Empirical psychology and its effect on natural moral realism

Version: Full Version

Citation: Lane, Ashley Alexander (2016) Empirical psychology and its effect on natural moral realism. MPhil thesis, Birkbeck, University of London.

Deposit guide

Contact: email

Empirical Psychology and its Effect on Natural Moral Realism

Ashley Alexander Lane

Submitted in partial fulfilment of the requirements for the
degree of Master in Philosophical Studies

Department of Philosophy
School of Social Sciences, History and Philosophy
Birkbeck College, University of London

December 2015

Table of Contents

<u>Abstract</u>

I argue that empirical research can have important implications for various metaethical theories. I take an area of recent psychological interest, moral judgement-making, and show how current psychological research affects Peter Railton's naturalist moral realism in particular.

I first outline Railton's naturalism and show that it is a viable metaethical theory that is amenable to psychological research. In particular, I outline Railton's conception of 'idealised agents' and their relationship to actual agents, which is important for his naturalism. I then consider G.E. Moore's Open Question Argument and Horgan and Timmons' Moral Twin Earth argument, which try to show that various forms of naturalism cannot work. I demonstrate that Railton's naturalism can avoid the arguments for reasons that lay it open to empirical resting.

I then consider Jonathan Haidt's psychological argument that we do not typically make moral judgements rationally, but through emotionally-based intuitions. As Railton's naturalism depends greatly on rationality, this might be a problem. However, Haidt does not show that we should make moral judgements non-rationally, nor that rationality is typically unimportant in moral judgement-making. Railton must take greater note of non-rational factors, but psychological research on intuitions may actually support his naturalism.

I next use R.J.R. Blair's work to show that psychopaths cannot make genuinely moral judgements. Railton implies that they can, but the empirical evidence is against him. It also shows that the psychopath's problem arises from an emotional deficit rather than from a rational one, and that we cannot make moral judgements without certain non-rational powers.

I argue that the psychopath thus creates problems for Railton's naturalism, because it raises difficulties for the relationship between actual agents and idealised agents. I end by looking at what other metaethical theories can profitably use empirical research.

Declaration

I confirm that the work presented in this thesis is my own and the work of other persons is appropriately acknowledged.

Signed:

Ashley Alexander Lane

Date:

Philosophers are becoming increasingly aware of how empirical research can be used to support arguments in moral philosophy. Typically such research helps with moral psychology, where even the most informal evidence can raise useful points. Take the Trolley Problem, which asks whether it is better to actively cause harm to avert a greater harm, or to refrain from acting and allow the greater harm to occur. Judith Jarvis Thomson discusses a variation of the problem: suppose that, in order to avert a greater harm, you must actively hurt someone because he has relied on your previous promise that you would not hurt him. May you do so? Jarvis Thomson thinks not, but when she asked others for their opinion she found that some people disagreed with her. Curiously, she says twice that she was 'greatly surprised' by this, which makes clear why moral psychology needs empirical research.[1] A statistical sample of one will tell you more about a particular person's psychology than the average person's psychology. A philosopher might not have the same intuitions that a layman would (or even that another philosopher would). The philosopher's answer may be more reliable, but if we want to discover how people generally act when faced with moral dilemmas then we cannot just ask what the philosophers would do.

So empirical evidence can help us understand how people react in moral situations. It might seem, however, that it can only help us understand what we *actually* do, not what we *should* do. Fortunately, doing something is not in itself justification for doing it, as empirical evidence can produce extremely disturbing results. One notorious experiment is the 'Millgram experiment', where subjects were told to administer electric shocks to someone in another room every time that person failed a memory test (The set-up was faked, but the subjects did not know this.).[2] The shocks became more painful with each failure. Even when the 'victim' was crying out, 26 of the 40 subjects continued to the highest voltage, 450v. The subjects did not enjoy doing this, becoming 'highly agitated' and showing 'extreme stress'. One subject was 'on the point of nervous collapse' and three had seizures. Yet they kept giving shocks. Nobody would want to take this as an example of ethical behaviour.

