological frame of reference.
understanding has affinities with the Smithian picture of sympathy, it is the suggestion here, following
Alvin Goldman and perhaps contra Smith, that this is not the kind or degree of simulation involved in
standard day-to-day social interpretation.

2.2 Nel Noddings and the Ethics of Care:
Understanding also has affinities with some alternative twentieth century ethical perspectives, such as
some feminist and so-called feminine ethical theories, which tend to emphasise the role of emotion
and interpersonal understanding, while downplaying the importance or usefulness of abstract rules.
One good example of this is Nel Noddings, who sets out a feminine ethics of care in her 1984 *Caring*,
which depicts and turns on what she calls the caring relation. She takes this relation of care to be
‘ethically basic’, not quite a good in and of itself but always good when exercised in regard to some
individual, and describes two kinds of caring, natural and ethical – the first is exemplified by the
unforced concern and affection she describes the mother as having for her child, and the second she
describes as harkening back to our memories of being so cared-for. Feminine moral action, then, is
motivated by ‘an evaluation of the caring relation as... superior to other forms of relatedness’, where
the desire for relatedness as such is taken as a basic human drive, and guided by ‘a picture of those
moments in which I was cared for and in which I cared’. She describes a binary dynamic of one-
caring and one cared-for, where the one-caring is engrossed in the other, adopting the advancement of
his interests as one of her own while being acknowledged by and so ‘completed in’ the one cared-for.
The way in which Noddings describes such ‘engrossment’, in particular, can bear a striking
resemblance to earlier talk here of thick, simulative understanding; ‘I must consider their natures...
and, although I can never accomplish it entirely, I try to apprehend the reality of the other’, which
‘involves stepping out of one’s own personal frame of reference into the other’s’. It would appear,
then, that while on the surface Noddings’ ethics of care is about a moral emotion – namely, care – in
fact it can legitimately be described as a member of the understanding tradition, since, for her, care
crucially involves the understanding and appreciation of other people, with a very similar emphasis on
imaginatively and emotionally rich engagement. This is especially so since this understanding
engrossment is required by *both* sides of those in her caring relation, either fully as one-caring or
minimally as one cared-for (as necessitated by the requirement of acknowledgement, which is more or
less to say appreciation).

113 Nel Noddings, *Caring*, p. 79.
114 Ibid., p. 83, p. 80.
115 Ibid., p. 14, p. 24. On the former page she even uses the phrase ‘caring from the inside’, which bears a
striking resemblance to Murray Smith’s later phrase ‘imagining from the inside’, meaning what we have called
identification. It must be said, though, that she assumes a rather more direct correlation between identification
and motivation that I do here.
116 In talking about mothers’ naturally feeling for or with their babies, she also suggests (as was stated in the first
chapter) that understanding need not *always* involve such imaginative transposition. Ibid., p. 31.
Other aspects of her position coincide closely with understanding as it is described in the first chapter, too. A central concern of her account, indeed, is the de-emphasis of abstract moral principles, and beyond this a claim that focussing on rules is not only unhelpful but actively harmful.\textsuperscript{117} She instead suggests the importance of detail and realism in ethical examples; ‘moral decisions are... made in real situations; they are qualitatively different from the solution of geometry problems’.\textsuperscript{118} So, to moralise properly, for Noddings, one must ‘concretise’ situations, making them specific, nuanced and real.\textsuperscript{119} This in turn suggests the usefulness of fiction as fodder for ethicists and moral agents generally, an idea we shall return to in subsequent chapters. Also important for Noddings, as for the understanding agent, is self-understanding, the attempt to clearly appreciate what sort of person one is, along with what one’s flaws and prejudices might be.\textsuperscript{120} Indeed she even emphasises the importance of understanding (in its generic sense) and experience of other cultures.\textsuperscript{121}

However, while much of what she says coincides with what I believe understanding and the need for understanding tells us, and while some of the limitations she discusses are of pressing relevance to the practice of understanding, she also says much which goes against it. Take for instance her claim that ethical action is ‘rooted in receptivity, relatedness and responsiveness’.\textsuperscript{122} At least on face value this makes the ethics of care look very much like an ethical theory \textit{built} on understanding. As I have already emphasised earlier in this thesis, one very often needs to appreciate who an agent is to an imaginatively considerable extent in order to respond to them properly, and this undoubtedly involves receptivity and responsiveness. However, the notion of emphasising relatedness is a potentially dangerous one, and not something that should be relied on, even though understanding can presumably involve seeing what is the same about oneself and another person. Noddings’ stipulation here might give her the advantage of an even stronger connection to moral motivation, it is true; quite possibly Adam Smith was right to suggest people are \textit{naturally} more motivated to help those with whom they feel some sort of kinship. But neither philosopher seems to properly realise or accept that this in itself need not be a happy fact. Early on in \textit{Caring}, Noddings says that ‘to be touched... I must see the other’s reality as a possibility for my own’.\textsuperscript{123} As a psychological claim this slightly nebulous assertion does not ring true, and as normative claim it seems positively odious. To the extent that it \textit{is} true, it might very well be something the understanding agent should try to \textit{overcome}, at least where it leads to marked bias and unfair treatment – which is to say, perceived relatedness need not be \textit{excluded} as a potential deciding factor when for instance there is no other differentiator between two

\begin{flushleft}
\footnotesize
\textsuperscript{117} See for example \textit{ibid.}, p. 2, 5, 24, 47, 56, and pp. 84-5.
\textsuperscript{118} \textit{Ibid.}, p. 3.
\textsuperscript{119} \textit{Ibid.}, p. 8.
\textsuperscript{120} \textit{Ibid.}, p. 108.
\textsuperscript{121} \textit{Ibid.}, pp. 184-5.
\textsuperscript{122} Nel Noddings, \textit{Caring}, p. 2.
\textsuperscript{123} \textit{Ibid.}, p. 14.
\end{flushleft}
groups each with a mutually exclusive claim to aid, but on its own it seems to pale into utter insignificance when compared with factors such as degree of suffering or numbers of people affected.

There are also some differences between the similarities, so to speak. For example, as already quoted, Noddings says that ‘although I can never accomplish it entirely, I try to apprehend the reality of the other’. 124 This is perfectly in line with the loosely fallibilist aspect of understanding, also found to some extent in Adam Smith’s work, which says we must remember our understanding of others, and even ourselves, cannot ever be complete. However, it is far from clear that this reflects a serious commitment of Noddings’ position, given that for the rest of Caring she talks in terms of ‘directly receiving’ people, and of the ‘transformation from individual to duality’, which suggests a naïve faith in people’s ability to understand and even substantially adopt other people’s perspectives. 125 Also, while her talk of concretisation is suggestive of the educative capacity of literature, she seems to believe that teaching requires intimate one-on-one contact; ‘I can lecture to hundreds... but this is not teaching’. 126 This is probably due in large part to her emphasis on reciprocation, which we shall discuss shortly.

If we look closer still we can see that Noddings’ care ethics sometimes makes suggestions that could well be inimical to the standards of understanding as they are construed in this thesis. Take for instance her claim that ‘the one-caring [should] interpret the words and acts of the cared-for in the best possible light’. 127 On the face of it, this looks like an impossible directive for the understanding agent, who is obliged to interpret people accurately. Understanding, as should be clear from the above, has nothing to say about what sorts of interpretations one should favour, in fact it might even suggest that favouring any particular sort of interpretation as a rule is counter-productive and irresponsible. However, Noddings adds that ‘it is vital that the one-caring not create a fantasy’, that he must ‘attribute the best possible motive consonant with reality’. 128 So perhaps this prescription is merely indifferent to our sense of understanding, insofar as while it does not actively clash, it recommends something which understanding does not, namely that one’s understanding must not only be as full and accurate as possible but must also be somehow morally improving for the one being observed – Noddings’ point is that we should help people to see how they can improve themselves by realising their ethical (broadly speaking) potential.

More puzzling from an understanding perspective is her stipulation that ‘the cared-for... has a special obligation to attribute caring motives to the one-supposed-caring’. 129 This particular requirement is owed primarily to Noddings’ concern that it is hard to go on caring for someone who

124 Ibid.
125 See for example Caring, p. 22, 30 and 61.
126 Ibid., p. 113.
127 Ibid., p. 123.
128 Ibid., p. 123, p. 193, italics added.
129 Ibid., p. 15.
does not appreciate one – this can change ‘care’ to ‘cares and burdens’. Again, this is not a concern to be found in understanding. Indeed, Noddings’ persistent emphasis on the importance of the one-caring being appreciated lends her theory a faint air of self-indulgent egoism. It is also illustrative of a pervasive difference between care and understanding, in that the caring relation is inherently asymmetrical. If two people interact, on the picture presented in this thesis, they both must be understanding toward one another, where this is a case of two independent and symmetrical obligations attaching to either party respectively, as opposed to a single unit comprised of two agents with asymmetrical obligations running between them. This difference can be watered down somewhat, but is ultimately inescapable; Noddings, for her part, does say that there can be pairs of people who respectively interchange between one-caring and one cared-for, indeed at one point she suggests they can both care for each other simultaneously. But still it is not clear this can be true of any particular action rather than describing a general state of affairs existing between two people.

As for understanding, it should be said that while all things being equal any two interacting agents will have an equal obligation to be understanding toward one another, it is probable that in some cases one party will have a greater obligation to be understanding, or an obligation to be more understanding. An example of this might be a student’s being under less of an obligation to understand her teacher than vice versa. Again, though, this does not help us to escape the fundamental difference that care is for Noddings inherently and essentially asymmetrical, while understanding is not.

It was alluded to above that Noddings’ theory is surprisingly preoccupied with the effect of the caring relation upon the one-caring. Noddings refers to the need for the one-caring to be acknowledged and so ‘completed’ in the other as reciprocation. When the cared-for does not reciprocate in this basic way, Noddings say that we can only ‘attempt to care’. Not only is this the most counterintuitive and inexplicable aspect of her theory, it is also the most problematic and the one most alien to the requirements of understanding. Think of Patrick Swayze in Ghost – he plays one half of a young couple very much in love, who is killed and subsequently follows his romantic partner as a ghost. Now, it is true that as the film goes on he develops the ability to communicate with her, but it is utterly counterintuitive, it is nothing short of bizarre, to suggest that he does not care for her until such time as he re-establishes communications. Or for a less flippant example imagine an inexpressive, socially maladjusted father (the kind Noddings often seems to envisage) who loves his child dearly but cannot for all his effort find a way to talk in the child’s language, his every attempt leading to misunderstanding and resentment. It seems somewhat cruel to suggest that this must

---

130 Ibid., p. 37.
131 Ibid., p. 69.
132 See especially Ibid., p. 69.
133 Ibid., p. 37.
134 Ibid., p. 14, p. 97. An example of this might be Willy Loman in Death of a Salesman.
mean the father somehow cares for his child less than a more naturally able parent, and it certainly does not seem understanding – after all, an imaginative construction of such a father’s perspective clearly includes a strong sentiment of care. In a similar vein, it seems perfectly natural to talk about caring for someone in a coma, or caring for someone too depressed to respond. One might say these are not ideal cases of caring, perhaps, or cases of caring in less than ideal conditions, but this is far from admitting it is a distortion of so great an extent as to actually disqualify them as cases of genuine care.

But perhaps this is unfair; after all, at the beginning of the first chapter here a slightly non-standard interpretation of understanding was argued for. Noddings herself lists some common sense definitions of care by way of contrast with her own, and so this problem should not surprise her. But recall that the modification made here to the commonsense idea of understanding involved an increase in the depth of understanding implied and the addition of an attitude of informal fallibilism, and while these are changes, they are very much changes consistent with the logic of the original concept; understand, just understand more and better, and understand that understanding is fallible. There is nothing internal to any normal or intuitively appealing definition of caring, by contrast, to suggest that the other person must repay one in some way, even if it is only by acknowledging one’s attention. I do not, what is more, see anything internal to the idea of understanding which would require that one be acknowledged by the other whom one is trying to understand, and this is the real objection at issue – the ethics of care depend on a heavy use of understanding, and so there is a lot of overlap between what a caring agent and an understanding agent should do, but care and understanding remain in this way very separate concepts. If someone does not, cannot or will not understand us, this does nothing in itself to lessen our obligation to understand him.

Not only does this reciprocation requirement introduce a fundamental difference between understanding and care, however, as already suggested it leads to serious problems for Noddings’ account, problems which understanding escapes. It is important for our purposes to call into question this idea that our moral obligations to a creature could be affected by its ability to reciprocate. After all, if it is reciprocity that makes the real difference between care and understanding then to defend the latter it should be explained why this difference is not a problem for us. In her attempt to deal with what ethical obligations we can have toward non-human animals, Noddings stipulates that ‘insofar as we can receive the pain of a creature and detect its relief as we remove pain, we are both addressed and received. There is at least this much reciprocity in our contact and, therefore, at least this much obligation.’ Her aim here is to leave us with some restricted obligations, no doubt because to do otherwise would strike most people as counterintuitive, while also holding that fully-fledged care cannot be exhibited toward animals, inasmuch as she holds they cannot properly acknowledge or reciprocate and so cannot enable the one-caring to be entirely ‘completed’. This claim in itself sits

135 Ibid., p. 150, italics added.
oddly with her suggestion that the gurgling of a happy baby is enough to complete one; surely a dog is much better equipped to appreciate us and reciprocate care than a newborn baby is. Beyond the fact that it appears to rest on something of an internal contradiction in her theory, though, it seems independently obvious that the extent to which it is possible for us to imaginatively assume the perspective of a creature is not a measure of its moral importance, as it arguably would be under Noddings’ approach. If we are in a situation where we suspect some creature to be of moral relevance, yet cannot usefully attempt to understand them, then so much the worse for us – this fact might somehow constrain or shape our practical obligations toward them, but it certainly does not affect the fact that they are morally important as such, even if they are incapable of care or any analogue thereof. And even should Noddings’ talk be limited to moral obligations rather than general moral worth, this still points to a formidable difference between care and understanding in terms of where and how our obligations fall.

Take her talk about cats. They are responsive cared-fors, she says, but their responsiveness is restricted. Ultimately, she claims that once one has cared for a cat, that becomes part of one’s ethical ideal, and the caring relation one has for that particular cat constitutes a formal link to other cats. In this way one becomes obliged to show care toward any cat one comes across. But, she says, ‘I was not obliged to enter the relation. There was no inevitable caring and being cared for with respect to cats’. She then goes on to suggest that one with such an obligation to cats does not thereby have obligations to any other species – using rats as an example. Looking at this imaginary rat she declares ‘I feel no relation to it. I would not torture it... but I would shoot it cleanly if the opportunity arose’.

From an understanding perspective, this seems glaringly unjustified. There are certain qualities which count toward a creature having moral worth, the most often appealed to of course being either the capacity to feel pain or the capacity for rational thought. As a concession to Noddings we might temporarily set these aside and imagine for the moment that what makes something morally relevant is its capacity to exhibit and respond to care. On any of these criteria, though, rats rate higher than cats – they are more sensitive, more intelligent and more social. Even if we only suppose that this is the case, it is still damning that it would have no effect on Noddings’ position, since by her it is a matter of personal choice which kinds of animal one accepts obligations to. So her feline allegiance seems arbitrary, and furthermore it fails to show understanding, because it does not properly take into account what we know about the perspectives of the relevant animals. And what about her stipulation that care between a person and a non-human animal is not inevitable? No doubt this is true, but the same is also true closer to home. Judging from history it would seem the natural reaction people have

---

136 Ibid., p. 120.
137 Ibid., p. 156. Again, the idea of restricted responsiveness or reciprocation suggests we cannot properly care for people in deep, extended comas or people who are severely depressed.
138 Ibid.
139 Ibid., p. 157.
toward those with a different language, religion or skin colour is one of the utmost animosity; at any rate it is obvious that care between people is by no means inevitable.

She later goes on to say that while plants cannot reciprocate, they are also ‘responsive cared-fors’, and that while they might very well flourish just as well under conditions of emotionally inert caretaking the one-caring, in caring-for her plants rather than simply looking after them, gets something more and special out of their flourishing that a more detached caretaker would not. Now, in and of itself this may very well be true, but it is most likely not something understanding would recommend, although it might have us excuse it in other people. This doting gardener calls to mind a kind of Aristotelian distinction between a fool’s happiness and eudaimonia, between ignorant content and genuine flourishing; should we encourage as fulfilment a form of happiness which so clearly flirts with delusion? I am assuming for the purposes of the present argument that though plants are sentient in some quite restricted sense they are not conscious. If I am wrong in presuming such, I at least have the excuse of being ill-equipped to determine that I am mistaken. But even if plants were conscious, of course, it would still be irrational to pursue a reciprocal relationship with one, regardless of whether one ‘got something more’ from the plant’s subsequent, effectively coincidental flourishing – since even minimally meaningful communication remains impossible. A man who lives off the land by tending to crops has as it were an instrumentally reciprocal relationship with his plants – they get something from each other – but this is not the same as a consciously reciprocal relationship, or even a conscious relationship which is reciprocal.

At the same time, I do not see that understanding would deny us the right to speak to our plants provided we were doing so in the spirit of make-believe – pretending to have a reciprocal relationship with one’s plants might very well be rational, or certainly not irrational, particularly if it enables us to gain more enjoyment from our plants, and more so again if in so doing we could not thereby end up acquiring any erroneous beliefs or inappropriate attitudes; perhaps a vague disposition to avoid undressing around tulips. It is far from clear, though, that this could justifiably or usefully be described as a relationship of care in Noddings’ terms. It could be described as an understanding relation, however, provided that it was informed by information about the plant that was, loosely speaking, as true as possible, and also provided that one could keep a handle on the difference between what was true of the plant and what one merely imagined to be true of it in play.

Generally speaking, then, the only way this reciprocity constraint could be seen as acceptable, I propose, is if it was rephrased to suggest that we have limited or no obligations to creatures that we cannot recognise as even being alive – which is to say, as long as we cannot reasonably be expected to think a given thing could possibly possess inherent moral significance we can be excused for not realising that we should ideally be affording it some kind of moral status. Imagine the discovery of a race of sapient crystals on another planet, for instance. It might be inconceivable from the perspective

140 Ibid., pp. 159-160.
of the first visiting astronaut that these apparently inert objects were conscious, particularly if the idea goes against his most certain beliefs. These crystals, let us suppose, communicate through incredibly minute vibrations that cannot be perceived except with very precise equipment. The hypothetical astronaut, on this understanding of the reciprocity rule, would not be morally at fault if he were to take a sample of this crystal, thereby unwittingly killing one of the sapient crystal beings, since it could not respond to him to signal that it warranted special consideration. But this is a problem with establishing that the creature is alive, however; once it had been established that the crystals were alive by some reliable means or authority, future agents who could have heard of this proof would have reason to treat the aliens differently, regardless of the fact that the human agents themselves could have nothing like a reciprocal relationship with them. The problem with such rephrasing, though, is of course that it remains misleading to refer to this as a rule of reciprocity or acknowledgement. The first astronaut would have reason to treat the crystals as morally significant if for instance some reliable source had left a note explaining that, though humans could not possibly communicate with these crystals, they were nonetheless very much aware and alive. So reciprocity really has no fundamental role to play whatsoever.\footnote{We shall return to considerations regarding understanding and alien perspectives in the following chapter.}

\subsection*{2.3 Richard Hare and \textit{Moral Thinking}}

Another prominent twentieth century philosopher who placed tremendous weight on a concept or process closely resembling the one here described as understanding was Richard Hare. To illustrate the extent and depth of this similarity it would be helpful to first provide a brief outline of the kind of picture he offers in his 1981 \textit{Moral Thinking}. Hare was a universal prescriptivist, which is to say he held that moral principles are prescriptions, or prescribed preferences, which must generalise across all relevantly similar cases.\footnote{The metaethical implication being a broadly non-cognitivist picture of morality whereby moral statements cannot appropriately be described as strictly true or false, while at the same time being subject to a certain standard of coherence and permitted some descriptive content – although Hare is keen to insist that one still cannot derive a moral principle from a naturalistic description of matters alone. See for instance his \textit{Moral Thinking}, pp. 216-8.} Since anyone holding a moral principle thereby effectively desires to impose it on all those possibly involved in situations of the same kind, Hare suggests, and indeed because people’s specific identities are not morally relevant so much as the generic capacities and roles they possess or occupy, the moral agent must employ principles which he would find most suitable were he to occupy all the positions involved.\footnote{\textit{Ibid.}, p. 42.} But already this bare description is getting ahead of itself – as we shall go on to see, the act of internalising other people’s positions or perspectives is left to the higher, more unattainable side of morality for Hare. So let us start again by describing the fundamental division he sees in moral thought, before going on to investigate in greater detail the ways in which the ‘levels’ he stipulates work and how they correspond to understanding.
Early on in *Moral Thinking*, Hare describes there being two levels of moral thought ‘which deal with matters of moral substance’, together with a third, the metaethical, ‘at which we are operating when we discuss the meanings of... moral words and the logic of moral reasoning’. It is with the first two levels that Hare is primarily concerned here. The first he calls the ‘intuitive’, and it is on this level he says that most moral decisions tend to actually be made, that is, most moral judgements are made according to one’s intuitions – something he takes great pains to endorse, for reasons that will become clear later, although at the same time he insists from the start that intuitions are worse than useless when thought of ‘as the basis of moral thinking’. The second level he calls the ‘critical’, and it is from this level that *true* moral principles result, although for various reasons this is not or cannot be the level at which we normally operate. What this critical level involves is very close to either simulation or understanding. Again we shall postpone closer examination for the moment, but essentially Hare describes ‘critical thinking’ as reasoning according to the knowledge of ‘what it is like’ for the people in some situation, where this involves reproducing their preferences internally in an emotively or phenomenologically rich sense.

On Hare’s proposed view these two levels are entirely separate and distinct, with different methods, different purposes, and different appropriate circumstances. Not only do both levels turn out to show some affinity with understanding, however, one might well suggest that from the perspective of understanding, or indeed that of the simulation theorist, the degree to which Hare separates them seems questionable, at least practically speaking. First let us look at the intuitive level more closely. Intuitional moral reasoning, Hare suggests, involves the use of ‘prima facie principles’ (PFPs), defeasible moral principles which in the natural course of things are not typically exposed to great scrutiny. The beginnings of these PFPs are instilled in children from an early age, such that they tend to be, rather than conscious rules, deeply engrained dispositions to have certain feelings in relation to certain things – although he is keen to point out, against the intuitionist, that PFPs can be submitted ‘to critical thought when... safe and appropriate’. So far this is all very much in the utilitarian tradition; Mill himself declared as vital ‘the general cultivation of nobleness of character’, such as an engrained aversion to telling lies, even on many occasions where they seem both beneficial

---

144 Ibid., pp. 25-6.
145 Ibid., p. 11.
146 Ibid., pp. 92-5.
147 Hare cites this term as a being derived from W.D. Ross’s own (quite different) talk of ‘prima facie duties’ (PFDs), that is, ‘the characteristic... which an act has, in virtue of being of a certain kind... of being an act which would be a duty proper [all things being equal]’. W.D. Ross, *The Right and the Good*, p. 19. See also Richard Hare, *Moral Thinking*, p. 38.
148 Ibid., p. 30, pp. 38-9. This claim is not *entirely* fair to Ross, though, who believed that PFDs did not actually possess true normative force except when the majority of them weigh in favour of some action – and because Ross thought both that any action could possess PFDs counting for and against it, depending on how it was described, and that we can never be sure of having discovered all the PFDs relevant to a situation, his picture in fact demands a tremendous amount of critical thinking (if not in Hare’s unusual sense). See W.D. Ross, *The Right and the Good*, pp. 19-20, p. 28, pp. 30-1, p. 33, p. 41.
and harmless. Indeed Mill also wrote that ‘indeterminate [moral] generalisations’, though imperfect and improvable, are essential in the face of practical constraints on utilitarian calculus. Hare arguably goes further in defending this sort of arrangement, though, in that he more specifically embraces the idea that the PFPs we live by will sometimes get things wrong, rather than merely admitting that they could be somehow improved – in other words, Mill is probably more optimistic about and motivated by the idea that our general principles can and should be brought in line with whatever the ultimate truth of morality might be.

Hare, for his part, uses several analogies to illustrate this point, his apparent favourite being the game of backgammon. As the die introduces an element of uncertainty, so too do we have to make our moral decisions with only partial knowledge of the facts and future. The backgammon player cannot devise a specific chain of moves which will win him the game, he must instead employ a step-by-step approach, acting according to general principles applying to his situation as it appears to be in a given turn. Even the best set of such principles cannot guarantee victory, even though overall they are the best means to pursue it. In the same way, the thought goes, our best PFPs will not always direct us toward properly right actions, and certainly ‘are not definitive of ‘the right act’; but if we wish to act rightly we would do well, all the same, to follow them’. Supplementary to this description of affairs, Hare offers two general reasons why the use of PFPs is a good thing. The first, as the backgammon example is also supposed to illustrate, is the practical reason already alluded to in the previous paragraph: it is far too difficult and time-consuming to work out in every case exactly which action will yield the greatest utility. The second reason is more psychological than epistemic; that people, particularly when judging in haste or under stress, have a tendency to bias their own opinions – here he uses the same sort of case as Mill, saying that, contrary to what we might convince ourselves of in a given instance, breaking with one’s engrained disposition against lying has a much greater indirect cost than whatever immediate gains one is thereby likely to make.

Now take Hare’s second, critical level of moral thought. As already stated above, he takes this to involve, as required by the universalising nature of moral prescriptions, the internalisation of the perspectives of those involved in a situation or decision – ‘under the constraints imposed by the logical properties of... moral concepts and by the non-moral facts’. For him one of the key differences between critical principles and PFPs is that, while PFPs are by their nature general, critical principles can be ‘of unlimited specificity’, where this does not include, as already mentioned,

---

151 Richard Hare, *Moral Thinking*, pp. 37-8. He later says, in the same vein, that good generals are the ones who win the most battles, not the ones with the best prima facie principles, although the best prima facie principles are those which, on the whole, win the most battles – *ibid.*, p. 52. Note though that this makes the following of PFPs seem curiously devoid of inherent moral worth.
people’s mere identities or other ‘individual constants’. The reason Hare emphasises this difference is to explain away cases where our apparent moral duties come into conflict, or appear as though they can be overridden by non-moral concerns. PFPs, by virtue of their general and thus rough nature, can easily come into conflict or otherwise fall short but, Hare maintains, properly speaking we cannot be obliged to do two mutually exclusive things (since ‘ought’ implies ‘can’), and so serious moral principles cannot come into irresolvable conflict. Critical principles’ capacity for specificity, then, is what allows them to avoid such a problem, and it is to critical thinking we must turn when we are unsure of whether we have chosen, or perhaps felt, the correct PFP, or when we need to decide which PFP to go with in a case of conflict. Hare further draws an analogy between imaginatively sharing other people’s perspectives and imaginatively understanding one’s own future preferences, also making the more explicit claim that by this means of internalisation ‘multilateral cases... reduce themselves... to intrapersonal ones’.

We have already seen that, if we are to conduct fully-fledged moral reasoning, Hare stipulates we must know ‘what it is like’ for the people involved in an emotively or phenomenologically rich way, and unsurprisingly he further stresses that this means one must imagine being them, and not just being in their position. This is all perfectly in line with understanding. However, plausible and reassuringly familiar as much of Hare’s story is for us, the standard he envisions for critical thinking arguably seems to put it out of our reach; he writes that, in considering whether or not something unpleasant should befall another man, ‘unless I have an equal aversion to myself suffering, forthwith, what he is suffering or going to suffer, I cannot really be knowing, or even believing, that being in his situation with his preferences will be like that.’ It is not immediately obvious what exactly he means by this – certainly he adds immediately afterward that he does not ‘wish to be taken as claiming that we can ever in fact have full knowledge of other people’s experiences’. And a few pages later he says he is not making the claim that such understanding demands ‘I have an aversion to his suffering as he is suffering’, pointing out that a talented torturer might well be spurred on by just such an understanding, and further suggesting that people are not ordinarily capable of such thoroughgoing empathy. So he must mean that in a given instance of observing Bill suffering x, or about to suffer x, critical thinking will lead one to wish not-x for oneself as much as Bill wishes not-x for himself. But while his qualifications seem to suggest simply that he doubts we can ever know exactly what Bill feels, that we can never be sure our desire for not-x matches his, understanding in this thesis goes

---

155 Ibid., p. 41.
156 Ibid., p. 27.
158 Ibid., p. 110, p. 124, italics added.
159 Again, as noted in the first chapter, we should understand this to mean something like ‘imagining to have (all) their properties’ rather than ‘imagining we are identical’.
160 Ibid., p. 95.
161 Ibid., p. 99.
further in suggesting, to an extent with Adam Smith, that our reactions to imagination can never achieve total fidelity with our reactions to real things (or rather, things we perceive as real). In this sense the analogy between understanding our future selves and understanding others need not be adversely affected, but it does mean that the imagined approach of $x$ could never yield the same degree of concern as when we (also) believe $x$ to in fact be imminent for us. Something like this serves as a plausible explanation for the psychological facts that by and large we seem to care less about distant misfortunes, either in the past or the future, and that people commonly procrastinate in the knowledge it will make future difficulties worse.