An easy way of separating our actual behaviour from morally ideal behaviour is to sever the link between what we do on one hand, and on the other hand what we should do (such as normative commands). It is summed up in the slogan 'No "ought" from "is"', which implies that facts about the

---

[1] Jarvis Thomson (1985), pp. 1411-1412.
[2] Millgram (1963).

world (including those about our psychologies and abilities) cannot entail any moral obligation. Neither will they tell you anything about the nature of morality.[3] Non-normative facts alone will never provide moral reasons for action or tell you what you should morally do.

Moral psychology has no such problem with using empirical evidence, and it need not only concern human behaviour. For example, both biologists and philosophers have worked extensively on altruism. The biologist Robert Trivers writes of altruism displayed in multiple species.[4] Richard Dawkins, an ethologist studying animal behaviour, argues that altruistic behaviour, whilst genuinely altruistic, developed because it nevertheless generally benefited the reproduction of the genes that promoted it. Furthermore, we have the biological and mental equipment now to ignore what promotes our genes' reproduction, so we can consciously act against their 'commands' for altruistic reasons.[5] We can also learn about our moral natures by looking at brain structure and chemistry. Simon Baron-Cohen reports that people with the MAOA-L gene cannot clear the hormone serotonin from their synapses quickly, which apparently makes them more aggressive than people with the MAOA-H gene.[6] Aggression can be morally problematic, raising questions of when exactly aggression is justified, and an agent with the MAOA-L gene may judge aggression as morally permissible in more situations (such as those of defence and punishment) than an agent with the MAOA-H gene would.

So moral psychology may benefit from at least three types of empirical research: study of human behaviour (as seen in the Millgram experiment), study of human biology (Baron-Cohen), and study of general animal behaviour and biology (Trivers and Dawkins). We can also use evidence from other sciences and the humanities, such as history and social anthropology.[7] Whilst using empirical research may be complex and difficult, the moral psychologist now has plenty of empirical tools to help her studies.

However, no 'ought' from 'is'; empirical research is apparently concerned with natural, non-normative facts about how people act, and not how people *should* act. Suppose that, as Jonathan Haidt argues, we generally make moral judgements based on our emotional states, rather than on rational, reasoned judgements.[8] How can that tell us anything about what we ought to do morally? Such findings do not by themselves seem to recommend or criticise making moral judgements

---

[3] The roots of this idea go back at least to David Hume, who thought that reason alone could not generate moral or non-moral motivations ((2000), section 2.3.3.4).
[4] Trivers (1971).
[5] Dawkins (2006), chapter 11.
[6] Baron-Cohen (2011), pp. 89-90.
[7] For example, Pinker (2011) uses both to argue that violence has decreased through history.
[8] Haidt (2001).

emotionally. Empirical research seems to give no help to metaethical questions. I will argue, though, that appearances are deceptive. If 'no "ought" from "is"' is interpreted as meaning that empirical facts about the world cannot tell us anything about the plausibility of certain metaethical theories, I believe it is wrong. I think that empirical research can have important metaethical implications.[9] For example, we could use the study of evolution of morality and moral behaviour to see whether evolutionary development can support or debunk a particular metaethical theory.[10] Alternatively, we could look at the study of modern human behaviour (such as through psychology and anthropology) and biology (such as through neurobiology). The two areas can easily overlap, but in this thesis I will concentrate on the second. I will argue that empirical discoveries about human behaviour can affect at least some metaethical theories, in particular natural moral realism.

Some metaethical theories have little, if any, use for empirical research. Analytical non-natural moral realism holds that moral properties exist and that one can define them solely by analytical or *a priori* means. Some forms of analytical non-natural realism, for example, can hold that moral properties have no important connection with natural properties. They thus cannot be affected by empirical research, such as from the natural and social sciences, because such research deals only with natural properties.[11] Other naturalisms can play out differently, though. We must differentiate between two forms of naturalism here. *Substantive* naturalism holds that the properties under discussion (whether moral or non-moral) are natural properties. The question is ontological – is property *x* natural? – and there may exist natural properties that we can never discover. Perhaps we will never have the practical means to discover a particular property; imagine that we could never find out the properties of the star currently furthest from the Earth, simply because we will always lack the technological means to do so, although such means could be developed. They just never are. Or perhaps we cannot even discover some natural properties in principle. There are simply no possible means to discover them. I take no stand on whether such properties actually exist, but a substantively natural theory could hold that they do. Analytical natural realism, which holds that moral properties are *a priori* natural properties, is a type of substantive naturalism. But it may not be a type of