Regardless of one’s level of scepticism concerning the possible perfection of such an imaginative move, though, experience seems to suggest it is fair to assume that people can, for the most part or in large part, perform relatively successful comparisons of the intensity of one another’s preferences. The essential impossibility of his critical thinking still stands, though, at odds with his otherwise down-to-earth strategy of mapping out the psychology of our moral principles and feelings. Despite his talking throughout Moral Thinking in terms of how we should conduct critical reasoning, he does at one point directly face up to this problem, reiterating an earlier stipulation of his that critical thinking is in fact moral thinking as a perfect moral agent, a superhuman archangel with superhuman powers, would do it. ‘The most that human beings can ask for...’, he writes, ‘is some way of approximating... to the thought-processes of an archangel’.¹⁶² This marks a significant difference between Hare’s approach and the one taken here. While Hare is content to talk about things we technically cannot do, with the assumption held somewhere far in the background that in practice we shall only approximate such doings, the attitude of understanding here described seeks more to be shot through with epistemic humility, to instead keep in mind that the psychological processes involved even in sound moral thinking are inescapably deficient. Looking at the basic position Hare ultimately advocates, rather than the picture he seems to invite, though, this difference all but disappears! Should understanding then merely be assimilated into Harean critical thinking?

The answer is no, not least because understanding does not appear to entail any specific metaethical commitments. Hare’s actual position is also, at least from a practical perspective, confused, as is clear from the arguments of the first chapter here. Recall Hare’s saying that PFPs are vital as a time-saving measure – given that actual critical thinking, as it can be performed by human beings, amounts to nothing more than thorough or effortful simulation, Hare goes against what later became common practice, in that this time-saving benefit has for some time been attributed to the simulative model of understanding others, specifically as an advantage over the theoretical. If simulation is the means by which we can arrive at better moral judgements, and if it is the means by which we ensure the correct PFPs are in play, it no longer looks convincing to suggest that moral thought should typically rely on a worse and, it appears, less efficient system.

¹⁶² Ibid., p. 122.
Recall too Hare’s suggestion that PFPs prevent bias: how he could make this claim I am not sure, even despite the obvious problem that the PFPs with which we are raised are often pre-packaged with all manner of pernicious social prejudices about gender, race, social strata and so on. Even if we were raised with, as it turned out, the best PFPs possible, it is mysterious why Hare should assume that we could not regularly apply these PFPs prejudicially. It seems clear that people tend to bias their judgements, to put it another way, but it is in no sense clear that PFPs are not frequently party to this very tendency. Bias does not merely affect judgement but can colour our very perception of what is around us; a selfish and greedy man is no doubt more inclined to see transgressions of his rights than others are, he does not simply prefer to always get the largest slice of cake – indeed his greed might be such that he takes more than his share without ever even realising it. So in describing both levels of moral thought Hare is guilty of dressing mutton as lamb, so to speak; he does not seem entirely ignorant of the difficulty we face as moral agents, but at the same time persistently conducts his arguments as though there is nothing to really worry about. And again, there is the psychological fact that our intuitive moral judgements are almost certainly simulative to a large extent, something Hare does not account for. It is all well and good to say we have a useful, automatic aversion to stealing, such that if one was to consider stealing it would cause one to immediately feel sceptical or ashamed. But simulation might well often be employed in our basic understanding of a situation, such that it could undercut our PFPs and determine whether we even construe a given act as stealing, as opposed to taking reparations, say, or a harmless liberty – there is no good reason to think it is only something we do in protracted moments of calm.

Some of Hare’s criticisms can also be used against him, here. He is disparaging toward both the subjectivist and the intuitionist, yet holds the former in higher regard because she is, as he says, ‘more self-aware’. Both positions deliver similar verdicts, that is, but intuitionism, unlike subjectivism, mistakes the same fundamentally subjective inputs as representing objective properties of the things and acts observed. In a similar way, then, it can be suggested that understanding offers a better model than Harean moral thought because it is more self-aware. This can be said for at least two reasons. First, it is less at variance with empirical facts about our psychology, where both understanding and critical thinking, if they are to have any relevance to us, must if nothing else be cast in terms compatible with what we know to be achievable for human psychology. The Harean hierarchy of critical thinking justifying and superseding our deeply felt yet defeasible moral principles does not hold in practice, because our critical thinking operates interdependently with our moral intuitions, where neither are beyond suspicion. And second, while Hare’s picture has room for this sort of restriction, in that he admits we can only approximate true critical thinking, in practice an ethic

---

163 Ibid., p. 77.
164 In any case, it is not clear why a perfect moral thinker could not or would not rely on perfect simulation and perfect theoretical reasoning.
that begins with and maintains a thorough awareness of everyone’s epistemic limits is less likely to lead to moral and epistemic arrogance; thus understanding encourages a more self-aware attitude in its practitioners. What is more, Hare also repeatedly stresses that sound critical thinking urges us not to think about cases which rarely or never occur, so it looks as though on Hare’s own terms it is counterproductive to judge acts or agents in terms some way referring to a perfect standard, as he apparently would have at least the ethicist do. But while perfect moral reasoning is perhaps a logical possibility, it is a practical impossibility. So there is no point in trying to develop a set of principles, prima facie or otherwise, that are actually the specific ones endorsed by ‘perfect critical thinking’, and it is imprudent to talk as though one has or can.

So much for competing pictures of moral psychology, then, but are there any more philosophical points of differences between Hare’s position and the one taken in this thesis, or differences in the specific normative claims proffered by either? Let us start by asking why Hare feels the need to establish a hierarchy of intuitively appealing, practicality-guided principles justified by superior, though imperfect, imaginative means. Again there are essentially two reasons for this – one, because he wants a story to explain why moral duties do not conflict, even though they seem to, and two, because he needs a story to explain away counterexamples which suggest utilitarianism can recommend actions that we strongly feel are wrong. In terms of the first reason, I do not see that we are forced to believe moral duties cannot conflict in the way Hare suggests. Genuine moral principles cannot conflict, he thinks, because should one be outweighed this would mean they could not be considered overriding. But there does not seem to be any compelling reason to accept this as true. It is normal enough to say that moral considerations are overriding, to be sure, but saying of a class that it is overriding of things outwith that class need not have any implications for whether or how things within that class should be ranked. This can be cashed out in such a way as to make it look strange, of course – take ‘overriding’ to mean ‘of ultimate importance’, for instance, and it does appear odd to say $x$ is of ultimate importance but $y$ is more of ultimate importance. But why should we have to do this? The set of ‘the best British walks’ presumably contains the walks those native to the British Isles should always choose when possible, all things being equal. It would be ridiculous to suggest that each member of the set must be equally ‘best’, however. Even then, supposing ‘of ultimate importance’ is the best interpretation of moral considerations’ overriding nature, to suggest that one moral principle might be more ultimately important than another need not be any more logically offensive than mathematicians’ regular talk about degrees of infinity.

As mentioned earlier, though, Hare has another reason to suggest that moral principles or duties cannot conflict, in that this appears to violate the rule of ‘ought implies can’. His example is of

165 See for example ibid., pp. 47-8 and pp. 141-2. It might be noted here that Adam Smith, contrary to what I am suggesting, thought ‘the wise and virtuous man directs his principal attention to... the idea of exact propriety and perfection’. The Theory of Moral Sentiments, p. 247.

166 Richard Hare, Moral Thinking, p. 57-61, p. 178.
promising to his children that they would all go for a picnic one afternoon (A), only for a long-unseen friend to briefly appear from far away, asking to be shown around town that same afternoon (B). Here it looks to be true both that one should do A and do B. At first Hare suggests ‘ought to do A’ means, in part, ‘ought not to do B’, but realising that not everyone will be convinced by this dubious move he then insists that the rule of ‘ought implies can’ is still broken by going from the two isolated ought-statements to the supposedly impossible conjunction ‘x ought to do A and ought to do B’ – impossible since A and B are mutually exclusive.\(^\text{167}\) His claim is that this is merely a case of conflicting PFPs, his position further entailing that there is, in fact, some definite answer as to whether x in fact ought to do A or B, inasmuch as true critical thinking, so he says, does not give rise to such apparent conflicts through its limitless powers of specificity.\(^\text{168}\) If we were to reject Hare’s suggestion that moral duties cannot conflict, then, we would have to offer some alternative explanation here. The easiest solution, although Hare does not mention it, would seem to be to suggest that impossible moral duties, if we should somehow acquire them, apply to us in that they can leave us with residual duties.

Suppose, as is no doubt reasonable, that ‘ought implies can’. In that case, the thought would go, once I acquire duty A I cannot take on any conflicting duty B, nor can anyone knowingly impose any conflicting duty B on me. As Hare’s example has it, though, the old friend doesn’t know about any pre-existing duties, he simply dumps moral claim B on me. Or suppose that I simultaneously receive two text messages making separate but equal claims on my attention – both my partner and my mother need to be driven to two different doctors immediately, say, where they live in opposite ends of town (assuming no other drivers are available and so on). It looks odd and needless to suggest that perfect critical thinking could only yield one course of action in solution to the dilemma; while Hare seems to assume one choice will always result in a higher chance of preference-satisfaction, it does not seem as if we need to concoct any outlandish examples to dispute this. In any case, this is the dilemma: I should do A, but equally I should do B, where I cannot do both. Now suppose we want to know what understanding would suggest about such a case. We would have to ‘imaginatively invest’ in this scenario, that is, really imagine what it would be like were we ourselves to face it, giving some attention to what the operative factors might be in determining our response to it, and looking not just at how we would resolve the dilemma in immediate terms but also at the more general effects this problem would have on us were we to genuinely face it.

Abstractly considered, as something of a logical problem which simply calls for a rational solution, we might think for instance that a duty to do A and B, where they conflict, must instead somehow really be a duty to do A or B. This is evidently the path that Hare chose. But again, applying the understanding approach, it seems that while we could not be expected to actually perform both A

\(^{167}\) Ibid., p. 27.

\(^{168}\) Where, again, for psycho-practical reasons our PFPs must be relatively simple and generalised, and therefore imperfect and overrideable.

56
and B we would still, contrary to what might appear rational on a narrow understanding of the situation, have to bear some emotional and behavioural burden as a result of not performing either one, even when the fact of their conflicting was not of our own doing. Further than this, it seems as though subsequently feeling or behaving entirely as though the unfulfilled duty did not, through its impossibility, apply to us, warrants blame. This need not be seen to violate the rule of ‘ought implies can’, however. In immediate, practical terms, it appears I must do either A or B, but doing one still does not entirely release me from my responsibility for having to do the other – since I could have, after all, fulfilled it. The unfulfilled duty does not disappear, but insofar as it specifically can no longer be fulfilled it transforms into one or more residual duties to apologise, explain oneself, or offer compensation of some kind. In fulfilling these duties I would, in turn, be fulfilling my duties, generally considered, toward the one who laid the unfulfilled claim on me. The number and kind of residual duties with which I am left will of course relate to, among other things, the question of if or how I myself am responsible for ending up with conflicting duties, since it is a standing duty of moral agents not to acquire obligations they cannot fulfil. But be that as it may, we can see from this that it is perfectly possible for a consistent picture of moral duties and principles to allow for the sort of conflict Hare thinks impossible.

What then about Hare’s attempt to escape the many infamous counterexamples to utilitarianism involving scapegoats, involuntary organ donors and so on? Here he maintains his characteristic self-assurance – insisting that he shows how such counterexamples can be ‘demolished’ – but again there are problems with his position, at least as he describes it, that he does not seem to adequately take into account. All such examples, he suggests, arouse our PFPs, where in any remotely likely case true critical thinking (we can only hope) will endorse our PFPs’ verdict. Let us focus on the sort of case Hare mentions later on in Moral Thinking, of what we might presumptuously call repugnant pleasures, the most straightforward example of which is sadism. To begin with, Hare dismisses as impossible the idea that one person could enjoy another person’s suffering so much as to have their enjoyment endorsed by utilitarianism, and this does not seem entirely unreasonable, at least in cases of intense suffering. But he then mentions the apparently more likely, and so more serious, case of mass sadism, as with Roman gladiators, foxhunting, bullfighting and so on, where the suffering of a relatively small number could be arguably outweighed by the entertainment of a very large audience. These latter, controversial cases of possible endorsement are especially plausible insofar as most ethicists, not just utilitarians, tend to take the moral claims of less cognitively sophisticated creatures more lightly, and so it would appear that at least a utilitarian defence of bullfighting or foxhunting is all but unavoidable.

---

169 This scenario does not apply to duties which are in themselves unfulfillable – such duties cannot come into conflict with others because they cannot get off the ground as duties at all.

170 Ibid., pp. 141-2. Harmless practical jokes may or may not be an exception to this.
Hare’s own response is two-pronged. First, he suggests that in determining what should be done, utilitarianism must not only assess a given course of action but also its alternatives. So even if utilitarian calculus came down on the side of foxhunting, say, it would still recommend entertainment that was comparably rewarding but without the negative element of suffering. This is not a promising solution, insofar as it throws the door wide open to problems of demandingness - superior alternatives no doubt exist to virtually all of mankind’s pursuits. Indeed it is also an unsettlingly contingent solution; could it not turn out that nothing matches the thrill of mortal combat, even after suffering is allowed for? His second response is in fact given to the slightly different problem of apparently repugnant pleasures which harm no-one, such as ‘deviant sex’ or extensive use of ‘the so-called pleasure machine’. Here he stresses that his version of utilitarianism is not cast in terms of pleasure, but rather of preference-satisfaction – and further, that preferences must be rational, that is, they must ‘survive maximal criticism by facts and logic’. This highlights something of an inconsistency within Hare’s account; he is evidently unsure about how to treat problematic preferences. In places he suggests that we should weigh mistaken preferences when they cannot be altered, and in others he suggests that we should only weigh rational preferences. Insofar as the latter approach seems to help with certain counterexamples, it does invite one to wonder why Hare should really consider himself a utilitarian at all, since internally-speaking utilitarianism does not have a satisfying means of distinguishing between pleasures, least of all by taking rationality to be a value. Consequently he only manages to distinguish pleasures to the extent that he relies on broadly non-utilitarian, rationalist standards. So while understanding and Harean utilitarianism overlap to a great extent (psychological pictures, justificatory problems and metaethical presumptions aside), this is another respect in which understanding is more self-aware, since understanding’s valuing of rationality can at the least be seen as a natural extension of its valuing intelligibility generally.

2.4 Closing Remarks:
We have seen over the course of this chapter, and indeed the one before, how the idea and emphasis of understanding has for some time existed in the work of a wide range of philosophers, all the way

171 Ibid., p. 142.
172 This need not be a fatal objection, at the same time, since Hare could say decisions should merely be relatively good, not relatively best, as it were. This might seem to jar with utilitarianism’s assumed prizing of maximisation, but given the sorts of concessions Hare makes to epistemic practicalities it could still be a coherent move.
173 This worry could be offset somewhat by the Epicurean perspective of possible pleasure being much less than possible pain, opening the way for a much heavier weighting of disutility, but aside from this being something of a dubious empirical assumption it still would not clearly resolve the matter.
174 Ibid., p. 143. Quote taken from p. 214, where Hare quotes R.B. Brandt.
175 See for example p. 106, p. 143 and pp. 179-81.
176 Mill notoriously tries to provide a means of distinguishing between pleasures that is doubly controversial, on the one hand for his argument being loose and on the other for it being questionably utilitarian – but we shall return to this problem in the next chapter.
up to the present day; no doubt signs of it can be found elsewhere, too, from Schleiermacher to Schopenhauer, Collingwood and many more besides. All of those philosophers discussed in this chapter have this in common, the belief that proper moral reasoning demands the imaginative assumption of other people’s perspectives. First we discussed Adam Smith for his unprecedentedly thorough and fascinating picture of moral psychology. Then we looked at Noddings, whose feminine ethics of care de-emphasised abstract principles and emphasised the imaginative understanding of others as a means to bring out the best in them and oneself. Finally we saw, in the work of Hare, that, contra Noddings, identification or thick understanding need not be at odds with the idea of moral principles, but that it can in fact be crucial to their ideal employment. Having thus set out what I hold understanding to mean and its importance to be in the first chapter, as well as exploring its presence in contemporary philosophical debate, and having in the present chapter described what we might think of as its philosophical heritage, as well as some reasons to believe it has progressed over time, we turn in the next chapter to see what ethical work it can do when applied to various questions and issues. We have already seen how understanding works as an interpersonal human practice, but questions abound concerning, for example, how we are to apply understanding to creatures at varying removes from our own psychology. Ultimately, once we have examined these issues, we shall return better equipped to answer the question of how and to what extent understanding differs from utilitarianism. But before any of that I shall first offer some brief arguments to show how the prevailing moral theories of today can – and must – make room for understanding on their own terms.
CHAPTER THREE

3. UNDERSTANDING AND ETHICS

3.1 The Prevalence of Understanding:

By this stage it should be clear that to treat a person ethically in an effective manner one has to have a reasonable idea of ‘what it is like’ for that person, where this can include not just a rough sense of their subjective experience but also external facts which are relevant to it. This is because, roughly, we need to know what her beliefs and feelings are and how likely or difficult it would be for her to change her beliefs and feelings, since these loosely determine action and, more importantly, strictly determine accountability or praise- and blameworthiness. While such a claim does not amount to an independent ethical theory, I think to this extent it is at least safe to say that we should test our moral theories for whether or to what extent they have room to accommodate or explain the importance of understanding. Any minimally adequate moral system must be able to provide some story about how understanding plays a role in moral thought and action; a moral theory someone could obey perfectly without understanding would simply not be up to the task of genuinely ethical interaction. If we are to avoid courting controversy with the apparent brashness of such a challenge, though, some preliminary remarks should be made about understanding and the main normative theories of today, as a means to deflecting any immediate objections before we go on to consider what more robust, original or distinctive guidance understanding might be able to give us in relation to various moral problems.

First, take utilitarianism. To pursue happiness, or indeed preference-satisfaction, for oneself and others obviously requires an understanding of oneself and others. This remains true regardless of what specific branch of utilitarianism one subscribes to, whether one looks at immediate consequences with the act utilitarian or more wide-ranging ones with the rule utilitarian. While ‘post-classical’ variants need not be concerned with hedonic value in a straightforward sense, understanding is still required to discern what rules or preferences are possible for whom. Beyond this, I take it that any plausible form of utilitarianism must find a way to somehow recognise certain preferences or elements of well-being as being more important than others. How exactly to determine when such interests are in play, and how to weight them against countervailing, lesser interests, are clearly very complicated questions, but it cannot reasonably be denied that doing so effectively often requires quite an imaginatively rich appreciation of the perspectives of those involved. So, in other words, any utilitarian must be understanding if they hope to track and predict the utility function of any action or set of rules, on pain of making more and worse mistakes in what they believe to be conducive to and/or constituent of such utility. There is then the further and readily observable fact that a great many of the most central and universal sources of human happiness – such as friends and family – are themselves inextricably linked with or dependent on social interaction, which we have seen to involve
the use of understanding’s so-called mindreading powers (both, we can allow, theoretical and simulative). It is so apparent that a workable utilitarianism must involve understanding, in fact, that it is surprising to see something like it fail to receive extended treatment until relatively recently.\textsuperscript{177}

Next, take Kantianism. Kant supposes us to have at least an indirect duty to our own happiness and a direct duty to the happiness of others, and as already suggested understanding is required to appreciate and anticipate what will make people happy, including ourselves.\textsuperscript{178} So Kantianism would at least prescribe agents an indirect duty to be understanding. But it can be further argued that Kantianism entails a direct, imperfect duty to be understanding, insofar as one could not rationally will a world in which one is never oneself understood; on pain of self-contradiction then one would have to form reasonably thick understandings of the other agents with whom one interacts.\textsuperscript{179} It could also be suggested that understanding is required by the formula of humanity, to ‘act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’.\textsuperscript{180} To properly treat someone as an end, it is not unreasonable to suggest, necessitates at minimum a moderately thick understanding of their perspective. Indeed it is hard to think of someone as having or exhibiting respect when they are not eager or at least willing to understand – just as a teacher might read a slight in the vacant stare of an uninterested student. This also underlines a difference between how utilitarianism and Kantianism would be likely to treat understanding; for the utilitarian, understanding is practically indispensable, though for classical, hedonistic utilitarianism at least it is of exclusively instrumental or derivative value.\textsuperscript{181} For the Kantian, however, to seek to understand someone and to be moved by their perspective could in itself constitute perhaps the most fundamental expression of a respectful or good will, and could therefore count in a more robust sense as something inherently good. The possibility of this greater Kantian role for understanding also parallels what many feel is an intuitive advantage of Kantianism over utilitarianism, namely that utilitarianism apparently has an unusually alienating tendency to view agents merely as receptacles, essentially insignificant parts of a total state of affairs considered impersonally.

Contractualism, in turn, could be described as a more understanding contractarianism – and in light of the importance given to understanding in this thesis it would be both helpful and suggestive if this comparison could be linked to the superiority of the former over the latter, as I believe it can. The

\textsuperscript{177} See previous chapter, section 2.3.

\textsuperscript{178} See for example Immanuel Kant, \textit{Groundwork to the Metaphysic of Morals}, pp. 64, 92. Kant further believes that principles guaranteeing happiness are not knowable for us; \textit{ibid.}, p. 82.

\textsuperscript{179} This might even lead to a perfect duty – that is, like promise-breaking, a duty we in some sense cannot coherently imagine being permissibly breakable – since a world in which no-one need be understood is a world we do not need to understand; that is, one that defies the rational objective with which we undertook to imagine it.

\textsuperscript{180} \textit{Ibid.}, p. 91.

\textsuperscript{181} I specify classical utilitarianism here since Hare and Singer both rely on something like the process of understanding as a fundamental part of their stories concerning moral thought. See for example Richard Hare, \textit{Moral Thinking}, pp. 42, pp. 110-1 and Peter Singer, \textit{Practical Ethics}, pp. 12-14.
most central difference between contractualism and contractarianism is the much thicker concept of reasonableness involved in the former as opposed to the much narrower idea of rationality employed by the latter, and this is a difference which translates closely into a difference in degrees of understanding. That is, reasonableness is a far more emotionally rich and psychologically realistic concept – an altogether more humane concept – than the sparse and minimalistically functional concept of mere rationality. While both suggest responsiveness to reasons, ‘the rational’ is often tainted by unassumingly psychotic undertones of egoism, particularly as social contract theories tend to cash out morality as an agreement for the sake of mutual self-interest; ‘the reasonable’ is far more suggestive of an inclusive picture of rational thought that sees other-directed reasons as naturally motivating, or as appropriate sources of motivation. Contractarianism is more error-prone from the start, then, because the sort of rationally self-interested agents it supposes do not of themselves pay much regard to the lot of those with whom they are bargaining – and this goes hand in hand with a diminished role for understanding. But morality is about more than mere coordination, it is more than a Hobbesian strategy to work out how we can reconcile ourselves to the regrettable torment of coexistence, at least between beings who are capable of more than mere self-interested coordination, as surely human beings are. So not only do contractarian agents have less room for the understanding of others, though they still need understanding to discern what others could ‘rationally’ agree to, such a position also recommends to us a poorer understanding of ourselves.

As one attempts to fill out the sparse and somewhat anachronistic idea of rationality to something thicker and more convincing, one comes closer to something like reasonableness. Take Scanlon’s story about the man who owns and controls the only water source in a desert region – the man can set virtually any price he likes since, as Scanlon says, ‘he is in a position to make it rational for his neighbours to accept whatever principle he chooses’. A reasonable agent, an understanding agent, seriously takes on board the perspectives of others and in doing so is thereby driven to find an acceptable mean between their own desires and the desires of those with whom they deal. This is part of what it means to be reasonable, as I see it – being reasonable suggests that one is not limited to one’s own perspective, it suggests that one is sensitive not only to reasons that apply to oneself, so to speak, but also to reasons as they apply to other people. A purely rational, self-interested agent has to grasp awkwardly for a reason not to break contracts when it would be to his benefit and could not have any further negative ramifications, such as spoiling his reputation, making him disposed to break

---

182 Though they do not explicitly link such reasons to the concept of ‘the reasonable’ itself, cogent examinations of moral reasoning according to something like such a model are offered by, for example, Philippa Foot’s “Does Moral Subjectivism Rest on a Mistake” and Derek Parfit’s On What Matters.
183 T.M. Scanlon, What We Owe Each Other, p. 193.
184 The link between Scanlonian contractualism and understanding is particularly encouraged, I think, by claims such as that ‘some of the most common forms of moral bias involve failing to think of various points of view which we have not occupied, underestimating the reasons associated with them, and overestimating the costs to us of accepting principles that recognize the force of those reasons’, ibid., p. 206.
contracts again or somehow inciting others to do so. This is not the case for a reasonable agent, and it is not the case for a properly understanding agent either, as they take seriously and are moved by the desires, needs and entitlements of other people. So we can see that contractualism’s greater role for understanding is closely connected to its superiority over contractarianism.

But perhaps most of all, talk about understanding is very comfortable for virtue theorists. What I mean to say by this is that virtue ethics seems to naturally give understanding greater prominence as an ethical activity itself, as opposed to something instrumentally, and so only derivatively, crucial to moral action. While Kant and Mill both allowed for the importance of proper moral habits and attitudes, to be sure, they did this more as way of dealing with peripheral fallout from their central theoretical commitments. But for a virtue theorist understanding is demonstrably of central importance to human flourishing, since its end is the proper interpretation and treatment of others and oneself; Aristotelian man is of course a social animal, and so understanding is itself a pervasive concern. It is something we are able to do, yet not something that is always natural or easy for us to do and requires practice; indeed it is probably most difficult when it is most important. Understanding, moreover, is important in the proper exercise of many important virtues — and not just virtues relating to purely social concerns, but also important virtues we would not typically think of as social or other-regarding, such as prudence. As noted in the first and second chapters here, we need understanding to imagine what our desires might be in the future, and so how to plan our lives (inasmuch as this is possible). Decisions about how much money to save, where to live, what sort of job to pursue and so on, are only better informed when they involve the proper exercise of understanding. At the same time understanding has its own set of subsidiary activities and its own distinct purpose, to the extent that it deserves to be considered very much a virtue of its own.

Having said this, a virtue theorist might worry that it is slightly less than clear how understanding can properly be expressed as having vices of excess and deficiency. Certainly it is easy to imagine someone who suffers from a deficiency in this regard. He who is not sufficiently understanding could be judgemental for instance, one who judges too quickly on too little evidence, or perhaps he even persistently ignores good evidence contrary to what his unreliable ‘gut’ tells him. But could there be an excess of understanding? My claim in the first chapter that understanding is always appropriate might invite the idea that understanding is invariably good, and that it makes no sense to think of an upper limit for appropriate understanding. This could be so particularly since understanding is self-policing; someone who is understanding will know when to stop trying to understand, say for instance because their colleague is a very private person who dislikes scrutiny, so surely if someone was extremely understanding they could only be better at understanding when

185 Although naturally it would be excessive to suggest that prudence must be exclusively self-regarding.
further or greater understanding is uncalled for.\(^{186}\) An excess of understanding, then, could only work, could only be bad, if really it was an excess of something like understanding but not really understanding – it would have to be something like prying, for instance, which is a much narrower concept than ‘overly understanding’. This need not be considered unusual, though, as something analogous could be said about many virtues; for example cowardice looks like the opposite of bravery, not just the lack of it.\(^{187}\) Excess of understanding, then, could involve a violation of the respectful and motivational tendencies I have suggested to be involved in the practice, whereby one’s thirst for knowledge about someone is not (sufficiently) dampened by internalising their justifiable or otherwise not inappropriate desire for privacy. It is also possible to imagine an excess of understanding in at least two alternative ways. Firstly, one could be too understanding in becoming condescending or arrogant, bored with those one can so easily understand; familiarity breeds contempt, as the saying goes. In this manner ‘fully’ understanding someone could lead one to lose interest in a person or take them less seriously, to avoid them or otherwise behave in a manner expressive or symptomatic of one’s lack of regard. Or secondly, one could simply be overwhelmed by the fallibilist dimension of understanding and become consumed by solipsism – in the face of genuine and fundamentally insurmountable epistemic doubts one could become paralysed by insecurity.

It should be obvious from all of the above that the importance of understanding can be comfortably expressed in terms native to most if not all contemporary moral theories. We can take from this that the proponent of understanding does not face any immediate or pressing objections from mainstream ethical theory, and further that emphasising understanding does not entail any controversial or counterintuitive claims which might lessen its attractiveness as a supplementary condition to any given ethicist’s approach. Indeed its undeniable appeal might in fact be its greatest weakness; understanding’s apparently universal applicability or relevance might give one pause to wonder if it must be somehow vacuous in order to be so unobjectionable. But nonetheless the examination of understanding does seem to yield important insights and useful guidance, both on a psychological and theoretical level. We have seen from what has transpired in the chapters above that understanding is a pervasive phenomenon, appearing in a wide range of debates and affecting not just our assumptions about how to conduct ourselves ethically but also our ideas about what kinds of things are important, about what subjects should recommend themselves to us, so to speak; a man can be an island, but he cannot be so in good conscience. We are obliged to take at least some interest in one another. In theoretical terms, we have seen that understanding provides a useful tool for illuminating and critiquing positions both which acknowledge or overlook it, not merely as a test for

\(^{186}\) Understanding that one should understand no further, so to speak, is not the same thing as ceasing to understand. It should be noted though that someone’s wishing not to be understood 1) still requires some minimal understanding in order to be appreciated and 2) could be outweighed by external factors, for instance by evidence suggesting they are hiding some nefarious plan.

\(^{187}\) Let us grant this minor point for present purposes, despite the fact that properly speaking rashness was, for Aristotle, the opposite of cowardice.
minimal adequacy but also as a means to assess one ethical theory’s performance over another. The task for the remainder of this chapter is to continue filling-out the theoretical yield from understanding. In this way we can establish a more provocative as well as a more productive picture, while also answering the need to provide illuminating distinctions between the recommendations of understanding as I see them and those of the most prominent contemporary moral theory we have seen to explicitly invoke a procedure extremely close to it: preference utilitarianism.

3.2 Understanding and Consciousness:

The supreme criterion for moral relevancy is notoriously a matter of differing opinion, the two traditional poles in analytic philosophy being those of, on the one hand, a hedonistic emphasis on pleasure and pain, and on the other a more Kantian valuing of rational thought. Both standards are subject to objections and counterexamples. Utilitarianism is famously seen as ‘ignoring the separateness of persons’, essentially by disregarding the importance of the distribution of happiness in favour of its net amount, potentially allowing for small numbers of abhorrent acts to be justified by sufficient gains and also opening the door to Derek Parfit’s ‘repugnant conclusion’, that a world of content yet barely aware zombies might be morally better than a much smaller world of fully aware human beings. Furthermore, while pain and pleasure often seem by themselves to be bad and good respectively, there are arguably conditions under which neither one should be considered as either in even a pro tanto fashion. Many people would believe for example that feeling a quiet inner glee at the suffering of another, even if this had no wider consequences or implications, could be wrong or could make a situation morally worse and in no respect morally better or ‘less bad’. Conversely, given appropriate conditions, pain might either count as a good thing or at least as something which is not in any sense morally bad or unfortunate. A joyless or even frightening ritual involving being tattooed, for instance, might be morally unobjectionable, at least provided it was entirely consensual – the presence of pain on such an occasion might very well be the or a specific reason why such a tradition would be undertaken and maintained, with no clear route to suggesting this must be, if morally permissible, justified by the wide-ranging or far-reaching effects on a group’s overall happiness.

The emphasis on reason, meanwhile, fails sometimes quite spectacularly to account for the duties many of us feel ourselves to have, in particular those toward the mentally disabled and non-human animals. Kant, for instance, notoriously fails to go beyond suggesting that kindness to animals is an indirect duty by virtue of the fact that animal cruelty might encourage us to be cruel to beings of actual moral relevance. As Jens Timmermann notes, this fails to convince not only because animals’ pain seems intuitively to be of direct moral relevance, but because Kant’s assumed analogous relation

---

188 See Derek Parfit’s *Reasons and Persons*, pp. 381-90.
189 This is so particularly given the sort of move Hare was mentioned as making in the previous chapter, where utilitarian calculus is described as taking into account alternative means of securing happiness.
between animals and humans (which supposedly encourages our treatment of one to affect our treatment of the other) does not hold between God and animals; yet it seems as though it would be wrong for God to torture them. Indeed, Kant’s argument not only rests on a dubious empirical premise but also on a misleadingly contingent one; were it to be the case that torturing animals actually helped us to treat people better, by relieving stress say, it does not seem at all obvious that we would thereby be free to harm them as we pleased. Others of a loosely rationalist bent have suggested that animals and the mentally disabled should be treated as though their interests were represented by a rational trustee, but it is not clear that this can be coherently defended. And some contemporary Kantians, in attempting to improve the Kantian position on such matters, have made suggestions that would seem to leave us with moral duties towards clearly inappropriate objects which nonetheless display ‘virtually’ rational powers, such as office computers.

There is a way to avoid these various problems, however. Reflecting on the process of understanding, looking at what we crucially (re)produce in understanding another morally relevant subject (that is, at what characteristics seem necessary to warrant moral claims), we can see that claims referring to hedonic values and claims referring to specifically rationalistic standards both depend on an underlying substrate: the conscious perspective to which such claims could be said to apply. If that is true, I suggest understanding directs us to see the conscious perspective as the ethical standard, as that which is of most fundamental moral value – and so to understand right and wrong ultimately in terms of what is conducive and/or contrary to the flourishing of a conscious perspective qua conscious perspective. This is a reasonable move for at least three important reasons. First, as just suggested, the conscious perspective is a necessary vehicle for both pain and rational thought in any meaningful sense, and so is something vital in either case. Second, this moves away from an overly sharp distinction between thought and feelings or emotions, a demarcation we might see as something of an oversimplifying anachronism of the Enlightenment. And third, this readily provides us with a natural means of distinguishing between animals as having varying degrees of moral relevance; roughly, by valuing their greater cognitive capacities according to some set of consciousness-relevant criteria. We can imagine a tentative such list as including elements such as intricacy, coherence, reflexivity and veracity, where these are understood to operate interdependently – so that the presence of any one in extreme amounts need not in itself count for much. Thus we could exclude from the start the possibility of counter-intuitively valuing creatures with only, as it were, pointlessly complicated brains, creatures with extremely meagre but highly organised thought processes, or

191 Scanlon, for instance, suggests the idea but is quite sceptical about its usefulness. See his What We Owe Each Other, pp. 183-4.
192 See Allen Wood’s “Kant on Duties Regarding Nonrational Creatures”. Here he questions what he calls Kant’s ‘personification principle’, saying we should not value rationality merely as it is found in persons, but rather in things which have ‘rational nature only potentially, or virtually, or having had it in the past, or having parts of it or necessary conditions for it’ (ibid., p. 197).
creatures subject to complex but meaningless hallucinations. It might also be the case that for instance a significant degree of reflexivity, which we might be tempted to think of as the most important criterion mentioned, could not make sense without a certain degree of intricacy and consistency to support it. These interrelated factors could in turn be seen as contributing to or symptomatic of various broad capacities for understanding the world, such as sensory perception or rational analysis.  

Something quite like this criterion has been suggested by Donald VanDeVeer in his 1979 paper “Interspecific Justice”, though VanDeVeer’s account suffers from some grave limitations. He proposes something he calls Two Factor Egalitarianism (TFE) as a response to what he takes to be deeply unpersuasive implications of Peter Singer’s utilitarianism. The idea is that in cases of potential conflict between members of different species TFE requires one to look at two factors: 1) the importance of the interests concerned, where these are divided between basic, serious and peripheral interests, and 2) ‘the psychological capacities’ of the creatures in question. ‘Basic interests’ are those which are essential to a creature’s survival, ‘serious interests’ are those needed for a creature to flourish, and ‘peripheral interests’ are whatever ‘frivolous’ or relatively trivial further preferences a creature might have. Under TFE, VanDeVeer says, basic interests should never be sacrificed for any peripheral interests, but by virtue of having greater psychological capacities one creature’s serious interests can outweigh another’s basic interests, and in the same way interests of equal seriousness should be conceded to the being with greater ‘psychological complexity’ (where, puzzlingly, he uses the terms ‘psychological capacities’ and ‘psychological complexity’ interchangeably). The advantage of TFE over Singer’s picture, VanDeVeer suggests, is that it preserves intuitively appealing outcomes that Singer does not, by for instance determining that should a man and a dog be trapped on an overloaded lifeboat (i.e. where their two equal interests in continuing to live clash) it should definitely be the dog who is sacrificed for the greater good. VanDeVeer does not rest his argument purely on an appeal to intuition, however, in that at several points he suggests he is concerned with ‘psychological complexity in so far as that bears on the capacity... to live a satisfying life’. He argues, then, that the interests of more psychologically complex creatures have higher value because in failing to secure them we thereby fail to secure a greater range of ‘opportunities for a satisfying life’ than could have been enjoyed by less psychologically complex creatures – he refers to this as an act’s ‘opportunity cost’.  

---

193 I also take it that understanding can support such a move for purely pragmatic reasons; it seems as though we would at least be improving our understanding of ourselves if it truly is the case that ‘consciousness’ can more efficiently and consistently capture our basic moral intuitions.

194 Donald VanDeVeer, “Interspecific Justice”, pp. 63-4, p. 68. At one stage he also emphasises the greater relative importance of creatures with ‘more complex psychological capabilities’, ibid., p. 70.

195 Ibid., p. 73. See also ibid., p. 78.

196 Ibid., pp. 71-2. It is not clear if VanDeVeer would prefer to cash out ‘satisfying life’ in terms of hedonic value or preference-satisfaction. In terms of his overall argument, it appears more likely that VanDeVeer is
Unfortunately for VanDeVeer, as a modification of Singer’s utilitarianism this fails to make advances of any kind.\textsuperscript{197} That is, Singer’s position does not make the counterintuitive recommendations VanDeVeer accuses of it, and VanDeVeer’s alternative strategy at best amounts to more or less the same way that Singer actually answers objections of the kind VanDeVeer imagines to apply. While Singer suggests that ‘the Principle of Equal Consideration of Interests’ means no one interest can ever outweigh an interest of the same kind by itself, he does believe that one set of interests can be outweighed by any sufficiently greater set of interests.\textsuperscript{198} VanDeVeer’s lifeboat example is intended to illustrate a case where there is a conflict of equal interests – the basic interest in continuing to live shared by the dog and the man who are both in danger of drowning – where he supposes Singer’s picture to offer us no sure guidance. But VanDeVeer’s suggestion, that human basic interests tend to outweigh the basic interests of dogs, is primarily supported by the idea that, essentially, \textit{humans tend to have more (possible) interests}. Consequently both he and Singer argue the same thing. It is the same claim Singer makes in arguing that person-qualifying animals are worth more than ‘merely conscious’ animals; as a creature with no sense of itself as a distinct entity existing over time, with a past and a future, cannot have preferences regarding its future, the killing of a creature \textit{with} such an ability will thereby, Singer maintains, violate more (‘central’) preferences.\textsuperscript{199}

We are still well within the logical confines of such an argument if we suggest that by extension this means the killing of animals with \textit{an even greater sense of themselves as agents existing over time} is worse because it will tend to result in the violation of \textit{even more preferences}.\textsuperscript{200}

VanDeVeer’s position also suffers to the extent that it suggests acceptance of hedonistic utilitarianism. While the majority of his discussion concerns interests specifically, where this would be more comfortable for preference utilitarianism, he supports TFE by suggesting that ‘...long term disutilities... [are] very much a function of the psychological capacities of the beings involved’; here he suggests this means that focusing on interests can mislead one about the amount of harm being done in a situation – because a (more complex) person would suffer more in the long run.\textsuperscript{201} But the emphasis on psychological complexity seems to be quite a precarious route for a hedonist to follow; VanDeVeer’s stipulation that greater psychological complexity must ultimately lead to greater dissatisfaction in the face of pain does not ring true, and beyond this it seems all but obvious that \textit{less}

\textsuperscript{197} For present purposes it is enough merely to compare VanDeVeer’s position to the one Singer currently holds.

\textsuperscript{198} See for example Peter Singer, \textit{Practical Ethics}, pp. 20-1, p. 49, or his \textit{Animal Liberation}, p. 8, pp. 228-9.

\textsuperscript{199} See \textit{ibid.}, pp. 20-1, pp. 228-9 or his \textit{Practical Ethics}, pp. 76-7, p. 80, pp. 111-2. We shall return to the idea of centrality of preferences in the next section.

\textsuperscript{200} Indeed, Singer accepts the plausibility of views offered by Roger Scruton and Gary Varner that are roughly similar to this, although more suggestively qualitative. See Peter Singer, \textit{Practical Ethics}, pp. 103-4, p. 122.

\textsuperscript{201} Donald VanDeVeer, “Interspecific Justice”, pp. 70-1.
‘psychologically complex’ creatures often seem much easier to make very happy.\textsuperscript{202} Even if VanDeVeer is right about psychological capacities causing greater sensitivity, then, while this might provide the hedonist with reason to prefer helping humans it would also suggest that non-human animals are arguably better off – indeed we might end up with cause to discourage intellectual activities or to selectively breed children for stupidity. Singer, meanwhile, is much more open-handed about the possibility of greater suffering in a given context going either way.\textsuperscript{203} This puts him in a better position than VanDeVeer, since evidence of early or great trauma is also readily observable in the behaviour of a wide variety of non-human animals; a visit to the dog pound would in many cases be enough to confirm that human psychology is not a prerequisite for lasting mental trauma. Even then, inasmuch as long-term mental trauma can also fail to manifest in overt or easily interpretable behaviour it seems as though this is particularly unsafe ground to tread, and so we are better off not resorting to it.

If we adapt VanDeVeer’s account to refer merely to the number and kinds of interests a creature is likely to have, rather than going on to emphasise the feelings of satisfaction or dissatisfaction associated with realising or failing to realise their interests, we get a more plausible if still quite contingent picture of the value of ‘psychological complexity’. As noted before, though, this does not in itself advance the position of the preference utilitarian. It also, in remaining utilitarian, fails to address as wide a variety of intuitions about the good as emphasising consciousness potentially can, by failing to allow rationality to be its own source of ethical significance. Any utilitarian could potentially argue for the value of rationality insofar as it is conducive to sorting out which pleasures or preferences to pursue and in what order, so to speak, but it would still be value of a decidedly derivative kind. We can plausibly see rationality, though, as a significant development in the nature of consciousness itself; a manner of perceiving and of organising thought that allows a creature to understand and respond to the world in new and significant ways. In particular we can also see rationality as making a vital contribution to the very possibility of moral interaction and thought, and so as engendering a special degree or kind of value. So the valuing of consciousness-relevant abilities \textit{as the consciousness-relevant abilities they are}, for making the special contribution to consciousness that they do, allows us to consistently hold hedonic values and rationalistic standards to be their own sources of moral claims in a way that is part of rather than purely derived from what we take to be of fundamental importance (that is, consciousness).

This also suggests, as mentioned earlier, a coherent means of valuing certain species over others without being ‘speciesist’; that is, without distinguishing between species on the basis of morally arbitrary factors. At the same time, this really is no greater than Singer and VanDeVeer’s

\textsuperscript{202} Indeed, as we shall see again later, Peter Singer directs much the same complaint toward Mill - see the former’s \textit{Practical Ethics}, p. 92.

\textsuperscript{203} See Peter Singer, \textit{Practical Ethics}, pp. 51-2, or \textit{Animal Liberation}, pp. 15-6.
ability to suggest species membership tends to but does not guarantee whatever degree of moral weight. *Mere membership*, then, still is not in itself significant beyond suggesting the *likelihood* of greater (or lesser) capacities, since not all species members possess all the capacities typical to their species; someone suffering from severe brain damage might, for instance, have little or no capacity for rational thought. In this sense we are in the same position as both Singer and VanDeVeer, in appearing to suggest that some unfortunate humans are in some sense less or no more morally salient or valuable than some of the great apes and so on. As Singer perfunctorily suggests, though, this is a comparison in which all things will not tend to be equal; such people also figure importantly in the lives of other fully-developed or otherwise ‘cognitively normal’ individuals.\(^{204}\) The further possibility is also open to us, however, to suggest that there are less contingent duties of care that exist between people with severe mental disabilities and those related to them in various ways – parents, for instance, might rationalistically be seen to have greater duties toward their offspring insofar as they are responsible for their existence. One could also argue, in this vein, that those who are not morally accountable for a child’s conception, such as victims of rape, are under less or no obligation to care for their child – although this in turn need not suggest that a victim of rape can treat their child as morally insignificant, but only as being no more significant than anyone else.\(^{205}\) Parents of children with severe mental disabilities, too, would then still have special and weighty obligations to them (regardless of their preferences), and we as members of the wider community would have a general obligation, recommended by understanding and the flourishing conditions of human consciousness, to help make life for people in such a position easier.

It is also important to construe the sort of capacities we have been talking about broadly if we are to establish a plausible scale of value. It might not be difficult to accept that our understanding of the world is more valuable than that of a pig, but it would no doubt cause discomfort if one were to suggest that being colour-blind or short-sighted makes one less ‘morally salient’. While in response to this worry we might place less emphasis on variation of effectiveness *within* a given category, however, it still seems that a case can be convincingly made for each of the senses, for instance, being too distinct to subsume under the capability of ‘sense perception’. Does this mean that those entirely lacking a capacity, such as the blind, say, or the profoundly deaf, are in some way less morally valuable? The first thing to suggest here is that lacking the capacity for awareness that sight provides can be seen as a moral disvalue without this suggesting that the claims of the blind are themselves less binding. There are at least two ways this can legitimately be argued on grounds consistent with the position I have been here expounding. One begins by making the much-maligned move of appealing

---

\(^{204}\) See for example Peter Singer, * Practical Ethics*, p. 101.

\(^{205}\) Under such conditions, though, it could be wrong to *raise* such a child; to the extent that loving and prioritising parents are necessary for a person’s psychological well-being, a rape victim who was not prepared to love their child *as their child* (morally speaking, as an adopted child) would be under a generic obligation to do something to contribute toward finding the child a suitable home. I avoid here questions relating specifically to offspring in utero.
to species norms. Singer explicitly objects to this sort of move, insisting that his principle of equal consideration of interests is not supposed to be ‘a description of an alleged actual equality’ – to try and base moral relevance on what characteristics a creature actually has, he believes, yields counterintuitive results.²⁰⁶ What is more, we can get wildly counterintuitive results from according a creature the treatment deserved by the typical capacities of its species if we invert the typical structure of such an argument; rather than treating the mentally disabled as morally identical to biologically typical humans, for instance, suppose a mouse was somehow given the capacity for a human degree of rational thought and agency. It would seem wrong, perhaps even monstrous, to treat it as though it was no more morally significant than a normal mouse. However, the appeal to species norms need not be so clumsily direct; we can grant the fact that $x$ belonging to species $y$ does not guarantee that $x$ has the full suite of $y$-typical capacities while suggesting that species membership is an appropriate relation to ground claims of being a worse-off member of that species, which in turn could serve to bolster the strength of their moral claims if we hold the common belief that help should be preferentially distributed to those who are worse- or worst-off.²⁰⁷ In this way the base claim-strength of people suffering from consciousness-affecting disabilities could be seen as equal to or greater than the claims of biologically typical people.

The other way to respond to this sort of problem is by again suggesting, not incompatibly, that there is special weight attached to those capacities which enable moral or normative thought – that is, those capacities which are foundational to our ability to value, protect and promote the interests of conscious perspectives. The capacity for rational thought is an obvious candidate here, and while Singer suggests that pleasure and pain are prerequisites for having interests at all we might more cautiously take them as being highly salient motivational forces.²⁰⁸ The power of sight, however, does not make any meaningful contribution to the possibility of moral thought or concern. Thus people with consciousness-affecting physical disabilities such as blindness can be considered equal in the most important sense, as moral agents, while still holding that there is something inherently and importantly less fortunate about their position.²⁰⁹ As before, these people could claim to be made significantly worse off by their disabilities, allowing them to make claims of at least equal strength to those of biologically typical human beings.

There are two things worth noting here. The first is that we should understand the capacities with which we are concerned to imply some degree of effectiveness; sight for instance is a valuable

²⁰⁶ Peter Singer, *Animal Liberation*, p. 5. See also *Practical Ethics*, pp. 24-34. This is an awkward claim for Singer to make, since he of course does rely on an alleged actual equality: that of an equal capacity to feel pleasure and pain, which he understands to be ‘a prerequisite for having interests at all’ – *Animal Liberation*, p. 7. See also his *Practical Ethics*, p. 50.

²⁰⁷ This is of course a distributive principle that can sometimes but not always be endorsed by utilitarianism.

²⁰⁸ See Peter Singer, *Practical Ethics*, p. 50, or *Animal Liberation*, pp. 7-8.

²⁰⁹ Note also that it seems harder for utilitarianism or deontology to explain why a perfectly happy and well-adjusted blind person might nonetheless be in a less valuable position.
capacity inasmuch as the information it provides us meets a certain standard of reliability. This does not mean we need to value *useless* or *counter-productive* capacities, such as the capacity to hallucinate, although it is not clear how we should treat capacities that might fail to correspond to reality while still contributing to the general flourishing of a consciousness – as some might suggest, for example, the capacity for religious experience to do. The second is that while the move previously made involving particularly important capacities allows us to preserve a degree of moral equality between those with different sensory capabilities, it does admit of one peculiar though interesting implication. Suppose either that there is one true religion or that different religions have a roughly comparable grasp of some religious truth. Some religious people describe themselves as having an acute awareness of, as it were, a different or separate reality, or of facts about it. If this reality or these facts bear some important relation to morality then, if these people are right, there would be a *significant* sense in which they would be more morally valuable than people who were not able to perceive or understand this religious dimension. It is not clear how much we can glean from this possibility, however, since it is hard to see how people could prove that their religious sensitivities were not mistaken, and thus that any preferential treatment was justifiable. Much also depends on the nature and extent of religion’s supposed contribution to morality; if it is possible for secular philosophy to cover ‘most of’ morality, or even its most important or central parts, then the value added by powers of religious insight might not be sufficient to warrant a significant difference in the claim-strength or value of the ‘perceiving religious’ as opposed to the non-religious or non-‘perceiving religious’.

However, despite all this there is one persistent, fundamental difficulty faced by an account of morality with consciousness at its base, it seems; the awkward fact that consciousness itself seems to evade all attempts at proper description. Indeed it has been suggested that the ineliminably subjective character of consciousness means that it is impossible to fully describe in external, third-party terms. Nonetheless, the exact nature, causes and purpose of consciousness continue to be a topic of heated debate in the philosophy of mind to this day. Without wishing to become mired in problems which we could not conceivably resolve in the space we have here, then, a sparsely functional but intuitive definition will have to do, without robust explanation of what further non-moral ideas or assumptions it may or may not imply – this also being a necessary strategy insofar as we are stipulating consciousness can be of various kinds and degrees, while trying to emphasise what is common to all or perhaps most of them. So, in that spirit, let us take our eminently defeasible working definition of consciousness to refer to something like ‘an experiencing perspective’, and by extension

---

210 The capacity to perceive a religious reality should probably be understood here as being distinct from normal faith or religious belief as we know it. It should also be noted that if it were typical or common for people to have religious perception then non-perceiver’s claims could appeal to their being worse off. There might of course also be ‘superethical’ religious reasons for valuing or disvaluing either group.

211 See for example Thomas Nagel, “What is it Like to Be a Bat?”. 

72
let us then take understanding to be broadly that process or set of processes by which we come to know about, and so how to treat, a given consciousness, and/or consciousness as such - present consciousness and, to varying extents, potential and past consciousness.\textsuperscript{212}

3.3 Understanding and Non-Human Animals:

Now, I have said above that understanding and the emphasis of the conscious perspective can allow us to value some animals over others, by appealing to the greater sophistication of their cognitive capacities and thus the greater sophistication of their consciousnesses. The ramifications of this difference in value could be taken in different ways. VanDeVeer’s claim is that greater sophistication makes for a fundamental increase in a creature’s importance, justifying this importance by suggesting that greater sophistication makes for a greater number and variety of interests. He does not explain why we should care about the variety of possible interests. Singer’s position is that our pleasure- and displeasure-derived interests or preferences are fundamentally identical to those of any and all creatures that can feel pleasure and displeasure, while also holding that typical humans have other, further sorts of interests that make us, all things being equal, more important than the non-human animals with which we are familiar (where this is supposed not to allow us to straightforwardly disregard the similar interests of other species). Unlike VanDeVeer, though, Singer pointedly includes a somewhat qualitative element to his story. On one hand he makes a faintly Benthamite move by appearing to emphasise something like the intensity and duration of a preference; in describing the importance of our future-orientated interests he says that they are ‘absolutely central to our lives... [our] most significant desires’, where the implication is that these more biographical desires are more important than the most central and significant desires of what he calls a ‘merely conscious’ creature – that is, a creature without any or much sense of themselves as a distinct entity existing through time.\textsuperscript{213} On the other hand, he also makes an appeal to a very Millian story about the ranking of preferences.\textsuperscript{214}

Mill, of course, suggested a rough bifurcation of higher and lower pleasures, where he believed most or all people with proper knowledge of both would prefer ‘the manner of existence which employs their higher faculties’.\textsuperscript{215} Singer, for his part, suggests that this is not a felicitous move for a hedonistic utilitarian ‘because it just does not seem true that the more intelligent being necessarily has a greater capacity for happiness; and... the fact that, as Mill acknowledges, this

\textsuperscript{212} While it certainly is not a straightforward issue, intuition suggests that we could have at least some moral obligations toward the dead bodies of persons, for instance, or bodies that will or might become persons, such as people in comas and perhaps foetuses.

\textsuperscript{213} Singer is also making a quantitative claim here, as previously suggested. See his Practical Ethics, pp. 76-7, pp. 103-4 and p. 122.

\textsuperscript{214} Ibid., pp. 90-3.

\textsuperscript{215} J.S. Mill, Utilitarianism, p. 139.
capacity is less often filled... would have to be taken into consideration'. But Singer does not seem to feel that this need worry the preference utilitarian; he suggests that we can legitimately rank typically human preferences over those of horses, for instance, in that we can imagine having the experiences of a horse but still prefer those of humans. However, this suffers from two of the same fundamental problems as Mill’s argument. Indeed, Singer suffers from the first such problem to an even greater degree; while Mill makes a contentious and unproven empirical claim, Singer apparently makes an unprovable claim on the basis of a nigh on unintelligible thought experiment. The second problem is the sheer contingency of the argument’s assumed result; even if it is true that most or all people would prefer the more distinctively human or ‘higher’ pleasures, we seem to be left with no reason to criticise or object to those who would choose differently. The mere fact that most people prefer $x$ has in itself no necessary bearing on the relative inappropriateness or worth of $\neg x$ – so if we are to draw such strong hierarchical conclusions from such an argument we must do something to provide reasons supporting the preference of the majority. And it is reasons for this presumed majority preference that both Mill and Singer do nothing to supply. Singer simply leaves the matter unresolved, saying ‘the question remains open’, and admitting that

In general, it does seem that the more highly developed the mental life of the being, the greater the degree of self-awareness and rationality and the broader the range of possible experiences, the more one would prefer that kind of life. Our emphasis on consciousness itself, meanwhile, provides a perfectly coherent and plausible explanation for why the majority should have such a preference, and furthermore affords us the capacity to criticise those who fail to share such preferences. So here we have an improvement on the utilitarian position. One might well wonder, in fact, why Singer himself fails to make the move from valuing preferences or interests to valuing preference- or interest-havers. At least part of the reason why Singer’s position is not closer to the one I recommend, it seems, is that Singer appears to conflate the notions of consciousness and sentience, where sentience is the basic ability for sense-perception and is particularly understood by Singer to refer to a creature’s capacity to feel pleasure and pain. To this extent it is not surprising that Singer fails to see rationality or indeed Mill’s general ‘higher faculties’ as extensions of and perhaps as improvements upon a basic characteristic – consciousness – of which ‘sentience’ is a more rudimentary constituent. But identifying Singer’s mistake here, though,

---

216 Peter Singer, Practical Ethics, p. 92.
217 Ibid., p. 91 – note that Singer believes horses to have some sense of themselves as distinct entities existing through time, and so that this is a more specifically qualitative claim than the one described in the previous paragraph.
218 Ibid., pp. 92-3.
219 Ibid., p. 92.
does little to shed light on his other fundamental misstep of valuing preferences rather than the kinds of things that can have preferences.