---

[9] I am not the only person to think so. For example, Sinnott-Armstrong (2006) and Darwall, Gibbard and Railton ((1992), pp. 188-189) explicitly call for philosophers to pay more attention to empirical research when working on metaethics.
[10] An early defence of this approach is Ruse (1988). More recent work includes Joyce (2007), Street (2006) and Boniolo and De Anna (2006).
[11] Other forms of non-naturalism, analytical or not, do not have to claim this. Such non-naturalisms claim that moral properties are non-natural, but that they supervene on natural properties. Supervenience comes in various forms, but the basic idea is that if two states of affairs have exactly the same natural properties, they must have exactly the same moral properties. There can be no moral difference without a natural difference. One can argue that non-naturalists have difficulty defending supervenience claims (e.g. McPherson (2012)), but many non-naturalists do endorse some form of supervenience (e.g. Shafer-Landau (2003), p. 77). They may also thus argue that empirical research about natural properties can help us find out about moral properties.

methodological naturalism, which makes the epistemological claim that we can discover a property (or show that it does not exist) through natural, empirical means. Methodological naturalism is most obviously used in the natural sciences. For example, we can detect the radioactive properties of an object with a Geiger counter, and empirical research showed that polywater (an alleged new type of water) did not really exist.

Analytical natural realism shows that substantive naturalists need not be methodological naturalists. Neither must a methodological naturalist be a substantive naturalist, as one can argue that empirical evidence demonstrates that particular properties do not exist. Moral expressivists can be methodological moral naturalists, even though they hold that no moral properties exist. Peter Railton (a natural moral realist) and Allan Gibbard (an expressivist) agree that '[t]oo many philosophers and commentators – we do not exempt ourselves – have been content to invent their psychology or anthropology from scratch…'[12] Both moral realists and expressivists, they think, need more empirical work and less armchair philosophy.

We must keep substantive and methodological naturalism separate to show how empirical research can help with metaethical theories. Whilst I shall consider both naturalisms, my main concern is methodological naturalism. Empirical research need not assume that all properties are natural, but since it works with natural tools it must be methodologically naturalist. I will examine a natural moral realism developed by Peter Railton that is substantively and methodologically naturalist, which claims that moral properties are natural and that we can discover certain aspects of moral properties through natural means. Railton himself states that he is more concerned with methodological naturalism.[13] He implies that to show that a natural property exists, it is best to demonstrate it empirically. Methodological naturalism is used to support claims of substantive naturalism.

I will use Railton's natural moral realism as a 'test subject' to answer the 'central question' of this thesis, which is:

Psychological (and biological) studies have shed increasing light on how we make moral judgements. Can the results, even if only in principle, be used to support or undermine natural moral realism? If they can, does this have any wider implications for metaethics?

---

[12] Darwall, Gibbard and Railton (1992), pp. 188-189.
[13] Railton (1989), p 156.

To answer this, my argument will go roughly as follows:

1) Natural moral realism (which I shall call 'naturalism' from now on) is substantively natural.

2) But there are serious difficulties if it is not also methodologically natural.

3) Therefore, in order to avoid these difficulties, a substantive naturalism should also be a methodological naturalism.

4) I am concentrating on substantive naturalisms that hold that we can identify at least some moral properties.

5) So if a substantive naturalism is also methodologically natural, it must hold that we can discover these properties through empirical means.

6) We can discover facts about moral properties through empirical investigation, such as psychological work on how we actually make moral judgements. This will tell us something about how we recognise moral properties.