We can see, in any case, that Singer’s semi-qualitative reasons for preferring distinctively human interests are unreliable in a way that mine are not. There are also some further ways in which the emphasis on consciousness, I would suggest, departs from the utilitarian picture he describes. For example, recall that Singer suggests the pleasure- and pain-derived interests of humans and non-humans are themselves equal. There are two possible corrections we could make to this claim on the basis of the position I am here endorsing. One would be to suggest that such basic interests could themselves, through being interlinked with the more sophisticated cognitive capacities of humans, be more sophisticated; roughly, the thought would be that we are able to feel more kinds of pleasure and pain in response to a greater amount of more complicated or subtle things, and/or that we can feel pleasure and pain themselves in a more 'cognitively dense' fashion. This suggestion, particularly its second branch, would require thorough empirical research in order to demonstrate for instance that our nervous systems and brains are linked in ways appropriate to suggest such a picture. The other possible correction is to suggest, not incompatibly, that the mere fact of pleasure and pain figuring in the mental landscape of a more cognitively sophisticated – and so more important – creature is sufficient to make that pleasure and pain more morally important even if it is no different to the pleasure and pain of other animals and has no wider ramifications. I assume that one or both of these suggestions is correct, and so I differ from Singer in claiming here that one of two formally equal claims can have a lower default value or weight merely because it is made by a creature with fewer or substantially less effective (moral) capacities for experiencing the world.

Where does this leave us? I submit that the observable facts from biology and behaviour in mammals are by no means exhibitive of such great differences that we have reason to think human bodily pain is qualitatively that different from the pain of a rhinoceros or cat, excepting what elements of it that might be due to our more sophisticated psychology. Nevertheless, being for example in a thoughtless, panicked state of all-consuming terror remains a seriously bad thing, and certainly not a state requiring advanced mental faculties. To this extent I would suggest that at least the stronger or more central claims of non-human animals still count as having considerable, if characteristically overridable, weight. Clearly there is a fundamental problems facing this sort of claim; owing to the differences in our physiology we cannot imagine what it is like for a dog to smell the wind or for a bat to navigate the confines of a dark attic, and if there are special emotions arising from such experiences then they are closed to us. So our already imperfect means of interpersonal interpretation are far more unreliable when applied to different species. But again, it seems fair to say that at least a basic range

---

220 In case it is unclear, what is meant by ‘distinctively human interests’ is those interests we have which are not shared by the non-human animals to which we are accustomed; these human interests might however be shared by aliens, for instance, with roughly comparable capacities to us, and ‘human’ should generally be here considered to also apply, by extension, to such creatures.
of emotional states are shared between us, and to this extent thick understanding can often be appropriately and usefully exercised toward them, for example to provide us with a phenomenologically rich imagining of animals’ pain that should, if we are non-egoistically rational in the way understanding supposes, make us more likely to take that pain seriously, where historically speaking the more natural reaction to animal suffering has been to disregard it as inherently inconsequential.221

So we have said both that non-human animals are legitimate sources of moral claims on us and that their claims should be considered weaker than our own corresponding claims. We can adhere to this picture while borrowing something like VanDeVeer’s hierarchical picture of interests as divided into the basic, serious and peripheral. This could lead us to expect, with VanDeVeer, that the non-peripheral claims of a human being could never be outweighed by the corresponding interests of a typical non-human animal, and so that should a person and an animal have a mutually exclusive interest in something, the person’s claim should in every case win out. However, like Singer we need to maintain an open mind in the event of any particular clash between the interests of humans and non-human animals. Suppose a man and a dog both have mutually exclusive desires for a certain ball lying between them. Understanding, by looking at the broader picture, tells us that given the difference in cognition and lifestyle between the dog and the person, and how many opportunities for real pleasure either has had and is likely to have, it is quite possible that the instance of ball-having will have more significance for the dog, even if the dog itself has a comparatively rudimentary sense of biography with which to appreciate such significance. It would not necessarily be impermissible for the man to take the ball in this sort of case, but it would certainly I think be callous and lacking in understanding, and so blameworthy to some degree. There are two things worth noting here – first, that this is a case where, even if we are better motivated by imaginatively entering into the dog’s enjoyment, a lot of the work is being done by our awareness of facts the dog cannot understand (i.e. thin understanding). And second, this could also be seen to again introduce a non-utilitarian distributive element to our story – although we shall leave an examination of this sort of move until we discuss aggregation more specifically in the next section.

Another difference between Singer’s preference utilitarianism and the sort of ethical stance that I suggest understanding recommends concerns the idea of replaceability. Essentially, Singer believes the painless killing of animals is not bad when the animal concerned lacks future-orientated interests, insofar as the subsequent pleasures and/or interests – no longer to be satisfied – can simply be replaced by creating another such creature.222 He supports this by suggesting that for such a non-biographical creature, living always in the moment, being anaesthetised or falling asleep is no

221 As Singer suggests, for example, those parts of our nervous system which relate to the feeling of pain are in evolutionary terms very old and widely shared. See his Practical Ethics, p.59, or Animal Liberation, pp. 11-5.
222 Peter Singer, Practical Ethics, pp. 105-7, p. 120, and Animal Liberation, pp. 228-9.
different to dying and being replaced. \(^{223}\) It might seem that understanding is in prima facie agreement with Singer here, but while thick understanding does not appear to yield a distinction we might appeal to the understanding of wider facts about such a creature, facts not necessarily accessible to it, in suggesting there is something wrong with such a claim. Though these creatures are none the wiser, we ourselves can see the ending of one chain of consciousness-instances, supported by one spatio-temporally continuous body, and the beginning of a separate one which might not be subjectively distinct. This idea of ‘replaceability’ strikes me as perverse; it goes against the situated, indexical nature of the conscious perspective that understanding would have us imagine, even if its internal states are not consciously linked. It may well be that the killing of an animal with no future-orientated interests is in some way less bad, but nonetheless understanding would have us ask not just ‘why have you ended a life?’ but ‘why have you ended this life?’ In a similar way I would say we should condemn a murderer even if they strike silently and instantly, to immediately replace their victim with a perfect copy. \(^{224}\) This position also allows us, unlike Singer, to distinguish between creating a new person and saving a temporarily unconscious man who has lost all his memories.

Understanding also, by recommending the value of consciousness itself, gives us reason not to use present livestock to very gradually breed animals of so little awareness as to barely warrant the term ‘consciousness’ – even if killing such animals would be less bad. It is not clear how other theories could object to this if it was done so slowly as to make no perceptible impact on the lives of the animals concerned. We see here that the value of consciousness can give rise to roughly three levels of good: at base, that which is conducive to the flourishing of consciousness, by extension the goods recommended by or crucial to the various consciousness-supporting capacities a creature possesses, and in particular those capacities which allow appreciation and support of the value of consciousness (where these can overlap or interact). \(^{225}\) There are no interests arising from a farm animal’s specific capacities that need be infringed by continuously breeding marginally inferior livestock, but our treatment of them is still constrained by wider facts about the value of consciousness itself. Apart from the caveat already mentioned regarding past, future and potential consciousness, we can take it that the non-conscious, be it alive or inert, does not qualify for moral relevance and thus we are under no immediate obligation to exercise even purely theoretical understanding towards it – although we would still have an indirect obligation to understand non-...

\(^{223}\) Peter Singer, *Practical Ethics*, p. 86.

\(^{224}\) Singer asserts that beings with future-orientated interests are not replaceable, but from the logic of his position this could only be because their exact interests could not be reproduced – if they could be, for instance in a perfect clone with formally identical memories, beliefs and desires, it would seem he could not restrict replaceability in this intuitively appealing way. See his *Practical Ethics*, p. 111, p. 120 and *Animal Liberation*, p. 228.

\(^{225}\) While we might think rationality a capacity crucial for moral agency, this position could also allow us to attach lesser value to motivationally important capacities such sympathy or thick understanding, which could be possessed to some degree by some non-human animals.
conscious things which seem as though they could have some relevant effect on conscious life (as also suggested with regard to the second level mentioned above).  

I have said that our treatment of conscious creatures should be constrained by the value of their consciousness. Even if we assume the claims of non-human animals have less weight than ours, then, I would further suggest that except in circumstances where our basic interests are at stake, understanding tells us that it is wrong to eat an animal that does not want to be eaten. Inasmuch as consciousness is a directed perspective on the world it would seem to contain a basic and on-going claim to existence; it’s busy being aware. Thus by and large we need not check to make sure that animals are interested in seeing tomorrow – there is a fundamental and universal reason for us not to interrupt a given consciousness in the absence of good countervailing reasons. If a person’s life did depend on the consumption of animals then she would be entitled to kill and eat such animals as she required for a healthy subsistence, with the proviso that the agent put due thought into painless means of slaughter according to the degree and kind of the animals’ (morally relevant) capacities. At the same time it is also consistent with this position to hold that some conscious animals could be so rudimentarily conscious that we could kill them for relatively trivial reasons – perhaps shellfish, for example.

It is also worth mentioning Martha Nussbaum here, as an example of understanding gone awry, or of a superficially similar argument taken too far. In her *Frontiers of Justice*, she provides an alternative approach to animal welfare, one heavily inspired by Aristotelian virtue ethics. She begins by providing a list of ten thinly-conceived capabilities necessary for human flourishing, as an alternative to traditionally conceived rights. In her penultimate chapter, she extends her notions of dignity and the capacity for flourishing to nearly all living creatures, and constructs a new list for non-human creatures which is intended to regulate our relations with them. Understanding arguably has here an impossibly large part to play; Nussbaum’s capabilities approach demands of one not only to recognise genuine cases of flourishing as one comes across them, so to speak, but further to plan out a map of what all such cases might involve or require. The pressure is on for the capability theorist, then, to actively try to understand as many kinds of people (and animals) as possible, a requirement understanding would avoid in the name of being psychologically realistic. At the same time, her claim that human practices affect the natural world at large, and therefore that we have some duties to wild

---

226 These would be duties to the conscious to understand (some of) the non-conscious, not duties to (some of) the non-conscious to understand it.

227 The example of shellfish is suggested by Singer himself – see his *Practical Ethics*, p. 60, or *Animal Liberation*, p. 174. Technically it cannot be denied that our knowledge of the internal states of shellfish just might not be sufficient to guarantee this is right. I can only suggest that such an objection presumes an excessive degree of scepticism in the face of the sorts of things we know about biology.

228 Martha Nussbaum, *Frontiers of Justice*, pp. 76-78.

animals generally, should I think be allowed. Nussbaum also specifies that the capabilities it is reasonable to expect a society to provide depend greatly on physical conditions; thus for instance a poor country should not be expected to immediately solve its problems, but rather to instigate changes that will ultimately bring about a better state of affairs. So in fact her vision is not quite as dramatically unrealistic as one might think at first blush, since it allows room to suggest our present duties could be fairly restricted in light of what we are at present reasonably capable of doing. However, the level of stability, affluence and technological advancement that would be necessary before a society could properly defend the welfare of wild animals in the manner she recommends, particularly without fear of doing serious ecological damage or neglecting important person-regarding concerns, seems so far off as to be effectively impracticable.

We are still left with something of an awkward case here, though. If understanding suggests that an animal’s suffering should be taken seriously as something morally bad, even if we don’t go as far as Nussbaum would wish, does that not still leave us with a vast amount of suffering in the world which we should reduce? It was established in chapter one that understanding requires of one at minimum to know about the cultures with which one personally comes into contact. We should extend this to include cultures affected by us, and perhaps even to unrelated cultures we are merely aware of as needing our help. The kind of understanding involved here may be relatively limited, though, since typically the ways we can influence such effects or provide aid are more indirect, they have less to do with the individual persons being harmed, and thus are somewhat less affected by cultural or biographical variation. So while it remains appropriate to imagine the perspective of a child labourer in a sports-brand sweatshop, inasmuch as it can spur one to action by making one appreciate the force of the reasons involved, and encourage one, if responsibly minded, to discover whatever available information there is which could fine-tune one’s response, the chances are that in such instances people can still only give fairly generic responses to generically conceived problems. But already the thought of being required to (sensitively) help other, remote nations or cultures would be enough to put some people off – what then about our responsibilities towards wild animals, when indeed the reliability of our imaginations is more tenuous?

Our defence here would have to be similar to that which we offered for Nussbaum: persons are most important, animals’ claims compete with other more pressing concerns and we do not have the resources to properly plan and manage broad scale intervention in the animal world. If we lived in

---

230 Her later suggestions that we police wild animals in order to reduce their suffering, or that we re-educate lions to chase balls rather than gazelles, are however bound to strike most people as, again, unrealistic. Ibid., pp. 370-1.

231 This becomes most apparent when one considers what the eighth capability on her list, ‘Other Species’, entails: ‘the gradual formation of an interdependent world in which all species will enjoy cooperative and mutually supportive relations’, ibid., pp. 399-400.

232 This is not to overlook the fact that cultural sensitivity, so to speak, is often important in such cases – it would be wrong for instance to carelessly and avoidably send pork to a starving village inhabited entirely by people whose deeply held religious beliefs forbade the eating of such meat.
a futuristic utopia where anything was possible then it would be wrong of us to allow suffering in the animal world, but until such time understanding must instead act merely to regulate what broad effects we ourselves have on wild animals, and to more closely regulate our treatment of domesticated animals. This can realistically be achieved by the division of ethical labour, so to speak; for most actual agents this would work out not as an ethical responsibility to alter their own (apparently non-existent) treatment of distant wild animals but instead to elect those with environmentally sensitive policies, or to prevent those in power from enacting or failing to discontinue policies which are excessively deleterious to some habitat at comparatively little gain to ourselves and so on. Contrary to the vision offered by Nussbaum, there need be no direct intervention between rivals in nature for the foreseeable future. This is a happy fact when it is not obvious ‘whose side’ to be on; should we care more about an owl or a mouse? We cannot treat one as blameless victim and the other as unprovoked assailant, obviously, because these are morally loaded terms which require a complexity of agency unavailable to the creatures concerned. Presumably we might argue that the pain of the mouse is worse than the owl’s gain, but this approach is complicated by the fact that the owl has no other means of subsistence, and furthermore it might in many cases be difficult to tell which creature is more cognitively advanced. In cases of impenetrable uncertainty, understanding advises us to abstain where possible – so in the case of the natural world, since virtually all the suffering is between non-agents and the cost of active intervention so high and difficult to properly execute, this seems to be one problem we are entitled to set aside.

Before continuing to the next section, however, it is worth returning here to a case considered in the previous chapter. Recall that in arguing against Nodding’s requirement of acknowledgement or reciprocity we imagined the discovery on another planet of sapient crystals with whom we could not communicate. Let us presume that the crystals’ language is for some reason totally indecipherable to us – as indeed it might be simply because of their totally alien perspectives, as suggested in Stanislaw Lem’s Solaris; it need not be due to technical complexity only. The point here is to say that not only can we not imagine their perspectives, they can give us no reports with which we could try to construct a purely theoretical understanding of their psychology. This naturally raises the question: what use is understanding in such a situation, where it is established that we have no chance of imagining or comprehending anything like the perspectives of such crystal entities? I claimed before that insofar as the first astronaut to discover these living crystals would supposedly have no reason to believe they were alive, and no reason to check this assumption, he could not be morally faulted for harming them, and this is precisely what understanding would tell us through highlighting the impossibility of his knowing otherwise. I also mentioned the possibility that there could be no way for anyone even to determine that these crystals were alive, apart from a note having been left by some reliable source. Inasmuch as we have reason to believe the note, and perhaps insofar as there are not countervailing reasons of sufficient urgency to press us into doing otherwise, we would then have to
respect these crystals as having moral worth. But how can we know what sort of treatment constitutes respect when we cannot understand anything about what it is like for them? Understanding then tells us, I believe, that we must leave them alone, that the best we can do for them is to not interfere in their lives. This is not an empty or vague recommendation, however, but rather a positive and unambiguous requirement that follows from the very idea of understanding, from the demonstrably doomed attempt at understanding and the consequently evident fact that there is a clear and fair possibility of unintentional but to this extent foreseeable harm. If harm should come to them when it was within the reasonable limits of our power to prevent it we could not be blamed for this, unless again it happened through something we had some reason to believe would be harmful to them.

This does leave us with a minor conundrum, however. Suppose a giant meteor is hurtling towards the planet inhabited by these aliens, one large enough to destroy the planet entirely. Suppose, too, that we can destroy this meteorite from a safe distance, at relatively little cost to ourselves. What would understanding have us do? In the case of beings we could to some extent understand, we could weigh the bad and good from their perspective against the bad and good from our perspective to see that, of course, we should save them by destroying the meteor. But what if these aliens are, in fact, a race trapped in something like constant agony, with no goods to enjoy whatsoever? What if, in other words, the meteor represents their only means of escape from lives not worth living? It makes little or no sense to imaginatively draw a comparison between the good of being saved and the bad of being doomed to perpetual torment for such beings; we cannot weigh the alternatives because we cannot even begin to know if we are doing so accurately. In the end it would have to come down to some kind of impossibly broad and general theory about how things are likely to be for them, or beings such as them, or beings generally. In short, there does not seem to be any entirely satisfying solution to such a problem. However, while this problem does show the usefulness of understanding to break down quite spectacularly, tellingly it still cannot undermine the basic need for understanding – of our own limitations at the very least – and more tellingly still it only shows understanding to collapse along with the very possibility of right action.

3.4 Understanding, Aggregation and Autonomy:

In his Moral Thinking, Hare provides an everyday illustration of critical moral thought – he imagines wanting to leave his bicycle in a space blocked by someone’s car, where he and the driver decide how best to resolve such a problem. In such a case, the story goes, the cyclist imaginatively entertains the perspective of the driver, who does not want to move his car, and weighs this sentiment against his original desire for the car to be moved, where the driver does likewise – the stronger desire wins and,

---

233 At the same time, it might still be good to make technically empty or fruitless gestures of respect; in a similar way one might light a candle in private for someone far away, knowing it would have no effect on them. But if this is a good then it is more probably to do with supporting either one’s own or a general tendency to be respectful, say, rather than a good which relates to the intended ‘recipient’ in any significant way.
Hare concludes, it is decided that the car should be moved. ‘We can see here...’ he says, ‘how the requirement to universalise our prescriptions generates utilitarianism’. Now, this comment is of course strikingly unwarranted, since the plausibility of whatever form of utilitarian calculus in entirely trivial situations can do little or nothing to suggest that utilitarianism as such is the correct moral theory. Peter Singer is, for his part, more cautious, explicitly rejecting the idea that morality’s universalising nature entails preference utilitarianism and saying instead that it is a ‘first base’ suggested merely by ‘simple, pre-ethical decision making’ – since one might coherently overrule preferences with an appeal to impartial rules such as rights. We can disagree even with Singer’s more restricted claim here, however, because we have looked at essentially the same process of imaginative comparison that both Hare and Singer describe and seen that it can underpin a fundamentally different approach to ethics, one that centres on a different, heterogeneous kind or collection of goods. It is true that we have discussed goods arising from different forms of consciousness, and it might well be the case that these could be translated into talk about preferences, or grounds for preferences, but even then we have selected a different fundamental value (i.e. consciousness as such) which allows for a stronger and more coherent ranking of goods than utilitarianism does. There remain two important areas in which we have further room to distance ourselves from the utilitarian, however: our approach toward the aggregation of multiple persons’ goods and the value we attach to autonomy.

Remember, Hare suggests that through internalising other people’s preferences ‘multilateral cases... reduce themselves... to intrapersonal ones’. This provides a seemingly innocuous example of how utilitarianism can fail to respect the separateness of persons, by too easily seeing people’s preferences or pleasure as significantly detachable from the specific people in whom they are manifested, and thus how preference satisfaction or pleasure are distributed. Once we shift our focus from a basic understand of the goods conscious creatures enjoy to the valuing of conscious creatures as such, however, the path is open to make some concessions to maximisation while retaining a sound rationale for placing certain constraints on how and what we try to maximise. Some maximising actions might coherently be incompatible with the basic valuing of consciousness, that is, since consciousnesses cannot be summed in the way that pleasures can; my pleasure and yours are themselves naturally or closely comparable, but part of the fundamental nature of consciousnesses is that they are indexically distinct, and as such inescapably separate. An action’s causing radical differences in outcome for separate individuals, for instance, can therefore be legitimately held to be an important fact in a way separable from how such an action figures in an impartial, overall calculation. This allows us to accuse Hare, Singer and other utilitarians of, unlike us, treating people

234 Richard Hare, Moral Thinking, p. 111.
235 Peter Singer, Practical Ethics, pp. 12-4. At the same time, Singer there leaves us with the implication that finding independent support for such principles is not an easy task.
236 Richard Hare, Moral Thinking, p. 110.
and other conscious animals – the things that really matter – as mere receptacles for things we see as themselves actually being of secondary importance. This is despite the fact that Singer specifically attempts to avoid such an accusation by pointing to pleasure’s necessarily depending on consciousness and the supposedly irreplaceable nature of future-orientated interests – a defence that fails since the former move suggests nothing and the latter move is delivered in the form of a relatively unsupported and counterintuitive assertion.²³⁷

Understanding itself – we might call it here broadly that process or set of processes by which we come to know about a given consciousness and the nature of its subject experiences – certainly does not entail the legitimacy of aggregation, since there tends to be no one individual enjoying the sum total of benefits yielded by a particular course of action. This might be seen to invite a Scanlonian picture where the reasons counting for or against some action must all be ‘personal’; that is, that the reasons for or against something must come from the separate claims of separate individuals, where in many cases this would require us only to compare the strongest reason someone has for something with the strongest reason someone has against it, leaving the number of people affected largely irrelevant. However, understanding also does not contain any special reason to completely dismiss the possibility that we might have other ethically coherent reasons for summing multiple people’s pleasures or what have you. Insofar as the valuing of consciousness would seem to generate agent-neutral reasons for protecting and promoting the existence and flourishing of consciousness, for instance, this could be seen to give us reason to prefer actions that lead to the greatest good for the greatest number of the more or most highly sophisticated consciousnesses. Note though that we already have an advantage over at least the hedonistic utilitarian here, since we cannot fall into the trap of the repugnant conclusion. It might be the case, though, that the valuing of consciousness leads to a requirement to bring about a world that supports the greatest number of people comfortably possible – if this is true then I doubt that it is problematic, as while it is perhaps a non-obvious obligation it is also not one that I could understand anyone finding positively objectionable, at least bearing in mind the caveat that it need not amount to a personal obligation to have as many children as possible, but rather the largest amount of children that one could reasonably expect to raise well, as well as such beings deserve.²³⁸ This, in turn, could be a plausible limiting factor because of its possible adaptability; the value of consciousness encourages us to bear children when we think it practically and emotionally feasible, but could also through its importance permit us to have children at times when we unavoidably cannot assure them certain significant goods, as might be the case if for example one was the member of an enslaved minority.

²³⁷ See Peter Singer, *Practical Ethics*, p. 106 and p. 111. See also the previous section’s footnote in relation to Singer and replaceability.

²³⁸ I should repeat here that children should not be raised by parents who do not properly love and prioritise them – so if one genuinely could not find the desire in oneself to have children at all then one would not be required to do so, at least provided that other people were having children.
To say this, though, is still to leave open the question of what forms of maximisation we should consider to be incompatible with the basic valuing of consciousness. However, recall that earlier it was suggested a consciousness consists of a directed perspective on the world which thereby necessarily contains its own (at least minimal) claim to existence. This means that in valuing consciousness we must also value, to some extent, a given consciousness’s self-directedness – or, in more familiar terms, its autonomy, where what we appeal to here is in fact a much more minimal and potentially or partially arational form of autonomy than that which tends to be valued by Kantians and others. It is a form of autonomy, in other words, which is better described as ‘internally directed’ rather than ‘self-governing’, one which does not rely on a Kantian picture of deliberate rationality as an escape to freedom and moral worth from the deterministic mores of contingent natural inclinations. On my picture we can still put a higher value on this kind of characteristically rational autonomy, of course, but in so doing we can still respect the non- or sub-rational directedness of less cognitively sophisticated animals. This achieves a high degree of plausibility by presenting us with very strong but conceivably overridable reasons not to treat people and animals in certain ways, where Kant’s purely agent-relative rules are counterintuitively inviolate and the utilitarian makes sacrifices too easily by failing to see any inherent worth in autonomy.

This sort of strategy also enables us to elucidate a curious phenomenon which Derek Parfit mentions but neglects to explicate – roughly, the idea that actual consent can sometimes make a moral difference entirely independent of what is considered permissible according to its not being rationally rejectable. To put it more clearly, Parfit suggests that it can be wrong to contravene the wishes of someone who, though fully informed and thinking clearly, is deliberately insisting on something irrational (at least, presumably, assuming that it affects only themselves). This is in the context of his suggesting what he calls ‘the Rights Principle’, according to which ‘everyone has rights not to be treated in certain ways without their actual consent’, where he offers very nearly no guidance on how we should pick out which ‘ways’ should be covered by such a principle.239 It is not the Rights Principle we need examine, however, but a subsequent elaboration he offers in connection with it, saying

When we ask whether people would in fact consent to some act, this is quite different from asking whether these people could rationally give such consent. We might know that certain people would not in fact consent to some veto-covered act, even though it would be irrational for them to refuse consent. In such cases, we might say, people have a right to be irrational, and to suffer the effects.240

---

239 Derek Parfit, On What Matters, Vol. 1, p. 194. The only guidance he offers is to suggest that such acts should be ‘narrowly described’.
240 Ibid., p. 195.
This is the phenomenon to which I allude above, something we might think of as ‘counter-rational consent’, in that it involves (refusal of) consent on the part of a rational agent who, while in a sense being rational or unobstructedly exercising their rational capacities, is consciously and deliberately making a decision which apparently is not rationally endorsable. But where Parfit is unabashed in leaving this as an intuitively derived and more or less mysterious moral opinion, it can be more fully explicated by the consciousness-valuing system I have argued that understanding suggests. This is because the valuing of consciousness as such allows us a plurality of goods (that is, those arising from and affecting consciousness per se and the various forms it takes, in a way which makes rationality important but not exclusively nor necessarily supremely so) and the watering-down of autonomy as a rationality-derived moral characteristic. It is not that we should let people have their rational way, so to speak, but rather that we should let consciousnesses have their way in a hierarchically structured fashion which shows preference to rationally endorsed or produced behaviour over non-rationally endorsed or produced behaviour. It is because people have the capacity for rationality, and to willingly and knowingly make concessions to irrationality, that at least their non-other-affecting irrational desires should be respected – and indeed treated as more important than the arational or irrational desires of non- or sub-rational creatures.

3.5 Closing Remarks:
It was my main aim in this chapter to illustrate understanding’s powers of versatility and insight, by showing that it has something interesting and useful to say about a wide range of problems and discussions, as well as to highlight the differences between the implications of understanding and utilitarianism – these things should by now be apparent. But the majority of what has been said thus far pertains to understanding’s role in ethics, and it is not with ethics alone that we are here concerned. It has been suggested by some that we generally respond to fictional characters in a way analogous to how we respond to real people, although it is of course important to note that this is not always the case; a typical adult responds very differently to cartoon violence than genuine violence, for instance, and it looks as if participatory forms of fiction such as storied computer games often encourage what would otherwise be psychotic violence toward fictionally innocent people. This can sit oddly, too, with the idea that fictional characters tend to operate according to the same psychological and moral laws that we do, an idea which naturally is relevant to the argument that understanding can be appropriately and indeed usefully applied to fiction. Nevertheless, there remains a clear prima facie case for the understanding agent’s needing to take an interest in fiction, since fictional scenarios appear to at least afford us with all manner of potentially valuable thought experiments, as well as the opportunity to, as it were, experiment with emotional reactions in a relatively contained and harmless manner. Because of this, in the next chapter we shall explore
aesthetic cognitivism, the idea that art’s primary value lies in its capacity to enrich our understanding of ourselves, others and the world. We shall do this both in terms of examining how the cognitivist claim is possible, indeed plausible, and how it should be more precisely understood, with particular reference to kinds of knowledge and asymmetries in our reactions to life and to fiction.
CHAPTER FOUR

4. UNDERSTANDING AND AESTHETIC COGNITIVISM

4.1 Understanding and the Cognitivist Gap:

It is interesting to note some of the recent work in philosophy and psychology very much supporting something like the cognitivist idea that fiction’s ability to teach us things is important to its value. In the first chapter we looked at ST and TT, the two theories of interpersonal understanding that competed through much of the last twenty five years. In the last decade, however, Daniel Hutto has put forward his own proposal, the Narrative Practice Hypothesis, which suggests that ST and TT do not describe something worthy of being called folk psychology, but rather that they pick out some of a ‘raft of abilities’ necessary for developing the more robust practice of genuine folk psychology. Folk psychology is, for Hutto, ‘a distinctive kind of narrative practice’ that describes people acting for reasons, where ‘proficiency in making isolated propositional attitude ascriptions... is not the same as knowing how these combine to form reasons’. This last qualification is important to bear in mind, as much of what else he says bears more against TT than ST. To iterate the point more fully, Hutto says of psychological principles that they are not learned as a series of rules but by [seeing] them in action, through narratives, in their normal contexts of operation... narratives not only show which features are constant to folk psychological explanation but also, importantly, what can vary in such accounts – such as the particulars of what a person believes and desires, how these attitudes can change over time and why, and also how character, history and other commitments might impinge on why a person acts as they do.

Stories that involve people acting for reasons are also useful because they are ‘public objects of mutual attention’; rather than the inherently private workings of ST and TT, reason-featuring narratives can be debated and compared interpersonally, allowing for the creation and spread of cultural, as opposed to merely personal, norms. This gives a good, cognitivist-friendly reason for the popularity of soap operas, and something of an answer to what Noël Carroll has called the ‘paradox of junk fiction’, in his wondering about why people enjoy fictions they consciously evaluate as being of poor quality; they provide a safe and approachable ground for the communal exercise of norms, reason explanations and values. They are harmless grist for the gossip mill, to be ground by

241 Daniel Hutto, Folk Psychological Narratives, p. xiii.
242 Ibid., p. x, 28.
243 Ibid., p. 29.
244 Ibid., p. 29, 38, see also Daniel Hutto, “Folk Psychology as Narrative Practice”, p. 12.
the behavioural expectations and general social sanctions which hold us together, which we can thereby share, debate and confirm.

Hutto and some of what might be loosely called his supporters say more that is in line with the general project or viewpoint adopted in this thesis. In particular, both he and Matthew Ratcliffe provide stories about the inadequacy of the traditional Humean picture of psychology, that which explains behaviour in terms of the interaction between desire and belief. Hutto, for example, describes a man eating acorns, claiming we would still be left puzzled even if it was explained to us that the man had a desire to eat acorns, coupled with the belief that what he possessed were acorns – this would not be sufficient to ‘explain an action in the strong sense of making it intelligible’. Similarly, regarding an example of his own Ratcliffe writes that ‘no simple statement of belief or desire will do the trick. What is required is an appreciation of how or why Bernard has come to believe and desire what he does.’ As mentioned in the first chapter, understanding is not limited to a sentential reconstruction of another agent’s motivating beliefs and desires; to thoroughly understand another is to acquire some sense of their internal life, typically in a much broader and imaginatively rich way than most analytic philosophy anticipates. This can be especially important in the sort of case Ratcliffe mentions ‘where belief incorporates feelings’. He unpacks such a case in the following:

Take a case where someone receives good or bad news and says ‘I just can’t quite believe it yet’. What is going on here? Does she believe that p or believe that not p? One can put a tick next to a sentence and assert that it is true but, at the same time, not fully accept the state of affairs in question. Genuine conviction, in some cases at least, requires feeling. Consider the belief that one will die... One might sincerely think that one dispassionately accepts the inevitability of one’s own death, but on those occasions when it fleetingly ‘sinks in’, it can seem suddenly unfathomable, unbearable. Both in unfeeling assent and in feeling realisation, one ‘believes that p’. But the two kinds of attitude are very different indeed.

Others, unconnected to this specific debate, have argued for related claims – Martha Nussbaum, for example, argues persuasively that her feeling of grief over the loss of her mother contains as part the propositional content ‘my mother is dead’, and that ‘the experience itself involves a storm of memories and concrete perceptions that swarm around that content... [adding] more than is present in it... [that] emotion is, then, cognitively laden, or dense, in a way that a propositional-attitude view

---

245 Daniel Hutto, *Folk Psychological Narratives*, pp. 7-8. The discussion here strongly echoes the spirit of G.E.M. Anscombe’s describing our need to assign an intelligible ultimate desirability condition to any apparent piece of practical reasoning – see especially her *Intention*, pp. 70-73.


247 Ibid., p. 391.

248 Ibid., pp. 391-2.
would not capture’.\(^{249}\) It seems then that there are thoughtful emotions and emotional thoughts, that it cannot be denied there are important mental/attitudinal states that straightforward sentential/propositional representation are not able to properly capture. As Jenefer Robinson says, ‘emotional understanding “regestalts” the world in a global way’; some aspects of our internal lives do not function as propositions so much as affecting, to describe things on the most crude and uninformative level, what propositions we are likely to assent to and in what way.\(^{250}\) These mental phenomena might range from moods – dispositions often too vague to be entirely put into words – to something like what Ratcliffe calls ‘existential feelings’:

Some of the psychological states we call ‘moods’ are intentional states directed towards a wide range of things, whereas existential feelings are ‘pre-intentional’... [they are] an experiential background that determines the kinds of intentional state that are possible... not a way in which a particular thing or even a range of things appear significant... [but rather] a shift in the kinds of significance that experience incorporates... Many people suffering from severe depression report a similar transformation of experience, where the possibilities of pleasure, practical significance and emotional communion with other people are removed from experience.\(^{251}\)

One need not endorse Ratcliffe’s precise picture, nor even care for his terminology, to understand and accept the point at issue here; there are clearly a great amount of psychological tectonics at work in a given agent’s perspective, and hence their thoughts, beliefs, desires and actions. These can include not only ‘existential feelings’ but also facts about how an agent’s beliefs and feelings interrelate, the complicated nature and history of their acquisition, and wider facts about how the relevant agent’s disposition is likely to affect them – and indeed facts about why the agent has the disposition she does!\(^{252}\) So when Jerome Stolnitz complains in the final paragraph of his “On the Cognitive Triviality of Art” that art’s truths are clumsy and ambiguous, that they are better arrived at systematically by other means, he is really quite wrong; it looks for example very much as though the finer appreciation of character, psychology and phenomenology is really best suited to the sort of protracted and indirect treatment they receive in fiction.\(^{253}\) This point even does something to support the cognitive value of some non-fictional artworks; a masterful portrait can, for example, demand and to some extent prompt

\(^{249}\) Martha Nussbaum, *Upheavals of Thought*, p. 65.
\(^{250}\) Jenefer Robinson, *Deeper Than Reason*, p. 128.
\(^{251}\) Matthew Ratcliffe, “There Are No Folk Psychological Narratives”, p. 395.
\(^{252}\) Perhaps ‘existential feeling’ actually means something like ‘disposition’, a less permanent or at least far-reaching or entrenched notion than ‘personality’, but on the way to such.
us to imaginatively fill in the psychological conditions supporting a certain, possibly ambiguous, expression.\textsuperscript{254}

These sorts of insights are also offered and investigated by other prominent philosophers. Martha Nussbaum, again, points out that ‘emotions... have a history... new objects of love and anger and fear bear the traces of earlier objects’; as one’s perspective is shaped by experience, one’s understanding is influenced by one’s past, and a conversation or piece of advice can only be understood narrowly and flatly without some idea of the life from which it is issued.\textsuperscript{255} It might not seem necessary to investigate someone’s personal history to understand why on a given day he takes out his bin or has a bath, because in general we will have enough in common to draw on or assume some set of standard reasons, but this kind of greater understanding becomes far more important when personal or ethical decisions or actions are involved. In short we generally have to try hardest to understand when it is most important not to misunderstand, not just because these things matter more but because they tend to be more influenced by the nuances of personal history. Again, these are the kinds of factors that require a story, and an imagination, to be properly understood.\textsuperscript{256} Jenefer Robinson, meanwhile, talks at length about the special capacity realist literature has for the depiction and elucidation of emotions as shifting, temporally-extended processes – she also stresses how novels do not do this by directly describing a changing set of beliefs on the part of a given character, but by depicting their internal lives in terms, roughly, of how things seem to the character.\textsuperscript{257} Admittedly both Nussbaum and Robinson conduct their discussions with particular focus on the emotions, while we are here concerned with understanding and interpretation in a broader sense, but it is an integral part of the position assumed here that the emotions, their motivational force and other aspects of agents’ psychological landscape not easily forced into the conveniently bare belief/desire mould, are something it is vital to understand if our ethical assessments and reactions are to be appropriate. As Nussbaum remarks, this ‘is not to say that moral philosophy should give emotions a privileged place of trust... [but still] we cannot ignore them, as moral philosophy so often has done’.\textsuperscript{258}

But while these kinds of insights enjoy relatively wide support, they also face some dissenters. Stolnitz’s attack on cognitivism has already been alluded to in passing, but it is also worth mentioning Lamarque and Olsen’s work in \textit{Truth, Literature and Fiction}, as it presents perhaps the most sustained critique of what we might call the cognitivist project. They are keen to point out that

\textsuperscript{254} Fruitful discussion of this sort of case can be found in Berys Gaut, \textit{Art, Emotion and Ethics}. Gaut also gives good reason elsewhere for thinking that interactive fictions can be a special source of a wide array of self-directed emotions from which, presumably, we could learn – see \textit{A Philosophy of Cinematic Art}, pp. 272-281.

\textsuperscript{255} Martha Nussbaum, \textit{Upheavals of Thought}, p. 175.

\textsuperscript{256} \textit{Ibid.}, p. 179.

\textsuperscript{257} See for example Jenefer Robinson, \textit{Deeper Than Reason}, p. 161, 165. This is in contrast to Lamarque and Olsen’s suggestion that in watching an Ibsen play we do not understand what it is like for the characters so much as we simply observe their behaviour – even though the studied observation of behaviour is obviously a crucial component in coming to know what it is like for someone. Lamarque and Olsen, \textit{Truth, Fiction and Literature}, p. 377.

\textsuperscript{258} Martha Nussbaum, \textit{Upheavals of Thought}, p. 2.
literature can be a valuable source of learning, but at the same time maintain that it would be inaccurate to describe literature as a source of knowledge or to suggest that the things literature supposedly teaches should be considered part of its worth qua literature as an art form. They suggest two formulations for the position that knowledge can be acquired from literature, a weak version and a strong version of what they term the Subjective Knowledge Theory of Cognitive Value. The former is that literature provides us with knowledge of ‘what it is like’. The latter is that literature does not provide one with specifically new knowledge, so much as it allows one to see what one already knows more clearly or in a different light. Neither, they suggest, passes muster. They hold that in order to be adequate, the SKT must meet three criteria: it must explain how the subjective experience provided by literature can constitute knowledge, it must explain the difference between apparent and genuine knowledge – distinguishing ‘between true and false’ – and it must ‘establish that literature in virtue of being literature promotes imaginative participation of the type that yields this kind of knowledge’.259

It bears remarking that this does not really threaten any of the claims in this thesis, since the claims made here are generally cast in terms of fiction’s helping us to better understand one another – these need not amount to claims about knowledge per se. Any claim that fiction can help us to know what it is like for others need not be taken literally; given the essentially private nature of subjective experience, and the consequent impossibility of confirming one’s impression with total surety, it might be best to take claims about knowing ‘what it is like’ as shorthand for some more relaxed state of warranted belief. So the position taken here is basically compatible with that of Lamarque and Olsen, at least in this respect. However, their complaints and stipulations also do not need to be taken as fatally damaging to something like SKT. Let us focus on what they regard to be the stronger formulation of SKT, that literature can illuminate what one already knows in some way. They admit it meets the first criterion of explaining how the subjective experience provided by literature can constitute knowledge, and that it helps in meeting the third by showing how literature does this in a literary fashion, but suggest that it fails on three further counts. Firstly, they say that it cannot distinguish between genuine and putative knowledge (which was their second criterion for SKT generally). Secondly, they say that while this highlights the educative power of literature qua literature, the specifically literary element to such learning does not contribute much, and that the learning thereby done in a given case cannot only be available through one specific piece of literature. Then, thirdly and finally, they say that the strong SKT ‘fails to make a distinction... between recognising a conception of a situation and adopting that conception’.260

259 Here ‘strong and weak’ is meant in the sense of ‘better and worse’, rather than the typical philosophical sense of ‘bold and modest’.
260 Lamarque and Olsen, Truth, Fiction and Literature, p. 371.
261 Ibid., pp. 380-4.
With regard to their first problem with strong SKT, it seems obvious that one can to some extent distinguish between genuine and putative cases of knowledge imparted by literature. Lamarque and Olsen, following Thomas Nagel, describe literature as being in this regard either ‘authentic’ or ‘inauthentic’, meaning roughly either inspired and moving or the opposite. Naturally a reader will have a sense of whether a piece is inspired or banal, but, they complain, either kind can just as easily lead one to see things in a new light.\(^{262}\) It is not entirely clear why exactly this should count as a problem, as this appears to translate fairly comfortably into talk about learning something from a book as opposed to learning something despite or due to it.\(^ {263}\) Aside from this, the reader would ideally know from her own reactions, and from conscious comparison against her own experience, whether the insight and depiction in a novel seems true to life, plausible or genuine. Granted, there does seem to remain some spectre of uncertainty, but this remains the case in all cognitive endeavours, whether one is studying Shakespeare or quantum mechanics.

Their second problem is no less puzzling, as the specification that literature yield a specifically literary kind of knowledge, or that it does so in a specifically literary way, seems totally unnecessary. They also complain that insofar as what is learned can be understood independently of the literature from which one learned it, which they say follows from the fact that it can be applied to the world, it ‘cannot only be available through one specific literary work’ – again, it seems all but mysterious as to why this should be a genuine cause of concern.\(^ {264}\) The reason for this complaint of theirs comes down, presumably, to their very specific interest in literature as a normative, institutionally defined category necessarily superior to simple ‘books’. If this class of works is so special, the thinking seems to go, any explanation of what makes it good must refer to qualities that set it apart from similar things which are not fit for entry into the set. Our focus here, of course, is on fiction more generally, and so the need for unique value seems far less pressing, but even so it simply seems false to suggest that literature’s virtues must only pertain to what makes them literature, or indeed to claim that what makes War and Peacevaluably educative must be totally unique and unrepeatable in any other work. Similarly, it seems uncalled for to demand that a major virtue of literature must be found in all examples of literary works. Their own reading of affairs, meanwhile, seems questionable; a piece of literature, they propose, should be assessed in terms of the perennial literary themes it evokes, where ‘a theme is not the kind of entity that can be true or false. Rather it is interesting or uninteresting’.\(^ {265}\) Literature and literary themes, then, seem to only be worthwhile on a sort of simplified, semi-Humean condition that they are simply considered such over time or by an extremely widespread audience. This is questionable because it obviously begs the question of what

\(^{262}\) Ibid., p. 380.
\(^{263}\) See also Berys Gaut’s example of the incompetent DIY plumber, who cannot teach but from whom we can learn. Art, Emotion and Ethics, p. 140
\(^{264}\) Ibid., p. 381.
\(^{265}\) Ibid., p. 437.
should make a given theme interesting or durable – a question Lamarque and Olsen refuse to answer.\textsuperscript{266} Surely the most obvious reason is that certain themes seem true to life, informative, insightful, or enlightening! In short, the value of literature is parasitic upon other, non-literary human concerns, and can be sensibly assessed in terms of how well it lives up to them. To try and define literature in a way which downplays this is undoubtedly to impoverish it.

Their third complaint, that SKT does not distinguish between recognising a conception and adopting it, seems equally stilted. Surely, to the extent that one can come to form an impression of a different subjective position on something, and have some awareness of how it relates to or could affect one’s own, this involves a working sense of the difference between recognising a conception and adopting it. But they go on to suggest that what is really at issue here is the fact that in coming to appreciate, for instance, Ibsen’s view of art, one is not thereby required to adopt it – the appreciation of it has no necessary implications for one’s adoption of it. In other words, they accuse the strong SKT of actively running together the recognising of a conception and the adoption of it, rather than simply failing to provide a means of distinguishing between the two. However, as just contended, SKT can distinguish between recognition and adoption, and in light of the fact that we are talking about, roughly, ways of looking at things, it does not seem odd or problematic that the former need not lead to the latter. They then suggest that this problem boils down to literature no longer offering an avenue to the acquisition of subjective knowledge parallel to philosophy, but rather one subordinate to it.\textsuperscript{267} Again, this simply does not seem like a catastrophic problem, even if it were true. As it happens, the above discussion of Robinson’s work in Deeper Than Reason should indicate that literature and fiction retain a very special ability to provide us with certain subjective experiences and insights, and thereby a special means of conveying and/or clarifying knowledge. Indeed, this also answers their complaint that the literary element of novels does not do much work in imparting knowledge; clearly, more sophisticated novels are better at more engagingly showing us more complex and sophisticated things – things that, while perhaps not impossible to learn otherwise, are perhaps best learned from (great) literature. But suppose this was not the case - it seems literally puritanical to insist that nothing should be able to do the or a job of literature better than literature. So it appears that Lamarque and Olsen’s attack on SKT is far from watertight, and thus that they doubly fail to pose a threat to the position detailed in this thesis.

What is important here, at any rate, is the idea that literature, and fiction more generally, along with the sustained and nuanced imaginative activity that it incurs, can at least indirectly improve one’s ability to understand others, regardless of whether or not this can appropriately be described as something involving knowledge as such. This is not to say that engaging with fiction is a

\textsuperscript{266} Ibid., p. 456. This also harks back to our dismissal of Mill and Singer’s move regarding the worth of different preferences.

\textsuperscript{267} Ibid., pp. 184-5. Note this seems to assume that philosophers never see positions they do not themselves accept as plausible.
sure-fire means of improving one’s capacity for understanding, or one’s character more generally. Many have pointed out what was mentioned in the first chapter, regarding identification, that fiction can also function to engrain one’s prejudices more deeply, or otherwise ‘narrow our horizons... [and] encourage the superficial stereotyping of persons’. 268 Others have expressed ambivalence toward the idea that empathising with fictional characters entails definite implications for our tendency to empathise with people in real life. 269 Robinson suggests that those who feel for fictional characters but not real people are ‘manipulated by the author to be (albeit briefly) better people that they really are’ – but this verdict does not seem certain. 270 Inasmuch as there seems to be something particularly reprehensible about an agent who weeps over some tragic fiction before and after turning a deaf ear to the pleas of starving orphans on the street outside, the sentimentalist’s sorrow seems utterly shallow and devoid of moral worth – the author of the given fiction might in such a case simply be providing her with a means of betraying how lost she is to genuine moral concern. 271 But, to reiterate, all this does not go against our position here, since I only make the weak claim that literature is a form of practice, that it can exercise and improve the kinds of imaginative capacities that are typically of great importance to interpersonal understanding. 272 Whether these capacities do in the event get exercised toward the better understanding of real people, or even if such understanding is always or merely typically conducive to the better treatment of others is, sadly, not something we are in a position to guarantee. 273

But the proposition that fiction presents a special and valuable source of imaginative and emotional training faces other limitations, some of which are potentially quite troubling. Nussbaum, for instance, points out that ‘some forms of life are... unavailable to us... medieval courtly love is not a live option in the present day, since we cannot share the metaphysical beliefs and practices necessary to sustain it... we can imagine ourselves in that world only on a very general and partial level, focusing on the ideas... that are still available to us’. 274 This particular worry, at least, is most likely a relatively minor one; the extent to which we can imagine ourselves, in a thick sense, in a very different time and society is not of great importance. What matters rather is our (admittedly imperfect)

268 Carl Plantinga, Moving Viewers, pp. 141-2.
269 Susan Feagin, Reading with Feeling, p. 100. Here Feagin talks about ‘the selfish sentimentalist’ who is extremely moved by fiction but who is entirely unmoved by the plight of people in real life.
270 Jenefer Robinson, Deeper Than Reason, p. 141. Robinson’s suggestion here does not follow from Feagin’s depiction of the selfish sentimentalist, but then she does not make it clear if this is supposed to be an elaboration or rather a contribution of her own.
271 Indeed, if one accepts something like the concept of fictive externalism which I argue for in the next chapter then it would appear to make little or no sense whatsoever to describe the sentimentalist here as having moral feelings.
272 This is the kind of exercise or practice alluded to in the first chapter, in response to worries about understanding’s demandingness and the efficacy of folk psychology.
273 Here I mean treatment in a sense beyond how one assesses others, naturally, since understanding people better does, in my view, constitute better treatment of a certain kind, even if it leads one to treat many people worse in a trivial, non-moral sense, e.g. by offering more warranted criticism and less unwarranted praise.
274 Martha Nussbaum, Upheavals of Thought, p. 170, emphasis in original.
capacity to imagine being someone in a very different time and society, and for this reason our greater capacity to entertain than to temporarily adopt foreign beliefs might be enough to get us by. What is of greater concern is a problem which lurks in the background of Robinson’s *Deeper Than Reason*, occasionally acknowledged but nonetheless downplayed – we might call this problem the cognitivist gap. Robinson mentions a more harmless-seeming manifestation of this issue when she talks about how, for example, a baseball-related passage in a book will be better attended to and better remembered by those who know more about baseball. The problem begins to seem much more acute, however, when she talks about how ‘a precocious little girl reading ‘The Ambassadors’” or ‘teenagers and young twenty-somethings who ‘hook up’ for the night at parties’ could not understand the sorts of emotionally educative literature she discusses because they describe ‘experiences outside [their] emotional range’. The problem, then, is that it appears fiction can only educate people about things they essentially already know, as indeed the strong SKT suggests, and further that it can only work at all from a substantial base of assumed knowledge, so fiction’s capacity to teach seems severely handicapped – in other words, if one has no real experience of something, fiction is not adequate to bring ‘mastery’ within one’s reach.

But how troubled should we be by this, exactly? It is probably far more of a problem for Robinson specifically than for aesthetic cognitivists generally, since Robinson’s focus is extremely specific and, consequently, her standards extremely high – she says herself that she is concentrating on ‘the great realist novels’ and, it follows from her later qualifications, those with an emotional repertoire sufficient to read them in a state of relatively comfortable recognition. But while it may or may not be the case that the kind of apparently profound sentimental education she discusses can only be had from great realist literature, it seems as though more garden variety learning can be had from more garden variety literature; certainly it is Hutto’s claim that even children’s stories lay the groundwork for a respectable capacity for folk psychological practice. This might suggest a deflationary analogy to us: if we cannot properly understand the sentimental masterclass supposedly offered by great realist literature without already possessing a wealth of emotional knowledge or experience, perhaps this is no different than in other subjects – no-one would expect to understand an advanced lecture on theoretical physics without already having a good knowledge of ‘the basics’. But it is not entirely clear that this analogy works, for two reasons. First, it appears Robinson seems to be talking, roughly, about experience of $x$ (or broadly $x$-like experiences) being necessary for a literary understanding of $x$ – the analogy just given describes, roughly, experience of $y$ being necessary for an understanding of something like $z$ or $Z(y)$. That is, if a novel describes some emotional journey or process, one can check it against one’s own roughly similar experience to see if it is accurate or illuminating. The advanced lecture on physics, however, relies on a knowledge of *related but quite*

different facts to allow for the processing of largely new information. The second reason is close to the first: a thorough knowledge of the basic tenets of physics might in themselves be enough for one to extrapolate more complicated information about the workings of the physical world, whereas it is not obvious that a thorough knowledge of young adult literature is sufficient to gain more nuanced emotional and psychological knowledge – or to put it in slightly different terms, it seems as though repeated close readings of Harry Potter will not make one any better at understanding Anna Karenina.

But the philosophical discussion must break down here somewhat, insofar as these questions seem to demand thorough empirical investigation if they are to be properly answered. Even so, the cognitivist need not despair, since it is hardly a conceptual requirement of learning that practical experience of the matters concerned not be necessary. If learning from literature requires experience of only broadly similar matters, or indeed even quite similar matters, it is still perfectly possible that literature could educate by means of helping one to understand or adopt a fresh perspective on what one has experienced. Perhaps this is not a proud admission if one takes a primary aim of education to be preparation – if one particularly cherishes the notion that to be forewarned is to be forearmed, so to speak. But there is nothing new or shameful in the assertion that some subjects require a special kind of acquaintance. Understanding other people becomes harder given greater differences, naturally, so really it is a fairly pedestrian observation that to understand fictional characters better we need to have had experiences at least broadly in line with their own.

What about the claim that other fiction can be good for other kinds or degrees of learning? Robinson does not really justify the oppressive narrowness of her investigation, dismissing ‘bad novels that try to teach us something and fail... novels that merely aim to entertain... [and] novels that are more like intellectual puzzles or games’. Her neglect of the third category mentioned here seems a particular unkindness from a cognitivist perspective, since it sounds like an excellent candidate for learning – or at least for the more general kind of cognitive exercise emphasised here – but then Robinson is specifically interested in the emotions, so we can excuse this in the context of her own project. At the same time, from a broader perspective, we should acknowledge that the sorts of things said above about how emotional development cannot be properly condensed into a list of propositions could apply, mutatis mutandis, to the development of certain non- or semi-imaginative cognitive skills. For instance, one might repeatedly be told about the basic tenets of logic without anything really ‘sinking in’ until one has had opportunity to put them into practice. It is not entirely clear what a novel resembling an intellectual puzzle would look like – there are children’s ‘puzzle adventure’ books which include puzzles one must solve as one reads through the story, or other examples further afield include things like ‘lateral thinking’-heavy adventure games such as Monkey

277 Ibid., p. 159.
Island, or storied videogames that feature more straightforward puzzles such as the Professor Layton or Puzzle Agent series – but it seems probable that it would or could afford such practice.

What about novels that fail in their attempt to teach? Presumably what Robinson means here are novels which fail due to some fault on their part, as opposed to those which fail due to some fault on the part of their audience.²⁷⁸ Such works have been dismissed elsewhere on cognitivist grounds, as when Berys Gaut uses the analogy of learning a valuable lesson on how not to fix one’s pipes from an incompetent do-it-yourself enthusiast; it is not appropriate to ascribe cognitive value to such works on the grounds that ‘there is not necessarily a [corresponding] cognitive or epistemic merit in the object or person from which one is learning’.²⁷⁹ But while it is thus clear that we should not praise a work, its creator or creators for teaching us something in such an instance, the fact that we can learn from it at all gives the understanding agent, as well as the cognitivist, at least a pro tanto reason to engage with it. It is important to bear this pro tanto qualification in mind, of course, or one would be left suggesting that an understanding agent should attempt all kinds of morally questionable or dangerous activities for the sake of the learning that could thereby be done – indeed, if a serial killer demanded before being judged that we ‘walk a mile in her shoes’, we would not want to occupy a position that leaves us feeling tempted or obliged to take her up on the request.

So what, then, about works which ‘merely aim to entertain’, as Robinson so loftily puts it. Satire presents a good counterexample to this sort of assumption, however, since the purpose of satire is, essentially, to make a point in an entertaining fashion. But perhaps Robinson means to dismiss not works that mainly seek to entertain, but works which seek only to entertain, to the exclusion of anything else. Here we should recall our previous considerations regarding soap operas; it seems as though the merely entertaining might still afford ample opportunity for the educative exercise of understanding. While it might seem true from experience that works made with the primary or sole aim of entertainment tend to have less cognitive merit to them, then, this should not lead us to believe that entertaining works must always be shallow or devoid of cognitive worth. Something understanding also reminds us to bear in mind here is that different people have different capacities, attention spans, preferences and so on, and so different kinds of fiction, though they might deserve to be considered excellent in their own right, might simply be unable to teach anything to large swathes of the population. And, of course, it would not be fair to suggest that it never takes any skill or artistry to make an entertaining or exciting film. So while we might think that ‘objectively’ some genres of fiction are more ennobling or worthwhile than others, it should be remembered that a work can be good not just for what it teaches, but for who it is able to teach.

²⁷⁸ One wonders if this is always a clear-cut distinction; perhaps authors who intend to teach have some obligation to make sure their lessons are approachable for their intended audience. So a great realist novel, for instance, might be a failure while containing no internal fault, by detailing an emotional development so rarefied that no-one could understand it.

²⁷⁹ Berys Gaut, Art, Emotion and Ethics, p. 140. Gaut’s position is very much cast in terms of pro tanto reasons, it should be said, and so he leaves it open for the cognitivist to praise a non-educative work on other grounds.
4.2 Fictive Distortion and Imagined Realism:

We can take it as established, then, that fiction presents a special opportunity for a particular kind of imaginative and emotional learning and/or exercise, even if it does not do so in an entirely unproblematic fashion. Before we tackle one of cognitivism’s potentially awkward complications – what I call the problem of imagined realism – in this section, however, some argument needs to be made concerning the manner in which cognitivism should be cashed out, in particular with a view to what sort of standards it should require. In recent times Gregory Currie, formerly something of an arch-cognitivist, seems to have become more and more sceptical about the possibility of our learning, in any significant sense, from fiction. In his 2010 Narratives and Narrators he voices concern about the situationist challenge to character, that external factors might be a better guide to behaviour than notions of fixed personality, and how this might affect the worth of fiction should character be a crucial notion to it. He does not seem overly pessimistic, though, concluding that ‘we may think of Character, not so much as a psychological-explanatory tool, but as a device for making vivid and coherent the interplay of... other, psychologically real factors’.\(^{280}\) At the same time he also acknowledges that not all forms of literature need emphasise character, and so that the threat posed to aesthetic cognitivism by situationism need not be disastrous in any case. Despite this, however, in a piece for the Times Literary Supplement in August 2011, Currie’s doubts about the possibility of our learning in any significant sense from fiction seem to have returned, though they are formulated in a relatively loose fashion. I believe that Currie’s intermittent lack of confidence in cognitivism’s prospects results in part from his not properly formulating the sort of standards the cognitivist should have in mind – more specifically, from his misplaced emphasis on realism. Before discussing the problem of imagined realism, therefore, the cognitivist’s need for realism as such must be addressed.