Steps (2) to (5) will be discussed at greater length in chapter two, whilst step (6) will take up the rest of the thesis. Before I start my argument, however, there are certain clarifications I can make to ensure that the aim of my thesis is not misunderstood. I will discuss such points here.

1) *What effect could empirical research have on naturalism?*

The central question makes no assumption that empirical research will be useful. It might be no help at all. However, I am more optimistic, and even if I am wrong I hope to show that we should take the question seriously. If I am right, there are three possible outcomes.

a) The research could provide support for naturalism. Naturalism would thus become a stronger metaethical theory, assuming that the research does not equally support another theory.

b) Empirical research conflicts with naturalism, but the conflict can be resolved by altering the naturalism. The naturalist theory that best fits the empirical evidence would be (all other things being equal) preferable to the others.

c) Empirical research conflicts irreparably with naturalism. This would be a strong blow against the naturalist.

In cases of conflict the problem may not always lie with naturalism. The empirical research might be wrong instead. However, the naturalist cannot just assume it is to get out of trouble. There must be good reasons to think that the research is at fault.

2) *Are psychological studies too crude to find out about moral properties?*

Psychologists can ask thousands of people about their moral standards, but those standards may be ill-thought-out, irrational, contradictory, hypocritical or simply wrong. People can express moral views which, when put to the test, they abandon. One can hardly demonstrate that moral properties exist with such research.

I will not use psychological research in this way. I am more interested in how people make moral judgements than with what those judgements are. What is important is the mechanism rather than the result. Railton and most other moral naturalists say that we can know what moral properties are, so we must have some means to do so. Psychological research can help discover what those means are.

3) *You are examining how we can know what moral properties are. Are you investigating a question of moral epistemology?*

Broadly speaking, yes. To answer the central question, I will be concentrating on how we make moral judgements. In order to make correct moral judgements, Railton holds, we need to know various facts and natural properties that are also moral properties. The question is whether we can make correct moral judgements by using his metaethical theory. This is a methodological and therefore epistemic matter. However, it could be that once we learn more about how we recognise or discover moral properties, we learn more about what sort of thing they are, given the way we discover them.

4) *Psychological research only looks at laymen, but laymen are not moral experts. You can only find out about moral properties if you are a moral expert who is not misled as easily as laymen are.*

This is unduly pessimistic. First, it implies that an incorrect moral judgement indicates a failure in the judgement-making method, when it could result from another reason. Maybe the agent does not have all the relevant information, or has incorrect information. Second, it is too dismissive of laymen. A moral expert (if any exist) will make fewer mistakes, but surely some moral questions are easy for most people. When Mr Smith the butcher claims that sexual abuse is vile, there is no obvious reason to think that he has bad reasons for his claim. He is not simply parroting someone else's view. Similarly, he may be able to solve various algebraic problems without being able to prove Gödel's incompleteness theorems. For some mathematical problems, he does not need to be an expert.

5) *You are looking in the wrong direction. Empirical investigation is important, but you should be looking at other kinds of research and not psychology.*

This objection concedes a lot, as it agrees that empirical research is useful for metaethics. The problem, it says, is not with my basic claim but with my particular emphasis. I am happy to say that research from any empirical area might be helpful, from history and sociology to the hard sciences. Each claim should be judged on its own merits, and we should not write off entire areas automatically. However, if any empirical area helps, it is psychology. Ethics is primarily concerned with human needs and relationships. As Railton links moral properties to our psychological and biological characteristics, it would seem that one way we could get to know these properties is by looking at what psychological experiments reveal about how we respond morally to various situations.

6) *Moral properties are not validated by empirical evidence. If Alan goes to jail for stealing – an empirical outcome – it is justified by our legal system but it does not justify the legal system itself. Similarly, we can agree that X being good is an empirical discovery without it justifying*

the definition we use of 'good'. We can agree, say, that charity is good given a utilitarian framework, but that does not justify the framework.[14]

This objection claims that empirical results may help our normative ethics (what acts are good within our moral framework?) but they cannot help justify our metaethical theories. However, I do not intend to provide a wholesale justification of a particular naturalism. The central question is based on the idea that we must be able to recognise moral properties if we can follow them. This is not the same as justifying such properties, though they are related questions. The ability to recognise moral properties helps to some extent to demonstrate the truth of moral realism. Neither is it only an issue for the moral realist. Expressivism may also depend somewhat on how we can and do react in moral situations, though that alone will not justify the expressivist theory. The central question does not assume that empirical evidence alone can justify a metaethical theory, only that it might support or damage particular theories.