To begin with, take Currie’s 1997 “The Paradox of Caring”, where Currie offers us what he suggests is a sub-division of the paradox of fiction, ‘the problem of personality’. He describes this problem by saying ‘we frequently like and take the part of people in fiction whom we would not... in real life’.\(^{281}\) His answer is to suggest that, in coming to like a fictional character one would not like were they real, we judge them according to the standards of a simulated personality the fiction induces us to assume. However, while Currie’s story looks plausible in the limited case he applies it to (C.P. Snow’s The Masters), we should not thereby assume that he has solved the real issue. For instance, on watching The Silence of the Lambs I found the only particularly interesting or likable character was Hannibal Lecter. While I would not like him were I to actually meet him, I do not think it is fair to thereby assume the film manipulated me into pretending to be someone who doesn’t think cannibalism is particularly serious as a character-flaw. Certainly the film doesn’t manifest the attitude that one should like him. The problem of personality is perhaps better put thus:

\(^{280}\) Narratives and Narrators, p. 212.

[PP]: we frequently respond to fictional characters differently from how we would if they were real people.

To a simulation theorist, this calls for explanation. If the psychological processes involved in reacting to a story about real people are the same, or largely the same, as those involved in reacting to a story about fictional people, why would there be a marked difference in our reactions? The simulation theorists’ first response, naturally, is to point out that ST has never made claim to describing the exclusive means by which we understand one another, and so it could be that something external to my Lecter simulation was at work. Even then, it’s not obvious that a simulation cannot have as part the idea that it is being done for something that is real or not real – there are no doubt an overwhelming number of factors which influence the extent to which we put effort into or take seriously the simulating of others’ perspectives, consciously or otherwise. But we don’t just react differently to fictional characters, we also react differently to fictional states of affairs. Imagine the entirely probable case of someone enjoying the bleakly dystopian vistas in the film Blade Runner. They see a dark, futuristic hellscape of fiery towers, filthy streets and desperate beggars, where there is no joy and non-human animals have all but died out. And what do they think when confronted with such a spectacle? “Cool!” This seems, I think to most of us, a harmless response. But were we observing these scenes in actuality, we would have serious questions to ask about this person’s sanity and motivation. So is the name ‘the problem of personality’ then a little misleading? Currie seems to name it thus because fiction can induce one to have emotional reactions contrary to what one would normally consider appropriate to one’s personality. But, as already mentioned, Currie himself has since grown very sceptical about the notion of personality or ‘Character’. So let us change PP, then, to the problem of fictive distortion:

[FD]: our fiction-responses often differ from our nonfiction-responses.

Should we be particularly interested in FD? It is, after all, a very broad and basic psychological claim. Nonetheless, FD raises concerns at least for Currie’s position of a year later, in his ‘Realism of Character and the Value of Fiction’. Here he claims that for it to be true that we learn from fiction, fictional worlds would have to be ‘systematically related to the real world’. This might not have to be considered overly restrictive, depending on how one construes the notion of systematic relation. Currie himself defines his realism here as the idea that fictions can teach us by virtue of their ability to provoke responses in us that are ‘similar’ to those we would have in responding to real life. Let us call this strict realism. What consequences does FD have for such a position? Well, while it is true that there are differences in our fiction-responses and non-fiction responses, clearly they also seem to
overlap much of the time. So perhaps we should reformulate the realist position Currie describes as a kind of pro tanto realism, thus:

[PTR]: we can only learn from a fiction to the extent that it, or the parts of it relevant to such learning, call on realistic reactions.

However, it is unclear if even this more moderate form of realism should be considered convincing. One could easily suggest that we can learn from fictions, and the aspects of fictions, which elicit unrealistic responses. After all, we can learn about the disturbing nature of jingoism watching a mindless Hollywood action movie that invites us to exult in the unreflective killing of Soviets or Arabs. One might no less think that we can learn from, indeed be taught by, satirical comedies that depict tragic or violent scenes amusingly, as arguably exemplified in La Vita è Bella, for instance, or even Team America. These are certainly cases where we are invited to react to fictional scenes in a manner different from how we would if they were really happening. Hence we can see that PTR, at least, is inadequate. But it is also very difficult to tell whether or not these cases still bear some ‘systematic relation’ to the real world. So it seems that Currie’s realism here is either implausible or frustratingly diffuse.

It is possible to take what we might call an anti-realist position, however, which explicitly rejects the claim that learning from fiction requires any kind of realism. For instance, Berys Gaut uses Ursula LeGuin’s The Left Hand of Darkness as an example to make the claim that we stand to learn, at the very least by means of contrast, from characters with an entirely different psychology to our own. 282 In addition to this, of course, it is often most important and most difficult to try to understand those who are very different from us; so stories which happen in unusual places to unusual characters can have special importance – variation and heightened difference could be useful means to taxing our imagination and expanding the horizons of what thought and behaviour we believe to be possible or intelligible. Indeed, if most normal social interaction happens according to assumed and unspoken social conventions, realist stories might often have a kind of weakness in that they do not tax a wide set of our imaginative capabilities. 283 This difference presents us with an advantage in fiction, whereas it would be something of an unhappy difficulty in life, because as Currie himself has pointed out we have the time and energy to freely try to understand fictional characters and states of affairs, potentially drawing on the aid of sensitive and well-crafted prose to help us to carry through imaginative projects that might be too protracted or difficult on our own. 284

283 At the same time it is these days true that the great realist novels tend to at least be set in a different century, so it would be misleading to suggest that they operate according to exactly the same set of social expectations as we do.
Such an anti-realist position is not entirely free of its own problems, however. A total disregard for realism in fiction on the part of the cognitivist does leave open the charge of triviality; it seems we could potentially learn about people and life from anything on such a reading of affairs. The anti-realist’s response here, I believe, should simply be to bite the bullet and point out that we can, in fact, learn from more or less anything. The limitations on what we can learn from which things are primarily set by the imaginative limitations of the person attempting to learn; it is only in a very loose sense that the properties of an object constrain what sorts of things could be learned from it. But more than learning, the cognitivist is interested in whether art works can teach us. And given that we can in principle learn from almost anything, the question of whether art can teach should not look at any given work’s subject matter, or the extent to which it depicts a realistic world or calls on realistic emotional responses. The question of whether art can teach depends on whether artists have any capacity to knowingly and deliberately manipulate their media for predicted effects, and if these effects can be conducive to worthwhile or edifying cognitive activity. The answer to both questions, transparently, is yes. Discussion of realism need not enter the picture here.

But while in principle we can see that, regardless of factual accuracy or general realism, fiction provides us with the material for engaging in imagining which we can test against our own knowledge and experience, if we are too credulous it seems as though fiction also presents a great danger of leaving us with false beliefs. This is the problem of imagined realism. Many fictions, perhaps most notably horror fictions, actually misdirect our emotions, teaching us to feel certain things toward certain objects in a manner we might well think of as inappropriate. This is not exclusive to the genre of horror, naturally; certain ‘aspirational’ teen dramas such as The OC or Beverly Hills 90210, for instance, could easily be accused of glorifying a certain degree of excessive self-obsession and materialism. In the previous section it was also suggested, contra Robinson, that ‘novels... more like intellectual puzzles or games’ could be a valuable source of intellectual exercise or learning. But this of course need not always be the case; an avid reader of Agatha Christie novels, for instance, might develop the belief that ‘in cases of murder, the culprit is always the one you would least suspect’. The most straightforward and plausible example of this sort of problem, though, is probably the horror film Jaws, which notoriously left an untold number of people with an irrational – or at least misled and inaccurate – belief that sharks are a great deal more malevolent and dangerous than they are in reality.

It is a sometimes stated ambition of horror directors and writers that they wish to make some mundane real life creature or situation seem so terrifying or bizarre that people never look at them or it in the same way again, and insofar as this aim is sometimes realised or is generally realisable it gives the cognitivist a good reason to be wary of the genre. But the understanding agent must surely try to understand and appreciate this problem, bringing with her to a fiction a sufficient degree of

\[285\] Ibid., p. 159.
knowledge to have some sense of what is exaggeration and what is not, or where this is not possible to
either make efforts to enquire as to whether she is engaging with a relatively realistic fiction or to
engage with the fiction in a state of epistemic scepticism (such that she does not hold it to be reliably
informative). It is only in failing all of these conditions that she should consider refusing to engage
with the fiction concerned. How pressing this matter is depends largely on the extent to which
practical consequences are involved, or potentially involved; it might well affect one’s life to think of
sharks as murderous, malicious killers who could turn up anywhere, and so one should take care to
view Jaws in the right spirit, but by contrast there is little pressure on most of us to know what the
lives of spies are really like, and so correspondingly there is less pressure on us to know just which
elements of an Ian Fleming novel are particularly implausible.\textsuperscript{286} As it stands, there is a pronounced
tendency in the horror genre, and among its aficionados, to view proceedings with a wry sense of
irony. Sam Raimi’s The Evil Dead series and Peter Jackson’s Bad Taste and Braindead are very good
examples of this, as indeed is the use of comedic musician Richard Cheese’s cover of Get Down With
the Sickness in the credits of the 2004 remake of Dawn of the Dead. So the invitation to assume a
protective attitude of epistemic scepticism like the one mentioned above is nowadays a staple of the
genre.

So understanding need not view horror fictions as actively pernicious, but the question
remains whether the understanding agent has any reason to actively engage with them. What could
there be to learn from horror films such as Jaws or The Exorcist? Surely not that the oceans must be
cleansed of the impending shark menace, or that we should live in fear of demons possessing our
children. And these are two examples of respected horror films. So most, if not all, horror fiction
seems prima facie to be delinquent in terms of life lessons – in fact it is somewhat interesting to note that
what seems to be the central factor in a fiction’s qualifying as a thriller rather than a horror is either
how cerebral or how realistic it is.\textsuperscript{287} So it appears that the more there is to learn from a film, the
further that film departs from the genre of horror. But we should also remember here Gaut’s claim
that we can still learn from the decidedly anti-realistic by means of contrast – although it is perhaps
worth noting that this might require a greater level of knowledge on the part of the understanding
agent, at least inasmuch as such learning demands more than the otherwise minimal requirement of
bland acceptance.

We should not forget, too, that the things we learn from fiction need not relate exactly to what
overtly occurs in the story. Night of the Living Dead, for instance, need not merely be seen as a story

\textsuperscript{286} As for the case of our Agatha Christie fan, it bears remarking that a rule such as ‘it’s always the one you least
suspect’ is in fact a good rule for understanding detective fictions. The error would lie in mistaking this literary
observation for a real-world psychological one. But again, this sort of apparent problem is less relevant if the
cognitivist emphasises beneficial cognitive exercise over learning per se, since this discourages the notion that
specific real-world maxims are or should always be constructed directly from fictive experience.

\textsuperscript{287} One notable exception to this being The Blair Witch Project, which was intended to be frightening primarily
because of its supposed realism or rather, to use a term I define in the next chapter, its ‘fictive transparency’.
about what it might be like defending a house from an encroaching zombie horde, or more generally about what might happen if a zombie horde really existed. Rather, it is probably most rewardingly viewed as a story about group dynamics under tense conditions. On a crude level, one might learn that groups should not split up when in danger, or on a cruder level still one might learn something about how to fortify a house against attack. But, what is more significant in emotional terms, one can also learn from it something about how people interact when afraid. If there is one thing horror fictions universally describe, it is people’s reactions to, expressions of and attempts to deal with fear. So there is a case for horror fictions teaching us something about the nature of our emotions. And such teaching carries with it a variety of less general lessons – Melanie Daniels teaches us in *The Birds* that it is a bad idea to investigate strange sounds on one’s own when there is a possibility of danger, Guy Woodhouse teaches us in *Rosemary’s Baby* that we should not let ambition obscure the importance of our loved ones, Mrs. Lee teaches us in *Dumplings* that a spouse who marries one solely for one’s looks will likely lose interest as one ages, and so on. Simply because the language in such fictions is hyperbolic, metaphorically or literally speaking, does not mean there is no message at all to be gleaned from them. Indeed there is some further opportunity for an increase in self-knowledge; by comparing and reflecting upon the different horror fictions with which one has engaged, one can gain some insight into what one’s inner fears might be, among other things by discovering which fictional situations one finds more frightening than others.288

Of course, even if one finds these arguments persuasive, and thus accepts that the understanding agent has *some* interest in consuming even horror fictions, this leaves open the objection that such an agent might still be almost invariably better off engaging with other kinds of fiction. After all, lessons about fear and so on are probably also to be found in more serious-minded fictions. But this is not a terribly troubling realisation; the more intuitively appealing position would seem to be that there is *some* good in horror and other less respected genres of fiction, rather than that they have a special and strong value which would leave us open to criticism if we did not engage with them. Considerations regarding personal preference and ability are relevant here, too – the kinds of lesson one values, as well as the degree of complexity or sophistication with which one is comfortable, are going to be a matter of individual taste and ability. Even then, those who usually prefer more highbrow, cerebral forms of entertainment might occasionally desire a story which appeals to simpler needs or fears, or that does so in a more straightforward or ‘entertaining’ fashion. Indeed, it is something of a fallacy to generally suppose that one thing must be made useless by virtue of there being another thing better in relevant respects, at least in the absence of some counter-intuitively maximising principle about how one should spend one’s time or make one’s choices.

There is another, largely empirical, worry that is relevant here, which Nussbaum alludes to, as in a way do Gregory Currie and Ian Ravenscroft. In considering tragedy, Nussbaum mentions that ‘the experience of tragic contemplation can even involve an aestheticizing of the person’s plight that has a most unwholesome moral character’. Currie and Ravenscroft talk of how ‘events, episodes, and even lives can be narrativised; the events are lived as narrative... [this] making a contribution to the motivation of the people who take part in them... [and if an agent’s] desire to see the narrative played out is strong, he may pursue a wicked course’. The general worry here is that fiction and fictive imagining can encourage morally inappropriate attitudes and behaviour – in the last two decades this sort of concern has in the media taken for its primary focus the playing of videogames. There may or may not be something to this worry, and as with the imagined realism problem much might come down to the responsibilities of the ‘consumer’ to keep track of how they believe themselves to be influenced by a given fiction or fictions. On the face of it, at least, it seems as though these kinds of gross distortions of one’s values and thinking are more likely to be a live concern for the already mentally ill. Having said this, while obviously reading fantasy novels does not make teenagers believe in dragons, a lifetime of exposure to fiction which depicts animals as unimportant, or women as being weak and dependant, could easily be seen to contribute to one’s general attitudes and behaviour. Indeed if most fiction supports the status quo it could work as a troubling force against what we might cautiously think of as the ‘moral advancement’ of society. But as with the question of what fiction or art’s exact power to educate is, to investigate these matters fully would require extensive and thorough empirical research of the kind that is not within the purview of philosophy. However, we can appropriately consider questions about the moral implications of certain general kinds of fictive imagining and activity – and in the next chapter we shall use understanding to explore what some of these implications might be.

4.3 Closing Remarks:
We have by now established the prevalence and centrality of understanding to ethical assessment and deliberation. In the present chapter we have now begun to move toward a picture of understanding’s usefulness to the philosophy of art – having also provided some discussion of thick understanding’s relevance to contemporary philosophy of art in the first chapter, by looking at some of the debate surrounding the concept of identification. Understanding obviously recommends practices, particularly low-cost or enjoyable ones, which can improve our ability to understand others, perhaps most especially those who are different from ourselves. To that extent this chapter’s discussion of aesthetic cognitivism was an obvious first step in establishing understanding’s strong connection to philosophy of art generally – in the first section we established the possibility of cognitivism, while in

290 Gregory Currie and Ian Ravenscroft, *Recreative Minds*, p. 204.
the second we did some work to make cognitivism more plausible and to clean up some of the complications to which it might be seen to give rise. In the next chapter we shall go on to see what work understanding can itself do within the philosophy of art, by looking at what it has to say about interpreting works and the characters they contain; what normative constraints we face and what it is that we should think of ourselves as doing, with a view to detailing what some of the central differences between understanding real and fictional people might be. This will bring us to a new ethical principle about imagining that I call fictive externalism, as well as justifying an inclusive and pluralistic perspective on the interpretation of works.
CHAPTER FIVE

5. THE ETHICS OF IMAGINING

5.1 Understanding and Fictive Externalism:

The previous chapter established that we can view fictions as a valuable means to learning, and clearly this gives the understanding agent a basic reason to be interested in engaging with them. We shall go further in this chapter, however, by showing more explicitly how understanding itself can be exercised and improved by engaging with fiction. But to begin with we shall first examine three fundamental differences between understanding as it is applied to real life and understanding as it is applied to fiction, differences which bring out an often overlooked ethical principle I call fictive externalism. We have already discussed the first such difference at some length in another form during the previous chapter; it is the fact of fictive distortion, that something’s being fictional often seems to itself make a large difference in our emotional responses to it. To illustrate fictive distortion as it relates to understanding fictional characters we need simply to refer back to our previous two examples from the first and fourth chapters. These were of course the charismatic Dr. Hannibal Lecter, who would have far fewer fans were he a real person, and Dr. Gaius Baltar, whose otherwise endearing cowardice would in real life make us breathless with indignant rage. It is worth mentioning here that Matthew Kieran provides a finer distinction within the phenomenon of fictive distortion, though, highlighting two asymmetries between our real life and fictive responses: asymmetries in valence and in response. That is, in engaging with fiction we sometimes have fundamentally ‘normal’ reactions, fearing the fearful for instance, where these reactions have an atypical valence – as when we enjoy being scared by horror movies. In other instances the asymmetry lies in our having an emotional reaction which is itself fundamentally different to what we would have if witnessing a given fictional event in actuality – it is this latter asymmetry which our examples of Lecter and Baltar are particularly meant to demonstrate, though arguably they could represent either or both.

In any case, it is worth noting in light of the arguments to come that in engaging with the fictional worlds of *The Silence of the Lambs* and *Battlestar Galactica* we are not meant to undergo a fundamental shift in relevant behavioural standards. In either fictional world murder, cowardice and betrayal are supposed to be just as bad as they are in our own, and people are not taken to have any qualities mitigating the seriousness of such things. Recall that in the first chapter I claimed this phenomenon of liking the unlikeable is often down to a given character’s being rendered in a more striking, insightful, or interesting manner than its fellows, and that in itself this need have no

---

connection to the character’s supposed moral worth, either objectively or relatively speaking. Kieran’s own explanation of differences in our fictive and real-world responses is somewhat similar, putting primary emphasis on the use of artistic techniques which, he suggests, can work to change our normative standards in relation to a fictional world and to make us sympathise more with particular characters. Regardless of the wording one chooses in explaining this sort of asymmetry, though, it is also very much worth noting the strong intuition that we are not under any moral obligation to hold these fictional characters’ faults against them – certainly nothing purely within the relevant fictions seems able to secure an obligation on our part not to, for example, love or ‘root for’ Baltar precisely because of his unscrupulous cowardice.

To examine the second difference between understanding people in fiction and in real life we shall begin by looking at the work of Alex Neill on empathy and identification, concepts the first chapter showed to be closely related to understanding, and in particular his application of them to the appreciation of the film Don’t Look Now in his article “Empathy and (Film) Fiction”. It is in this piece that Neill makes a claim supporting the relevance of our present line of inquiry, broadly speaking; that ‘empathizing with a fictional character is not… radically different from empathising with an actual person’. What this amounts to, for our purposes at least, is a side-stepping of the paradox of fiction, that is, the paradox of how it is that we come to have feelings for and about things which we know not to exist. It should not be especially puzzling to note that one is moved by Othello’s grief and regret insofar as the process of understanding and feeling for a fictional character is essentially the same as that by which one understands and feels for real people, for people in the past, in the future, and in counterfactual situations. As discussed in the first chapter this process, in philosophy of art typically referred to as ‘identification’, can be taken by us to be a subsidiary part of understanding.

Now, naturally a proponent of identification such as Neill does not have to claim that there is no difference in our reactions to life and to fiction – indeed such a claim seems manifestly implausible, as established in the previous chapter – the suggestion is simply that we should not be surprised at our being moved by fiction as such, that it is a natural and generally harmless consequence of our imaginative abilities and therefore that it does not constitute at least a troubling kind of irrationality. Unsurprisingly, on a cognitivist understanding of the value of fiction one can go further than this, and suggest the important learning that stands to be done as a result of our being so...

292 In much the same way we probably all have a friend of whom we do not think very highly in moral terms, yet whose company we enjoy because they are more interesting or amusing than many of our other acquaintances – though the extent to which we can overlook wrongdoing in fiction is of course far, far greater. 293 Matthew Kieran, “Emotions, Art, and Immorality”, p. 693. It is probably the latter capacity of artistic techniques which is in play with the two examples given here, since one tends to dislike the same flaws when manifested by other characters in the same fictions. 294 Alex Neill, “Empathy and (Film) Fiction”, p. 191. 295 As Gregory Currie points out, the sort of importantly simulatory psychological processes described and emphasised in this thesis satisfy ‘Moran’s constraint’, that our answer to the paradox of fiction should do something to shed light on our feeling for real but non-existent people. Currie, “The Paradox of Caring”, p 64.
emotionally affected by fiction in fact makes it rational, instrumentally rational, for us to be so affected; it is or can be rational for us to indulge such reactions. Nevertheless, while Neill’s position seems fundamentally sound, he arguably overlooks the importance of this response-difference in his reading of Don’t Look Now, which he suggests is a case where our empathising with – that is, ‘feeling the same thing with’ – a character is essential to our properly engaging with or appreciating the story. The film involves a married couple, the Baxters, moving to Venice temporarily, where they meet a seemingly psychic pair of elderly English ladies who claim that they can make contact with the spirit of the Baxters’ deceased daughter. The wife, Laura Baxter, eventually comes to take their claims more and more seriously, ultimately seeking to contact her daughter through them. Neill’s suggestion is that we must imaginatively engage with Laura and empathise with her terrible grief and secret hope or else ‘the rest of the movie will be lost to us’ because we would merely see her as being ‘pathetically deluded’.296

This is not a felicitous analysis because, as already suggested, it arguably underestimates or overlooks a difference between identification, or for our purposes understanding, as it works in fictional as opposed to real-world contexts.297 Neill suggests that we would evaluate Laura’s behaviour as desperate or irrational unless we really appreciate and share her feelings, in which case we would deem her to be reasonable. It seems very possible, however, that it is not the completeness with which one imagines her perspective that rescues one’s opinion of her actions, but rather facts about the fictional world in which she is contained. In the real world, for people of Laura Baxter’s background, it is desperate and to a certain extent pathetic to go to psychics in order to contact the dead. This remains the case no matter how intelligible her reasons for so doing might be. Indeed, it appears that this is expected to be our initial reaction during the film – a great deal of the dramatic tension in Don’t Look Now results from the fact that we are torn between, roughly, two interpretations of Laura’s behaviour, namely that she is being either reasonable or unreasonable. We only truly judge her to be reasonable when and to the extent the film establishes that it depicts a world in which psychics genuinely can contact the dead. It is not simply that we need to empathise with Laura to understand or condone her behaviour, but rather that in order to fully empathise with her we must first decide that her beliefs and concerns are not in fact misplaced or misguided. Imaginative projection into her position does contribute to such a decision, but it is by no means entirely constitutive of it, since some things are taken as so basic, obvious or important that no amount of personal perspective-taking can justify overlooking them (recall that this is why understanding need not advise us to forgive an indoctrinated Nazi soldier for failing to realise he is faced with a real and urgent case of wrongdoing). From this example we can see that the standards by which one judges the behaviour of

296 Alex Neill, “Empathy and (Film) Fiction”, p. 181.
297 It also comes up short because it obscures the fact that identification is in large part an evaluative practice, though perhaps not primarily one as Adam Smith would arguably have us believe.
fictional characters are more malleable than for real people, insofar as there is more room to change or reinterpret what we take as basic in fictional worlds.\(^{298}\)

This leads us to the third, very significant difference between understanding as applied to fiction versus real life: the natural inefficacy fiction has for causing moral obligations. Recall our *Blade Runner* fan of the previous chapter, who greets fictional scenes of desolation with extreme pleasure and interest. Earlier, I said this seemed harmless. But naturally this need not be the case; it is easy to imagine someone enjoying such scenes in what seems like a morally problematic fashion, or at least a fashion that gives cause for unease. Similarly, one can imagine there are ‘normal’ Hannibal Lecter fans and unhealthy Lecter enthusiasts. So, when is it blameworthy to imagine \(x\) and when is it not blameworthy to imagine \(x\)? One immediate worry relates to possibilities for real-world behaviour. One might object to certain kinds of imagining on the grounds that they promote overt behaviour that is morally questionable. One could then say that, more specifically, at least those imaginings are wrong which are entertained by people who are thereby (significantly) more likely to bring about bad things. But there are two problems with this response. Firstly, this worry rests on an empirical assumption that has not been adequately settled or detailed. Secondly, this seems inadequate as an explanation in that it looks as if we can preserve our moral condemnation of some imaginings even when they cannot contribute in any way to overt behaviour. Berys Gaut gives a case illustrating something to this effect in “The Ethical Criticism of Art”, where he describes a prisoner in permanent solitary confinement who engages in violent rape fantasies about exclusively fictional women. So why do we think such imaginings are wrong? Gaut’s position is that the prisoner’s imaginings

...are partly directed toward kinds... he is imagining them *as women*... [and this is] essential to his imaginative project... [so] by virtue of adopting such an attitude toward his imagined women, he implicitly adopts that attitude toward their real-life counterparts – and so reveals something of his attitude toward real-life women.\(^{299}\)

Gaut also advances a related view in his more recent work, in discussing the distinction between what it is fictionally the case that one has done in a given work and what one actually does in making something fictionally the case. Here he says that ‘a novelist should not feel guilty simply because he has written a fictional work in which something morally bad happens. Rather, it would be morally

\(^{298}\) This sort of move is also made by Matthew Kieran in “Emotions, Art, and Immorality”, where he points out that moral standards idiosyncratic to a given fictional world often make for an apparent difference in our response to it, pp. 689-90. But this malleability is of course not limitless, as evidenced by the puzzle of imaginative resistance. See for example Kendall Walton, “On the (So-Called) Puzzle of Imaginative Resistance”.

problematic if he enjoyed making morally bad things happen in the work’. This suggests that in the prisoner case we should condemn not only the attitude he implicitly takes toward women, but also the attitude he adopts to rape as such in relishing his imaginings, and indeed the explicit aim he has of indulging such an attitude.

It is not entirely obvious whether Gaut’s attitudinal approach merely operates alongside concerns about the promotion of bad behaviour, or if it in fact underlies and thus supersedes them. If imagining $x$ will make me more likely to $x$, for instance, then with or without my knowledge of this it would in an objective sense be bad for me to imagine $x$; I would be increasing the chances of $x$ coming about. But on a more subjective and immediate level any blameworthy wrong I thereby commit could probably be described in attitudinal terms, for instance as a lack of concern for the people whose interests $x$ affects. One might also wonder if Gaut’s example is subject to a deflationary story about counterfactuals; perhaps the only reason we feel the prisoner’s imaginings are wrong is because we cannot shake the assumption that they indicate a high likelihood of his performing certain wrongs should he be given the opportunity. But we need not settle these matters here, however, as in either event the above is sufficient for us to extract fictive externalism:

\[
\text{[FE]: the moral character of an imagining must be determined in reference to the non-fictional.}
\]

So, the question of whether it is morally wrong for someone to imagine that $x$, or imaginatively engage with a work that depicts $x$, is then always answered by reference to the real-world attitudes and responsibilities of the imaginer. Note that this circumvents discussion as to whether or not a work is morally and/or aesthetically at fault by inviting or inducing one to genuinely adopt morally reproachable attitudes. I can do something wrong in reading an innocent children’s book about bath-time, and I need not be doing something wrong by listening to a death metal song ostensibly endorsing infanticide. Two imaginative episodes can have precisely the same content, yet one be wrong and the other not, for instance if one is undergone by a rapist for the sake of sexual gratification while the other is done by a criminal psychologist for the sake of curing or capturing a rapist. Indeed, given that the very same imagined object can represent or express multiple things, a given imagining might well be blameworthy in one sense and not in another. Thus my imagining of Hannibal Lecter might be unremarkable in that it does not reflect any genuine attitude about cannibalism on my part, but blameworthy to the extent that it might somehow reflect my being...

---

300 Berys Gaut, *A Philosophy of Cinematic Art*, pp. 273-5. The distinction he discusses is that between media which can simply inform one that one has fictionally done certain things, e.g. in a theatrical aside, and media in which one has the power to actually influence what happens in the fiction, e.g. in videogames.

301 Gaut seems to feel that this would be a question of responsibilities *toward the real-life counterparts of given fictional beings*, but the reality is much broader; obligations relating to real-life people, institutions or things can serve as wrong-making factors in the imagining of overtly quite different objects.
chauvinist, say, or unhealthy respectful of dedicated egoists. It could of course also be positive, if for example my liking Lecter demonstrated that I had resisted the prejudicial urge to dislike all criminals soon after the mugging of my parents.

As in the above example concerning heavy metal, it is also worth noting here that FE doesn’t entail realism in reader responses or their assessment; the point is that to morally assess some piece of imagining we must look at the attitudes and/or consequences it really involves, not that attitudes can be straightforwardly read. Enjoying fictional violence does not necessarily imply endorsing violence in the real world, and so FE does not ask us to assume that someone’s apparent fiction-responses will automatically be mirrored to any given extent by her nonfiction-responses. A more in-depth case study should help to show the importance and usefulness of FE. Gaut once again provides a useful example for us to draw on, this time in *Art, Emotion and Ethics*:

Consider, for instance, a cartoon where Jerry puts Tom through a mincer in the course of one of their incessant battles. One would not want to say that this involved inviting the audience to enjoy acts of physical mutilation: cartoon characters do not suffer even fictionally, and instantly after his shredding Tom will re-form into one whole and very annoyed pussy cat. By moving the fictional world far enough away from the actual world in its basic properties, actions that would be morally wrong if done to real beings would not be wrong if done to their fictional counterparts, because of the morally salient differences between the properties of real beings and those that fictional objects are represented to have in these worlds.\(^\text{302}\)

What should we make of this example? It should be granted, certainly, that Gaut’s move can work; he is right in that sometimes what look like cases of morally unfortunate reactions can be explained away by appeal to relevant differences in the fictional world.\(^\text{303}\) But it is not clear that this move is sufficient to completely rescue the enjoyment of *Tom and Jerry* here. It is certainly true that Tom comes to no permanent harm, and in this sense it is clear that his fictional world is quite different from ours in terms of its basic properties, but at the same time it seems obvious that Tom is in fact supposed to feel genuine, and intense, pain and fear – hence all the screaming, throbbing, and glowing red. But not only would their pain-behaviour make little or no sense if they did not feel pain, neither would the entire structure of their resentments and interactions. The entire programme would make no sense were it not premised on the fact that the two characters dislike and want to visit harm and unpleasantness upon one another. Should we say then that *Tom and Jerry* is morally problematic?

---


\(^\text{303}\) This is similar to the move I made in regard to *Don’t Look Now*, and also indeed to the one made later by Matthew Kieran, as alluded to in a previous footnote. See his “Emotions, Art, and Immorality”, pp. 689-90.
Gaut warns against ‘bizarrely censorious attitudes’ in these matters, and it might appear that we are in danger of this if we condemn such a beloved children’s program. \(^{304}\) So, what would FE recommend?

The answer, first of all, requires us to remember that the moral character of our imaginings is dictated not by their content, but by our real-world aims, attitudes and responsibilities. This picture does quite a good job of explaining whether or not to condemn *Tom and Jerry*, and furthermore of explaining when and when not to condemn enjoyment of *Tom and Jerry*. First, it seems as though *Tom and Jerry* itself would be reprehensible if its creators enjoyed the thought of inflicting pain on animals, and/or if the manifested or implied creators seemed to enjoy the thought of inflicting pain on animals. Second, apart from that, it would seem that it is morally problematic to enjoy *Tom and Jerry* when this enjoyment is grounded on the fact that one really enjoys the thought of inflicting pain on animals, or the attitude that suffering in animals is amusing or of no consequence, and so on. Intuitively, this is quite appealing - the thought of someone enjoying a silly cartoon does seem quite different from the thought of someone watching *Tom and Jerry* with a genuinely sadistic glint in their eye. So we can track and explain the difference between what feels like an innocent laugh and what feels like a malicious one.

Still, though, is this enough to entirely rescue *Tom and Jerry*? Mice and cats do not run on two legs and they cannot make use of tools in any significant sense, so even for these reasons alone we might have cause to think our attitudes toward Tom and Jerry respectively probably do not have implications for our attitudes toward real cats and real mice. That is, one could not watch *Tom and Jerry* without being cognizant of the fact that one is watching an unrealistic, fictional representation of two characters only very superficially resembling a cat and mouse. However, without wanting to employ an unfair *reductio*, one might suggest that the cartoon could yet be cruel in much the same way that golliwogs are racist. Neither one is meant to be taken seriously, or as a direct representation of reality, but still both are products, indeed they are arguably icons, of a time and society that culpably failed to fully appreciate certain important ethical issues, those of racism and animal cruelty, and so it is no easy matter to distance them from the attitudes with which they are thus reasonably associated. In this sense it appears as though one can have ethical responsibilities stemming more from facts about one’s society than oneself; if one lives in a society that is racist or cruel to animals, or has a history of such, it is very tempting to believe one would face a heightened obligation to avoid fictions which assume they can unproblematically depict such things in an uncritical fashion, despite one’s own attitudes being different. Aside from issues relating to these sorts of wider, external duties, though, this touches on another important element of how we are to understand the ethics of imagining – the matter of what we might call *fictive transparency*.

Take the controversial and enormously popular multi-platform videogame series *Grand Theft Auto*, in which players the world over tend to go on murderous rampages, laughingly killing

---

\(^{304}\) Berys Gaut, *Art, Emotion and Ethics*, p. 249.
fictionally-innocent pedestrians in a bid to see how long they can evade the increasing waves of law-enforcement agencies sent to stop them – indeed most iterations of the series feature a ‘rampage’ or ‘kill frenzy’ mode in which one is typically tasked with killing as many people as possible within a given time limit. The people one kills in such instances scream, run, and often do not return from the dead – in short, their fictional properties are essentially supposed to be those of real people, and what is more we are apparently invited to imagine them as such, and as inhabiting a moral world much like our own. Again, FE would have us judge those who engage with such fictions in terms of the genuine attitudes they thereby indulge or reveal, although it bears remarking that the GTA games are more akin to faux-immoral dark comedies than thoughtlessly violent action movies. However, one of the more recent iterations of the GTA series, GTA IV, bears one striking difference from other GTA games in that it makes a large and deliberate step in the direction of realism and, in a manner of speaking, moral responsibility. It is more realistic on two fronts: its graphics are, for the first time, no longer cartoonish, and it features a more immersive level of detail, that is, one can engage in a wider variety of pursuits, with often vaguely realistic consequences (one stumbles and vomits if one gets too drunk, for example), together with slightly more realistic dialogue and more convincing voice acting. There is thus an arguable shift in emphasis from representational forms to represented objects, that is, GTA IV and other more realistic games and fictions tend to have a higher degree of fictive transparency.

Note, though, that fictive transparency need not be cashed out simply in terms of straightforward realism; inasmuch as there are few conceptual constraints on what could invite us to take a fiction seriously, we should say that fictive transparency is the extent to which a fiction invites or causes us in any way to understand it or some pertinent parts of it as though they were real. To this extent the form/object distinction might hold true, but begins to appear misleading; a videogame might feature prominently stylised graphics, for example, yet centre on the moral implications of one’s choices in such a moving and convincing way that one instinctively reasons in an earnest manner, as though say a cartoonish child’s suffering is naturally the same as a genuine child’s suffering. Conversely, a game might feature lifelike graphics and physics, yet have a story and characters so bizarre that one could only think of it as an unreal fiction. So the extent to which a fiction induces one to see it in terms of either its represented content or the manner of its representation generally is a sort of cover for the underlying ways in which different aspects of a fiction might draw on serious attitudes. Here ‘realism’ should refer, then, not to how closely something resembles its closest real-world counterpart, but the extent to which something calls on attitudes as though it was itself real. High fictive transparency, then, quite naturally undermines the conceptual defence that there are no necessary connections between what attitudes one has to a given set of fictional objects and what attitudes one has toward the real-world objects to which they correspond.
However, while a high level of fictive transparency should be of concern to us in the ethical assessment of a fiction and its enjoyment, it does not in itself guarantee the possibility of wrongdoing, and it certainly is not enough to entirely turn the tables and ensure a necessary connection between one’s attitudes to the fictional and one’s attitudes to the real. This could only be done by a Nozick-style experience machine that, Matrix-like, immerses one to the point where one is not at all aware that one is not engaging with the real world. Even the fantastical holodeck on the USS Enterprise does not meet this criterion, unless one enters and uses it unawares; in other words, one’s attitude to a fictional object only guarantees the same attitude to relevant real objects when one is fully under the impression that the fictional object concerned is, in fact, real. This is not to deny that an increase in fictive transparency can or does lead to an increase in the likeliness of such attitudinal correspondence, but in order to be understanding one must look insofar as one can at how seriously a given person actually takes a given fiction, not to mention our needing to remain cautious here if we are to avoid unwelcome implications regarding the appropriateness of censorship and the like.

The above might appear to conflict with what was said previously about our responsibility to be sensitive to the attitudinal flaws of the society in which we live – if understanding would have us assess specific agents’ fiction-attitudes and their actual personal implications, case to case, what difference should it make to Smith’s compartmentalised and thereby innocent enjoyment of Tom and Jerry that those living around him enjoy it for the same reason that they do not care about hurting real animals? But the two cases need not come apart. There are two factors in play, here: one is a matter of appearance and the other is a question of support. In terms of appearance, imagine the following example. Suppose that one has a fondness for an old golliwog doll that absolutely in no way corresponds to any racist attitudes on one’s own part. If one wanted to keep such a doll, it would have to be in the knowledge that it could give some people cause for discomfort – while such people would, of course, be under an obligation to be understanding and thus to be, insofar as they can, sensitive to the fact that one is not a racist, this might not always be possible for them, and so it comes down to a calculation of the likeliness and intensity of other understanding people’s discomfort against the extent to which the doll is personally significant to one. An understanding agent at least could not display such a doll, secure in the conviction that she was not racist, because an understanding agent must anticipate and coordinate with the likely perspectives of others. Hence the possibility that other epistemically responsible agents could be made uncomfortable by it could not be ignored. In terms of support, then, one’s non-blameworthy enjoyment of certain fictions or fictive objects are still responsible for the extent to which they perpetuate a wrong or bad system. Thus it would be wrong to buy a golliwog doll, even if one did so without the faintest hint of racist sentiment, insofar as doing so would support an industry and a cultural practice which is, in fact, racist – where

305 This appears to be a sentiment shared by Matthew Kieran; see his “Emotions, Art and Morality”, pp. 699-701.
this is constrained by facts about how likely or possible it is for one to realise this (which in this sort of case seems very possible, at least today).

There are two main worries facing a principle like FE. The first question is whether or not FE is accurate – are there cases in which we need not refer to the real world in condemning some piece of imagining, or in other words can there be inherently blameworthy imaginings? The second question concerns redundancy – if FE is true, is it telling us something that is not trivially or otherwise uninterestingly true? Those who are more inclined to press the first worry tend to question whether a morally neutral imagining can really be separated from the blameworthy attitude of which on some occasion it is the expressive vehicle. This raises a very difficult question about what exactly we take an imagining to be or involve. If two imagined scenes contain all the same imagery, yet have different emotional implications for the two imaginers or are imagined for different reasons, how justified are we in describing them as imagining the same thing? Their acts, generally considered, should probably be considered separate; in imagining $x$ for different reasons there is an undeniable sense in which they are doing different things.\footnote{As Anscombe might say, they would give very different answers to the question ‘why are you imagining $x$?’} But it does not necessarily follow from this that what they imagine should be considered different. So how does this affect the argument?

If FE is true, one might say that there are no morally bad imaginings as such, only morally bad attitudes which certain imaginative habits or actions can betray – let’s call this strict fictive externalism, or \textit{strict externalism}. If this first worry about FE is well-placed, however, \textit{fictive quasi-internalism} would be true; there would be some set of imaginings $y$ that, \textit{inseparably linked to certain blameworthy real-world attitudes}, would be sufficient to guarantee wrongdoing. Even if this sounds convincing to some, though, it would not be a serious loss, as FE itself does not claim there are no blameworthy \textit{imaginings}, but rather that imaginings can only be blameworthy by virtue of implications regarding the non-fictional; this is true even of an imagining that could not exist were it not to express some blameworthy attitude. Thus quasi-internalism seems more a subdivision of FE than an alternative to it. Fully-fledged \textit{fictive internalism}, on the other hand, would be the position that some things are wrong to imagine \textit{simpliciter}. I do not know who would seriously argue for such a position, and in particular I am not sure how such a position could allow for us to reproduce or otherwise imitate an inherently wrong imagining without ourselves doing wrong. Perhaps the fictive internalist would be happy to suggest that some wrong imaginings are ‘morally incomprehensible’; that a good person cannot imagine or be prepared to imagine some things. But such a stand seems extremely austere.

This leaves the second worry, that FE might simply point out something trivial. It is not just that we cannot describe the moral character of an imagining without some reference to the non-fictional, one might complain, we cannot adequately describe the moral character of \textit{anything} without reference to the real world. One might go on by further suggesting that no action, sentence, or any
other piece of behaviour, overt or otherwise, can be considered blameworthy without reference to the attitudes of the relevant agent. Alternatively, one could insist that the attitudes of an agent are irrelevant to, or at least not the be-all and end-all of, moral assessment – perhaps because one thinks attitudes only matter as indicators of likely behaviour or outcomes. FE does not in itself prove such views wrong, and to the extent that FE could be adapted to refer exclusively or primarily to the real-world consequences of imaginings it still wouldn’t be telling narrow consequentialists anything new or surprising. So why bother with FE? Because in practice people very often forget that FE, or something like it, obtains. A theatre might put on a notoriously offensive play, for example, fully expecting everyone involved to appreciate which aspects of the play warrant condemnation. The play might even be presented as a study piece, or in some other way that indicated its morally reprehensible aspects were not being endorsed by those producing it. Nonetheless there would be a good chance that some people would gather to protest against such a production. ‘Even though no-one here intends the attitude we find offensive,’ they might say, ‘simply supposing some things is sacrilege!’ Or suppose one is sitting with one’s partner at the end of Titanic, and is met with a sharp, tearful rebuke upon chuckling at the demise of Leonardo DiCaprio’s character Jack Dawson. Some people would suggest that one’s partner thereby implies an immediate normative claim, that it is wrong to laugh at Jack Dawson’s death. Such a claim would, according to FE, be false – the claim would instead have to be that in so chuckling one exhibits some insensitivity toward genuine death or suffering, for instance. We can see from the plausibility of these examples that FE is psychologically significant. I suspect the reason it is particularly easy to mistake fictional people and matters for having independent moral significance, over other objects lacking such moral significance, is that the point of fictional people, at least generally, is that they are like real people to us – in simulatory terms, there is nothing closer. This in turn points to the psychological fact that the wall between our beliefs and our imaginings is not entirely impermeable. So while FE does not ask us to draw direct conclusions from what people imaginatively enjoy or endorse to what they would enjoy or endorse in the real world, we are left with reason to believe there will often be some correlation between the two.

But there are still some relevant questions left to ask here, although they more concern how to understand the consequences of FE than whether or not FE should be accepted. One such question is how we are to understand the notion of wrongdoing internal to a given fiction or imagining – if FE suggests that a so-called ‘imaginative wrong’ can only be done by virtue of its implying something about the nonfictional, can we still say for instance that Macbeth does wrong in murdering his king? To pose this as more of a direct problem, should we resist FE because it appears to do violence to our ordinary talk about fictional characters and events?

This problem is largely illusory, though, or at the very most it is symptomatic of a fundamental difficulty endemic to all our talk about fictional events. The sense in which Macbeth’s regicide fails to be a genuine wrong is the same as the sense in which all of his actions fail to be
genuine actions of any sort, inasmuch as Macbeth and his actions do not, properly speaking, exist, and so cannot be described in normal (as opposed to normal-sounding) terms. So, conversely, we can say his regicide is wrong in the same way that we can say he becomes the king of Scotland. At this point a distinction introduced by Peter Lamarque and Stein Haugom Olsen is useful: that between internal/intrinsic and external/extrinsic perspectives on a fiction or fictional character. From an external, real-world perspective, a fictional character is nothing more than a collection of uninstantiated properties arising ‘from their linguistic origins’. Internally speaking, though, from a point of view within the fiction containing them, they are real people, even though externally they do not in this sense exist. In the same way, then, we should say that when we identify Macbeth’s actions as being wrong we are speaking, as it were, from the internal perspective.

An alternative approach would be to suggest that such claims are essentially counterfactual; that ‘Macbeth was wrong to kill Duncan’ really means something like ‘if Macbeth and Duncan were real, then it would have been wrong for Macbeth to kill Duncan’. This however leads us to another, two-sided problem. First we must ask how we are to deal with fictions that involve their own, different morality. Suppose Macbeth was in fact set in an alternate universe, where either the moral law (so to speak) stipulated that regicide was always acceptable or where it was impossible for people to know regicide could ever be wrong – let’s call this play McBeth. Our counterfactual assessments of McBeth’s moral fibre then becomes very muddled. We can easily say that internally McBeth does no wrong in killing Duncan, but as a counterfactual it is not obvious if we should or can hold the fiction’s moral laws constant. Or imagine another version of the play, O’Beth, where morality is not supposed to be any different yet the play manifests the attitude that O’Beth’s actions are good. Again, it is relatively straightforward to say that externally the fictional character O’Beth has ‘a good man’ as one of his properties, where we can dispute what the internal fact of the matter is without any great conceptual headaches. In counterfactual terms, though, we again run into confusion over which aspects of the fiction we should translate into reality – it is simply much more difficult to know when or how we can hold our understanding of the work over what appears to be that of its author. The second aspect of this problem, as discussed by Peter van Inwagen, is that it is not even clear that the character Macbeth (or McBeth/O’Beth, no less) can sensibly be understood in counterfactual terms at all, and for similar reasons; fictional characters have crucial ‘literary properties’ (such as being the main villain in the fiction x, or appearing in chapter nineteen) which cannot be ascribed to real people. So, on balance, it appears safer to go with something like the picture offered by Lamarque and Olsen.

307 Peter Lamarque, Fictional Points of View, p. 13. See also Lamarque and Olson, Truth, Fiction and Literature, pp. 142-3. It is in the latter text that the words ‘intrinsic’ and ‘extrinsic’ are used, but this eventually gives way to the usage of ‘internal’ and ‘external’.

Another, distinct question results from what some might consider a counterproductively unlikely example. Taking FE as true, what should we say about a fictive internalist who enjoys imagining torture in some way entirely insulated from his real-world attitudes? This sadistic internalist does not have pro-attitudes toward actual torture, yet fully believes that imagining torture is wrong in itself – all the while deliberately indulging his desire for such imaginings. Can we then say that this is a counterexample to FE, that his imaginings are wrong though they do not reflect any blameworthy attitude toward the nonfictional?

The most immediate objection to this apparent counterexample is to say, as already suggested, that it is too implausible to make much or any difference. Indeed it might be impossible for a human being to have fictive and real-world preferences that are entirely insulated from one another in the manner this counterexample suggests. But the viability of this move is something more for psychologists to assess than philosophers, relying as it does on empirical information not easily accessed from the comfort of the armchair. A similarly immediate response would be simply to deny that this sadistic internalist is doing anything wrong; the thought being that this man’s ethical confusion should not affect the truth he fails to see. But this might well strike some as a hasty and presumptuous move; there does seem to be some intuitively appealing sense in which, though this man might not be doing anything objectively wrong, still he seems to be doing something blameworthy.

However, even if it is granted that the sadistic internalist is guilty of some moral transgression worthy of our attention, this imagined case still fails as a counterexample to FE. The source of his wrong in this story, if there is one, is his deliberately going against what he might somehow have good reason to think are his moral obligations – not in his actually performing some specific action which morality advises against. To insist on his imaginings themselves as being the source of his wrongdoing is to ignore the quite specific explanatory story by which we were even able to arrive at such a supposed counterexample, to simply deny FE and its plausibility at the cost of embracing a quite counterintuitive and unnecessary understanding of the case. That is, for the counterexample to get off the ground at all we must include information about his unusual beliefs regarding the nonfictional, information crucial to our thinking he does wrong. In other words, this story depends on the sadistic internalist’s wilfully engaging in imaginings he regards as inherently wrong – and thus indisputably involves his having a cavalier real-world pro-attitude toward flouting his real-world moral responsibilities.

Although it should be clear from this that FE has a perfectly adequate means of deflecting such a counterexample, there might still remain some who cling to the intuition that it somehow goes through. Perhaps such people think for instance that the above counterexample involves two loci of wrongdoing, his cavalier attitude and his sadistic imaginings, where FE can only do half the explanatory job, as it were. It is prudent, therefore, to offer here some other explanation of why we
should accept something like fictive externalism over something like fictive internalism. The most basic means of pursuing such an explanation is to point out that fictional characters and states of affairs do not exist, and to suggest that because of this they logically cannot be the source of any kind of constraint on us. But this will not do – after all, we appear to have all sorts of moral obligations in relation to things which do not exist, perhaps the most widely recognised being our obligations toward future generations as yet unborn. There are two responses to make in the face of this problem. Firstly, it is fair to say that there are two kinds of nonexistence in play here; contingent or impermanent nonexistence and necessary or inherent nonexistence. Fictional characters cannot exist under any circumstances, and an existing thing can at most resemble a fictional thing to some fundamentally incomplete extent. Morally speaking the whole point of future people, or indeed people in counterfactual situations, is that they could in theory come to exist, and furthermore it should be apparent that any given moral claim they place on us varies in strength according to how likely it is that they will exist. Secondly, as has been remarked several times in previous chapters, ethical relevance is usually cashed out in terms of a creature’s possessing certain qualities, typically rationality and/or the capacity to feel pleasure and pain – or, as I have argued, consciousness. Again, hypothetical people potentially possess these qualities, whereas fictional characters cannot genuinely possess any qualities whatsoever. So we can see without any significant theoretical commitments that moral obligations involving the fictional, where this does not have any implications for the nonfictional, cannot make sense. Thus even if we were to change our sadistic internalist example, so that he was not an internalist at all and thus was not consciously doing wrong, an intuitive assessment of wrongdoing in such a case very much appears to lack any rational justification.

This might be seen to raise an awkward question for our overall endeavour here, though. Since fictional characters are inherently non-existent, and so cannot truly possess any (morally relevant) qualities, they cannot be an immediate source of moral obligations; at most they can become involved in some moral obligation by virtue of something non-fictional. So we might for instance suggest that failing to understand fictional character x shows a tendency to fail to understand real people like x or in x’s situation, or we could say for example that an English teacher has an obligation to understand x because x was created by a student in her care. Recall also that in the first chapter I suggested there is a basic moral imperative to exercise understanding toward others in that misinterpreting someone, or at least misjudging them, does them a disservice. Clearly then, all things

309 More fully, we should say that the moral claim strength of hypothetical people varies according to 1. the extent to which we have control over matters, 2. the seriousness of their complaint, should they in the event have cause to make it, and 3. the likelihood of their in the event having cause to make it.

310 This is similar to Gaut’s abovementioned move regarding the incarcerated fantasist – as our sub-optimal attitudes can be reflected in our failing to understand or imaginatively engage with a given fictional character, so too can our imaginative habits indicate our morally sub-optimal attitudes to real-world objects. The latter of my two examples above, though, also illustrates why this matter does not boil down to merely a question about fictional counterparts of morally significant objects. See Berys Gaut, “The Ethical Criticisms of Art”, pp. 187-8.
being equal, we cannot have such an obligation to be accurate in our understanding of a fictional character. So does FE imply that should we not, after all, bother talking about understanding and fiction?

5.2 Understanding and Interpretation:

The answer, of course, is no. While all things being equal we cannot have an obligation to understand fictional characters, recall that I have said understanding picks out those things which could be useful to our better understanding people and/or creatures of moral significance. Given that fictional characters are for the most part expressly modelled on people, it seems likely that practice at one tends to count as practice at the other – where this assumption is of course supported by the discussion of the fourth chapter. To this extent all things will tend not to be equal, as it were, and we are left with a strong if indirect obligation to exercise understanding toward fictional characters. This should not seem odd - recall further that it is an already stated and accepted limitation of understanding that it cannot perfectly reproduce the perspective of another in its entirety. From this it follows that whatever perspective we imagine for another, it is almost certainly the case that the imagined perspective does not apply to them, insofar as it does not perfectly apply to them. So obviously the point of being understanding must be to try to understand others as fully and properly as possible, bearing in mind the sorts of conditions that have already been discussed regarding privacy and so on. If understanding allows for some degree of discrepancy between what we imagine in understanding and the target of our understanding, then, it is open to suggest that as long as our understanding aims at something sufficiently like a real person or consciousness it is appropriate. Trying to understand a fictional character, then, is not inappropriate, it is not a waste of time, insofar as either broadly the same psychological and moral laws we think of as holding to us hold to them, or because their psychology, when different, demands comparison to our own.\footnote{This is due in part to the fact that fictional characters are largely ‘supposed’ to exist as people, internally speaking.} This is much the same reason why it is not a waste of time to think about people in the distant future, or the distant past. And, as Aristotle says, fiction is not so much about things that happen as the kinds of things which might happen, so we should be no more deterred by the inherent non-existence of fictional characters than by the contingent non-existence of people in counterfactual situations.\footnote{‘The poet’s function is to describe, not the thing that has happened, but a kind of thing that might happen’, \textit{The Complete Works of Aristotle}, p. 2322 (Poetics 1451a36-b2).}

Another suggestion presents itself here, though, since fictional characters are not real yet their creators are. Is our primary or sole responsibility when understanding fictional characters to try to enter into the imaginings of the artist who created them, or in other words, must our attempt to understand a fictional character somehow be an attempt to understand its creator? Fortunately, understanding is not bound to endorse such a heavily intentionalist stance. I would suggest, rather,
that understanding sees the author’s intentions as one important aspect of a work qua something intentionally created, one that should not be entirely ignored. But it by no means suggests that the facts about language, fiction or art more generally forbid there being other, equally or even more profitable or appropriate interpretive strategies. In fact it might well be seen to actively invite this view, given the emphasis placed here on art and fiction’s capacity for enriching understanding; arguably we have in this sense some obligation to get as much from them as we can. But we shall return to this sort of suggestion later on this chapter.

In the meantime let us take it as established that understanding is apt to provide worthwhile insight into fictional characters, remembering that one should not from this begin to form the genuine belief that there is, as it were, a real or ontologically robust object of one’s understanding, as opposed to something more like a set of cues for imagining. Thick understanding here depends for its appropriateness on the kind of make-believe attitude which, as the second chapter suggested, allows us to rationally talk ‘to’ plants – but while being understanding toward plants is merely permissible in optimal conditions, we have seen the understanding of fictional characters to be generally advisable, prudentially speaking, and morally important or commendable. So if we can sensibly be understanding toward fictional characters, then, what does understanding have to tell us about how to understand them? Things should remain more or less the same as before, one would expect, with the difference of our being motivated by more pragmatic concerns than moral ones. We might draw one possible difference from Lamarque and Olsen, who mention that genre conventions can influence our understanding of fictional characters, as when one resists suspecting the first person to seem guilty in a detective story (since initial misdirection is a standard trope in such fictions). However, one could argue this sort of practice is analogous to adjusting one’s understanding of real people according to relevant cultural standards – culture here acting as the analogue for genre – for example, in a polite society one might have to expect that, perhaps confusingly, not all offers of help are given in the expectation that one would ever accept. And what about the sort of worry raised earlier, about limiting our understanding of others; suppose a given novel or painting depends for its proper reception or appreciation on one’s failing to entirely understand some of the depicted characters. Suppose then that one reads this book or looks at this painting, wishing as moral agents generally should to be understanding. Should one, according to the spirit of understanding, try to understand these characters as much as one can or should one refrain from ‘figuring them out’ past the point which would be optimal for appreciation of the piece? The first thing to wonder in the face of such a question is probably how we are to tell what the optimal level of understanding for these characters is. More specifically, are the signs we must watch out for to be found inside or outside the individual characters? Granted, it might in a sense be part of Smith’s

313 Lamarque and Olsen, Truth, Fiction and Literature, p. 95.
314 For present purposes let us excuse the murky idea of ‘optimal appreciation’ as such.
character that he should only be understood so much, but this is not something we would appreciate in imagining his perspective (although it is instructive for conducting the project of such imagining), and it certainly does not make sense to perform the sort of move we did in the first chapter regarding those with accountably wrong beliefs; we cannot suggest that it is an aspect of Smith’s perspective of which he should be aware.

This would be a problem for identification, but understanding is a broader concept and can include facts that are unknown and unknowable to the object, provided at least that these facts do not form part of one’s thick understanding. The question remains, though, of where the knowledge that these characters should only be understood so much is to be found, if not in the characters themselves. It might be possible to express it as a quality of a given character, but it looks as though it must be a quality defined and set by wider facts about the story in which she is situated. So it is these wider facts that ultimately determine the appropriate level of understanding — or, as Lamarque and Olsen put it, the external perspective determines ‘the kind of involvement appropriate from the internal perspective’.315 One’s trying to understand a character, then, is directed by that very fact to understand the overall story which contains them. This is roughly analogous to the idea that wider moral concerns might potentially influence the point at which we should stop understanding someone, and quite closely analogous to the fact that we must understand something about a person’s background and culture if we are to properly understand them and the degree to which they are likely to welcome such understanding.316 The difference, of course, being that if Smith is a private fictional character his fictional wish for privacy cannot itself produce any obligation on our part to restrict our understanding of him — such could only result from a pragmatic constraint on our understanding stemming from our general desire to understand the story properly, or an indirect moral constraint insofar as disregarding Smith’s wish might for instance cause us to pay less attention to real people’s desire for privacy.

Let us turn here to elucidate two concepts which could use further clarification — what exactly does it mean to gain insight into a fictional character, and what does understanding have to tell us about the ‘optimal appreciation’ of a piece? Jenefer Robinson, in Deeper Than Reason, suggests that it is an error to pause and reflect upon a story ‘after every page... [because] we are likely to arrive at mistaken beliefs’.317 This seems to have implications for both questions; perhaps we gain insight into characters by forming accurate beliefs regarding them, and perhaps it is only upon finishing a novel that we can be sure how much and at what stages we should have understood the characters — although by this stage it would of course be too late! But on inspection these suggestions do not get us very far. The idea of forming accurate beliefs about a fictional character does not seem significantly less murky than the idea of gaining insight into them, and the notion that we should finish a book

315 Ibid., p. 156.
316 Indeed, this also points toward the idea that to properly understand what an author meant by something we need to understand things about the context, cultural and otherwise, in which she was writing.
317 Jenefer Robinson, Deeper Than Reason, p. 156.
before making any settled judgements concerning it might tell us something impractical about how to reach an optimal appreciation, but without really telling us what such an appreciation consists in or is defined by.\footnote{Even then, it is not clear how appropriate Robinson’s claim is, here. Insofar as we must form working hypotheses about a character, real or otherwise, as we deal with them, her suggestion does not seem plausible other than in the trivial sense that we should not form final, settled beliefs until all the evidence is in.}

Now, it would not be fair to suggest that ‘accurate belief’ about a fictional character is \emph{no less} murky than ‘insight’ into a fictional character. I might believe that Sherlock Holmes lived at 221b Baker Street, where this belief is straightforwardly accurate (as far as fiction and truth usually go), without this having any special implications for my degree of insight into the character Holmes; in other words it seems that while ‘accuracy’ is not an unproblematic notion, it perhaps includes more easy cases than ‘insight’. The picture gets somewhat murkier as the beliefs concerned become more speculative, indeed one might say it gets murkier as belief approaches insight, so we must try more directly to explain how ‘insight’ is to work here. To gain insight into Sherlock Holmes, let us say then, would involve acquiring a sense of what we have reason to believe would be his internal life, which would most probably go hand in hand with an improved ability to predict and interpret his actions, and indeed to fill in the likely actions, thoughts and feelings missing from a history of some past series of his doings.

But where does this get us, the understanding reader must ask. Is it not a mistake to call the sense we thereby construct accurate or insightful when that \emph{actual} thing they aim at, Sherlock Holmes, thinks and feels nothing? Here we must again return to Lamarque and Olsen’s distinction between internal and external perspectives, and remember that we are responding to something like an arranged set of properties, or cues for imagining. The set of scenes containing or relating to some character \(x\) form the basic parameters for what construals of \(x\) are acceptable – one’s idea of what a character is like, for instance, should not contradict \(x\)’s exhibited behaviour without some justificatory story. As Lamarque and Olsen put it, in regard to what sorts of imaginative filling-in by a reader are legitimate, one’s understanding must fall within what is ‘\emph{authorised} by the content itself and its presentation, where the relevant authority resides not only with the author but with the genre, literary tradition, historical context, and so on.’\footnote{Lamarque and Olsen, \textit{Truth, Fiction and Literature}, p. 89.} Within these parameters we might also add a constraint of plausibility (relative to what appear to be the internal standards of the fictional world) versus something like ‘dramatic value’.

Take Marv, the protagonist of the film \textit{Sin City}: given that he spends the film beating and killing various people in a murderous quest for revenge, one might think it impossible to describe him as ‘a good guy’, but such is not the case. To do so, on one level, one could point to textual evidence illustrating reluctance on his part, or remorse, or some sort of extenuating condition (mind control or some such) that could preserve the \emph{fictive coherence} of one’s interpretation – in other words, to show...
that the description ‘good guy’ is appropriate according to terms internal to the fictional world.\textsuperscript{320} Psychological laws might hold in the fictional world which do not in real life, for example; the sight of the colour purple could in a given fiction send even pacifists into a non-blameworthy psychotic rage, where this would not be a serviceable explanation in real life. On another level, one could of course also point to general facts about the fiction, but not fictionally true in its world, which do not directly pertain to Marv: perhaps one could say that in our world one would not call him a good guy, but that the fictional world he inhabits is so gruesome and violent that he can in a sense be thought of as ‘a good guy’ within it. But neither approach need be, it is worth noting, a typically \textit{fictive} justification insofar as they bear some similarity to our ability to judge real people in terms relevant to their own, different contexts – perhaps the primary difference being, as has already been mentioned more than once, that we are generally prepared to uncritically accept a wider range of differences in fictional contexts, and that in fictional contexts we usually operate more on pain of pragmatic mistakes than moral ones. But barring these kinds of more or less fictive determinations about the meanings of behaviour, different moral norms and so on, or in other words assuming a relatively realistic and close to home specimen of fiction, one can generally think of one’s insight as accurate to the extent that \textit{were} the imagined scenes genuinely occurring one’s assessment of them would be reasonable.\textsuperscript{321}

Given that fictions cannot practically fill in all the details relating to the worlds they depict, and so that the ‘reader’ is invariably left to fill in (through unconscious presumption or otherwise) many such details, this looks as though it still leaves us with a lot of room to differ in our respective interpretations of a given character. Should we be concerned about this? I do not believe that we should – in fact, this should be greeted as good news. There is no special reason why we should expect or demand that a single text have \textit{only} one valid meaning or interpretation. If one values the learning, or more broadly the \textit{cognitive exercise}, involved in engaging with art or fiction, it seems at least on face value that the more substantive interpretations a work supports the more valuable opportunities it affords for discussion and thought.\textsuperscript{322} Indeed it also seems as though in life one often cannot be certain of the understanding one has of real people, \textit{so realistic practice at understanding might not involve definite answers so much as the weighing of possible interpretations}. Having said this, it is of course one glaring disanalogy between literature and life that the former typically involves much greater access to the thoughts and feelings of characters, and so this aforementioned possibility of ambiguity could be seen as an important step in mitigating the ‘stabiliser effect’ such increased

\textsuperscript{320} This sort of claim is also offered by Matthew Kieran – see his “Emotions, Art and Morality”, pp. 689-90.
\textsuperscript{321} This should not be confused with the ontological position that characters possess properties insofar as if the characters were real they would genuinely have them, which has suffered criticism in the past. See for example Peter van Inwagen, “Creatures of Fiction”, p. 306.
\textsuperscript{322} Note that such a work should support interpretations that are interesting and justifiably supported by the text to a reasonable standard, otherwise Rorschach diagrams for example might counter-intuitively qualify as an art form \textit{par excellence}. 

124
epistemic access has; otherwise, like training wheels on a bicycle, the inner monologue of a fictional character might make them easier to understand while not doing much to improve our ability to genuinely read thoughts in other people.\textsuperscript{323}

Of course, if this picture is simply too vague for some, there are further limitations one could reasonably seek to add. As alluded to before, for example, given that a novel is an intentionally produced object it does not seem unreasonable to feel that one’s interpretation should be in some way constrained by facts about the author. For reasons along these lines many suggest that one’s interpretation of a work, and hence presumably one’s interpretation of the characters contained therein, should be limited by what one thinks to be possible of the author.\textsuperscript{324} This would mean that one could not read a Shakespearean character as being a Marxist caricature, for example, although one could probably still read a character as embodying something like a very generally Marxist position. But such examples need not always seem so intuitively appealing – it might seem obvious to some that we should not read a given text as espousing very specifically a political position that was not developed until some centuries after the author’s death, but this intuition will in most cases probably wane as the conflicting characteristic of the author becomes more personalised and obscure, and the contrasting reading becomes more interesting or otherwise aesthetically or academically fruitful. This is particularly so given the pressure the understanding agent is under to find cognitively rewarding or enriching interpretations of characters and artworks, and conversely to disapprove of or avoid cognitively or morally pernicious ones. So if a story involves a character $x$ who could, coherently with the rest of the text, be read as showing some insight into feminism, it is not entirely clear how much we should reject or suppress this reading if we discover that the author was contemptuously chauvinistic in his innermost private life.

This problem also depends in part on the purposes one has in engaging with a given fiction or artwork. It might very well be more permissible for a politics student to force a Marxist interpretation of Shakespeare than for a historian to do so, for example, given the greater relative yields each might thereby achieve within the contexts of their fields – and beyond this one’s general ideas about the purpose and nature of art will naturally colour one’s intuitions further.\textsuperscript{325} I take here something rather like what David Davies calls a ‘consumer-based’ position, which primarily emphasises the possibilities for enjoyment and enrichment to be had by those engaging with art.\textsuperscript{326} The main difference here being a relative absence of an emphasis on enjoyment as such, according to the priorities of understanding and cognitivism. The term ‘consumer-based’ is also mildly unfortunate,

\textsuperscript{323} Note that, as with Gaut’s fictional properties argument, this does not lead to a radical difference in how we understand, or at least how we assess and react to, fictional characters as opposed to real people.

\textsuperscript{324} One example being Robinson, \textit{Deeper Than Reason}, p. 192.

\textsuperscript{325} One might well suggest, for instance, that formalists need never give much or any thought to what an artist’s actual intentions might have been, except perhaps to track which artists are likely to produce other works of a similar quality (thereby discovering whose work to avoid and whose to seek out).

\textsuperscript{326} David Davies, \textit{Art as Performance}, p. 76 n. 12.
though, given its commercialist undertones; such a position should not involve market-based considerations (such as value for money or production cost), and beyond this to describe one engaging with fiction or art as simply ‘consuming’ it seems to suggest a misleadingly passive or unthinking activity. But these problems aside, something like my qualified version of the stance it describes leads me to hesitate for fear of setting too strict a set of standards according to which one may or may not interpret a work, since the possibilities for enrichment that a work offers must vary according to person, time and other contextual factors.  

Suppose one still wonders by what standard one is to estimate the accuracy of one’s insight, of one’s beliefs, even if one is not troubled by the fact that strictly speaking there are no psychologically internal states which they aim at or seek to reproduce. Suppose one reads Sherlock Holmes stories devotedly, committing all of one’s imagination, intellect and effort to constructing as vivid and as faithful an impression as could be had corresponding to the character as described in Conan Doyle’s stories. Then, about three quarters of the way through the catalogue of Holmes’ adventures, in a needless epilogue to one such story, Holmes asks some Americans to stay and converse with him because, he professes, “It is always a joy to meet an American… for I am one of those who believe that… our children [will] some day [be] citizens of the same world-wide country under a flag which shall be a quartering of the Union Jack with the Stars and Stripes.”. Suppose, further, that this not only seems entirely unprovoked and tangential to the usual development of such stories, but that it even seems to strongly conflict with one’s sense of who Holmes is. What is one to make of such an occurrence? Should we revise our understanding of Holmes, or should we overlook it as some sort of mistake on the part of the author? If we should revise our understanding, what implications does this have for the accuracy of our previous understanding?

Let us briefly delay the answer to these questions, because all of this very much feeds into our earlier question of how we are to cash out the idea of ‘optimal appreciation’, at least according to the understanding perspective – or in other words with a view to appreciating art in an ethically conscientious fashion, as an ethical exercise or at least one with ethical implications. This is an inescapable question, as surely as entertainment-value cannot justify unfairly slanderous interpretations of a work and artistic licence cannot justify the pursuit of ethically questionable art

---

327 It is a corollary of this view that the extent to which a work x should be valued by person or society y can vary over time, but this does not or need not imply that as it were the objective value of work x need change, since it might make more sense to see that as determined by contextual factors local to its time of production.

328 This example has been altered somewhat for present purposes – Holmes says this toward the end of The Adventure of the Noble Bachelor, which is actually in the first collection of Holmes stories.

329 This is indicative of the wider problem of how to react when certain evidence goes against an understanding of someone that one has confidence in – again, perhaps the central difference being that here it is primarily a pragmatic rather than an ethical issue.
forms. So what, then, does understanding tell us about ‘optimal appreciation’ of a piece? Recall that we have said fictions cannot be the immediate source of moral obligations. So what generic obligations does one face upon picking up a work of fiction? As previously suggested, if there is one characteristic common to all artworks it is that they are on some level intentionally produced – even found art is, by virtue of being presented, imbued with intentionality, and even art made through some deliberately randomised process is initiated, presented or conceived of intentionally. This element of intentionality is a crucial difference-maker between art and the naturally beautiful. A majestic sunset and the sound of gently falling rain can both be beautiful, but neither on their own can be considered art. Now, understanding warns us to exercise some minimal degree of caution whenever interpreting the acts of agents, particularly their intentional acts. Since it does someone a disservice to avoidably and significantly misinterpret their words or actions, it naturally does them a disservice to avoidably and significantly misinterpret those actions of theirs which they feel to be invested with particular meaning or significance, as surely works of art tend to be. So an understanding agent, on deciding to engage with a given fiction, should also thereby resolve to have at least some minimal interest in what its author could have meant in creating it.

Note though that this of course is not a commitment to (wholesale) intentionalism. It means that authorial intention, when broadly accessible at least, cannot ethically be ignored, but this leaves ample room for alternative interpretations of a given work, provided such interpretations are rewarding according to cognitivist criteria and do not dress themselves as representing the intended aims or meanings of the work’s creator. Even then, of course, given the natural obscurity of authorial intention there remains room for debate about which meaning should be assigned to the work even on just the more restricted part of this approach. But regardless of this difficulty within intentionalism, the text an author produces, for instance, is a thing separate from herself, and as I have suggested artworks are in large part so valuable because they ideally serve as ‘springboards’ to constructive imagining, that is, certain kinds of imaginative and/or emotional learning or exercise, not necessarily reducible to internally possessed meanings of the work. In light of this we ‘the consumers’ should be allowed to partially appropriate works, as long as the nature of this appropriation is properly appreciated and the part of the author, the creative agent, is not neglected, does not disappear. Even if it is the artist’s expressed wish and view that artworks either generally or in the relevant instance should be interpreted with no heed to the artist, as responsible agents we face an obligation to at least check to see if this is the case. So, if it really improves the story in some meaningful sense, for example by making it more intellectually stimulating rather than simply enjoyable, we can for

---

330 The latter claim here calls to mind David Bowie’s concept album *Outside*, which described a cyberpunk (then) future in which a member of the art police investigated which apparent crimes should instead be considered art.

331 Barring some highly contentious theistic argument, at least – but even here this would be beside the point as such would rely on God’s intentionality in creating natural features $x$ and $y$, etc., rather than on any person’s, and so a clear distinction would remain between art and nature.
example think of *The Turn of the Screw* as employing an unreliable narrator, while bearing in mind on some level that ‘it is not meant to’. Let us call this *compatibilist pluralism*. It is not driven by simple hedonic interests, but rather the cognitivist desire to ‘get the most out of’ a work in broad terms, sensitively and respectfully – that is, the compatibilist pluralist seeks to interpret a work not simply for the sake of what the author meant, though she does and must care about this, and not for the sake of what is merely enjoyable, but rather for the sake of what is enriching, enlightening, meaningful and so on.

To put it directly, then, in terms of the understanding perspective, an optimal appreciation of a piece is any cognitively rewarding set of appreciations that incorporates or acknowledges some reasonable knowledge of the author and what she might have intended – we have no reason or need to think that understanding can provide us with an exhaustive theory of art, and so the factors which should determine one’s further interpretations of a work are up for debate. The fact that this description is quite broad, in other words, is of no consequence. But again, what does this mean for understanding and the extent to which a character is to be optimally appreciated? This becomes a particularly thorny question in light of the cognitivism built into understanding’s aesthetics; if some, though not necessarily all, of art’s value comes from its capacity to teach or enrich, and if the understanding agent has something of an overriding, that is, an ethical interest in exploiting art’s capacity for such, what is one to do when caught between enjoying a work and learning the most that one can from it? In practice, though, this is probably not a very difficult problem. Appreciating and learning from a fiction is not like reading a set of instructions or a list of facts – what one learns from a story can take a great deal of time to ‘sink in’, or might only be arrived at after a process of due reflection, or might very well exist and operate subconsciously for the most part. Consequently it seems likely that in normal conditions one could afford to put off greater understanding for the sake of more immediate enjoyment. If, on the other hand, enjoyment of a fiction or fictional character were to rely on an destructive, prejudicing or otherwise actively unenlightening view, then of course such enjoyment should simply be rejected, at least where it is seriously invited as opposed to ironically entertained.332

So finally we are left with the questions we put off answering a moment ago – what, that is, should we make of our problem concerning Sherlock Holmes and Arthur Conan Doyle? If a well-established fictional character exhibits some uncharacteristic behaviour in a way that the fiction does not explain or justify, how should we take this? The most natural first step here is, perhaps, to appeal to the idea of realism. If Holmes is very much a certain sort of person, and on one occasion exhibits some uncharacteristic trait, we can object to this on the grounds that it makes Holmes as a character

332 I would suggest that most or many cases where such unfortunate attitudes are ironically entertained actually attempt to say something positive and constructive about such views being, in fact, wrong and unfortunate. This is why for instance many ‘shocking’ or ‘nihilistic’ comedians such as Frankie Boyle see themselves as in fact being quite morally conscientious.
somewhat fractured, unnatural and unconvincing – it feels like the voice of the author, or at any rate a foreign voice, being forced upon him. But how fair is this complaint, exactly? It does not seem unrealistic, surely, to suppose that someone, a real person, could on a given occasion behave in an uncharacteristic manner, or briefly take up opinions they would later reject. In fact it seems unrealistic to think of people as totally coherent! So there is some force behind the suggestion that our example could even be an example of good writing, not bad. But of course, this depends; there can be cases where unusual behaviour in a fictional character can be put down to realism, when it seems to be done deliberately, and times when it can be put down to bad writing, when it seems the work is unaware that the behaviour is unusual, although specifying in more precise terms when to tell these two apart is clearly a more difficult question, and perhaps one better left to literary theorists.

Having said this, it would appear all the same that our example is quite likely to be a case of bad writing, given that Holmes’s words appear to be presented as a standing opinion of his, rather than a temporary aberration of mood or belief. In any case, let us suppose such for the moment, that in this scene the author (actual, postulated, manifested, or implied) intends (or appears to intend) for Holmes to voice a genuine set of beliefs and desires which, though thoroughly uncharacteristic, are not deliberately portrayed as such in the light of any textual or biographical evidence. We should accept that this constitutes a flaw in the work. But how exactly should we spell this out? Can we say that Arthur Conan Doyle has misunderstood Holmes? In this limited sort of case it seems we can, even if on a basic level it would be impossible for an author to radically misunderstand one of her own characters. Note that this sort of move need not rely on a presumption of realism in the relevant fiction or its characters, even if this looks like a safe bet in a Victorian detective story – this is important because, as mentioned in our discussion of Marv, facts internal to the fictional world, or facts about the genre in which it is situated, can militate against the straightforward transference of real-world psychological standards (in our Holmes case, standards concerning the coherence of attitudes). What is necessary, rather than psychological or general realism, is a broad kind of coherence; insofar as a fictional world might for instance deliberately involve an entirely erratic psychology, a prima facie case of coherence would then probably count as fictively incoherent. And we can see that even if such atypical psychology is stipulated in a fictional world the reader might still be able to sometimes better keep track of its consistent application than the author.

In saying this, as in saying that Conan Doyle has here misunderstood Holmes, we affirm that our understanding of a character can at least sometimes be held over that of its creator. So our

333 Indeed, this fact forms the basis of situationist criticisms of virtue ethics’ emphasis on character.
334 It should be noted that while we might employ different everyday psychological standards in understanding someone from a very different context in real life, these standards would still be ‘real-world’ psychological standards inasmuch as there would still probably be some difference in what we would consider plausible or possible in life as opposed to fiction.
335 This is less restrictive than Lamarque’s suggestion that authors are typically limited by ‘constraints of plausibility, verisimilitude, and aesthetic coherence’. Fictional Points of View, p. 37.
previous understanding of Holmes was accurate, and we are left with an ugly anomaly to work around. What should we do with this discrepant scene? Can we say that it somehow ‘does not count’? Can we say that the Holmes it contains is not, in some way, ‘the real Holmes’? This seems a little trickier, especially if the fiction concerned were to later rely on developments in the relevant scene (although this is not the case with the example we are looking at). It is perhaps analogous to watching a film where an actor is replaced in one scene; the character isn’t really the same, but we can appreciate them as a sort of stand-in, so to speak. Ultimately, of course, this could lead to the downfall of a fiction as a successful work, but in itself this is not an observation which need trouble us, nor is it one we need do anything to explain away. More importantly for our purposes, it again reinforces the notion that we can have a better sense of a fictional world, at least sometimes, than the one who created it, and this in turn lends support to the idea that a work, though made by a specific agent or agents, can sensibly be appropriated to some extent by its audience – and so that an understanding reader, while she should note the author’s intentions, should also be interested in the interpretations of others. Hence compatibilist pluralism is not merely permissible, but advisable.

5.3 Closing Remarks:
This discussion in this chapter began by describing three core differences between understanding people and understanding fictional characters. The first of these was the fact of fictive distortion, that we often have asymmetrical nonfiction-responses and fiction-responses, either in terms of the emotions a thing evokes in us or the valence such emotions have. The second was that the standards by which we estimate and interpret things and people are far more malleable when we are engaging with the fictional. The third was fiction’s moral inefficacy, as represented by the principle of fictive externalism; fictional people, unlike their real-world counterparts, can never in themselves be the source of genuine moral claims. It was then shown how we can sensibly talk about being understanding toward inherently nonexistent people, and how we can think of ourselves as having insight into them. After this some work was done to explicate how the understanding agent should approach the optimal appreciation of artworks, and in particular how this relates to the extent to which we should try to understand fictional characters. Finally, this led us to consider how understanding demands that we approach the interpretation of art as such, although this was done in terms of minimal ethical criteria; room was left for additional determinations according to other, fuller theories of art specifically. We have throughout been guided by what understanding recommends, not only in terms of what we should do, but in terms of what we should think of ourselves as doing, and it has proved to be quite a fruitful endeavour. All that remains now is for us to look back over our entire project in order to take stock of what we set out to do and what we have accomplished.
The argument in this thesis has been primarily concerned with one thing: exploring the basic reality of what does and should go on when we interact as moral agents. It has been conducted in the light of a central concept, understanding, which we have filled out to the point of knowing what it involves, what it demands and what sorts of peripheral activities it recommends. We began by outlining just such a picture; in 1.1 it was established how understanding should be defined, examples were given to illustrate how it works and the kinds of decisions or assessments it would have us pursue and avoid, and its usefulness and relevance was defended against a number of opening attacks. In 1.2 we saw how identification, a large part of the overall concept of understanding, has been debated in the philosophy of art, which showed it to be a wide-spread and constructive idea, as well as laying the ground for my later discussion of aesthetic cognitivism and imaginative exercise. 1.3, then, saw a more precise explication of some of the psychological processes involved in understanding by looking at the debate surrounding simulation theory, a concept again with close ties to understanding and identification. This cemented the understanding approach with a basis of creditable empirical research, as well as demonstrating the important motivational potential of what I call thick understanding.

But while the first chapter showed how we should think of understanding, together with where and how its closest conceptual cousins perform in contemporary debate, the second chapter established that understanding can also be seen as part of an extended philosophical tradition. In 2.1 we looked at the work of Adam Smith, arguably a more ambitious precursor to simulation theory, which detailed an intricate picture of moral and social psychology with something very much like understanding at its core. Not quite satisfied with some of Smith’s empirical assumptions, though, and his relatively modest aims in terms of establishing positive normative determinations, we moved on in 2.2 to Nel Nodding’s ethics of care, which again turned on people’s ability to movingly enter into the positions of others. Finding her account to involve important infelicities, though, regarding requirements of reciprocation and arbitrary non-human allegiances, we then looked in 2.3 at Richard Hare’s ethics. Here we found a story on some level very close to our own about how interests or preferences should be weighed and acted upon, though again the need for change was established by pointing out some questionable elements to the philosophical picture driving his account; in particular it was found that his hierarchy of moral thought was, in realistic terms, quite misleading, and his credentials as a utilitarian were called into question.

After this 3.1 showed that, regardless of which currently popular normative theory one might subscribe to, one must make room for and employ the concept of understanding or something
equivalent to it. 3.2 then saw understanding put to work on the question of what makes for moral relevance, where it was found that understanding can be seen to lead to new and interesting results which are in many ways more appealing than those of the main normative theories of today. 3.3, in turn, thoroughly established the superiority of this approach to its closest contemporary cousin, preference utilitarianism, where both it and the previous section developed a highly plausible approach to our treatment of non-human animals. 3.4 continued in a similar vein, arguing for a moderate, constrained approach to aggregation and showing how a broader and more consciousness-orientated concept of autonomy can explain some of our otherwise puzzling moral intuitions.

After the groundwork of 1.2, 4.1 saw us move our primary focus away from ethics. Here the pursuit of understanding led us to first explain how and why we should see fiction and art as offering a means to a special and valuable kind of learning or enrichment, in particular by looking at the arguments of Martha Nussbaum, Jenefer Robinson, Peter Lamarque and Stein Haugom Olsen. Once the connection between understanding and cognitivism was established, and the usefulness of cognitivism defended, it was clear that understanding could be fruitfully applied to matters in the philosophy of art. Before rushing ahead, though, in 4.2 we did some work to show how exactly we should conceive the standards that cognitivists should look for, and to clear up some of the residual worries that a cognitivist or understanding aesthetician might have in the face of fiction and art’s loose connection to the truth.

Finally, in chapter five we set out to see what insights understanding might grant us when applied to art, fiction and fictional characters. In 5.1 we began by setting out three core differences between understanding as it is applied to real-life and to fiction; first, our emotional reactions are prone to differ both in kind and in valence, second, we have a much greater capacity for changing even our most fundamental assumptions when engaging with fiction, and third, fictional characters are in themselves unable to be the source of ethical constraints on us. This last difference was also expressed in the principle of fictive externalism, that the moral character of an imagining must be determined in reference to the non-fictional, which was then given an extended defence. Afterwards, in 5.2, we first allayed the worry that understanding might not after all be appropriately directed toward fictional characters, before explaining how we can coherently think of ourselves as gaining insight into inherently nonexistent creatures. The chapter then drew to a close by offering some remarks on how we should understand the optimal appreciation of art according to understanding, and by developing a pluralistic notion of interpretation supported by understanding and cognitivism.

The overall result of these arguments has been a wide-ranging and informative examination of the social and epistemic practicalities faced by any truly conscientious moral agent. None of us can be perfectly good or perfectly understanding, it seems, but through the pursuit of sensitivity and imaginativeness there is yet hope for our improvement.


Neill, Alex, “Empathy and (Film) Fiction”, in David Bordwell and Noël Carroll (eds.), *Post-Theory: Reconstructing Film Studies*, Madison: University of Wisconsin Press, 1996.