7) *Why choose empirical research as a way of defending or attacking Railton's naturalism?*

This question misunderstands my aim. I am most concerned with how empirical research affects Railton's naturalism, rather than whether his naturalism is the best metaethical theory or not. It is not the only naturalism I could have chosen as a 'test subject'. Richard Boyd is another realist with a fondness for methodological naturalism, and he does not rely on analytical definitions of moral properties.[15] Boyd claims that moral goodness is made up of a cluster of 'important human goods, things which satisfy important needs'. Determining what these needs are is 'a potentially difficult and complex empirical question'.[16] This is another realism based at least partially on empirical investigation.

'Knowledge of fundamental human goods… represents basic knowledge about human psychological and social potential. Much of this knowledge is genuinely *experimental* knowledge and the relevant experiments are ('natural' occurring) political and social experiments…'[17]

---

[14] Copp (1990).
[15] Boyd (1988), p 199.
[16] Boyd (1988), p 203.
[17] Boyd (1988), p 205, emphasis in the original.

Boyd claims that we find out what is good in the same way that we find out about scientific properties, through observation. 'Goodness is a property quite similar to the other properties studied by psychologists, historians and social scientists...'[18] There are various ways that we could empirically investigate Boyd's realism. For example, he claims the moral intuitions can be reliable because they are 'a species of trained judgment'.[19] Really? What actually happens when we use moral intuitions? How reliable are they, under what circumstances? Jonathan Haidt uses empirical work to argue that most moral intuitions are emotional reactions that we cannot rationally justify.[20] How could Boyd respond? What happens when moral intuitions conflict, as happens in the various Trolley Problems?

Boyd's realism would have done just as well as a 'test subject' as Railton's naturalism. Alternatively, I could have avoided moral realism and chosen a form of expressivism to test instead. I will discuss in chapter two why I chose Railton's naturalism, but I have no particular aim to attack or defend it.

8) *Could your project have any bearing on normative ethics?*

Metaethics asks questions such as 'do moral properties exist?' and 'what do moral terms mean?' It will not tell you what you should morally do; it will not generate normative commands. That is the business of normative ethics. There is no necessary link between metaethical theories and normative theories. For example, a moral realist and an expressivist will disagree over the existence of moral facts, but they may both agree that one should promote the greatest good of the greatest number. However, at least sometimes a metaethical theory will have normative implications. Railton's metaethics, for example, gives great weight to the well-being of society's members, which would appear to be more favourable to consequentialists. A.J. Ayer argued that moral statements are nothing more than expressions of emotion, and thus could not have any truth value. Therefore the statement 'You have a duty to honour your parents' is nothing more than an expression of approval towards honouring your parents.[21] Duties do not actually exist, so Ayer's emotivism implies that strictly speaking deontological theories are wrong.

---

[18] Boyd (1988), p 206.
[19] Boyd (1988), p 207.
[20] Haidt (2001).
[21] Ayer (1936), chapter 6.

So if a metaethics is affected by empirical psychology, and has an influence on normative ethics, then the psychology might indirectly affect the normative ethics. However, psychology may affect normative ethics more directly. Whilst I will not say much concerning normative ethics, I should say that philosophers have been using empirical psychology to criticise or endorse normative theories. Joshua Greene, for example, has used psychological arguments to argue that utilitarians are more likely to use reason in moral judgement-making than deontologists, which he believes is in the utilitarian's favour.[22]

To answer the central question, I have divided the thesis into four parts.

In chapter two I outline Railton's naturalism and defend it against two arguments that target various forms of naturalism, G.E. Moore's Open Question Argument and the Moral Twin Earth argument. Railton's naturalism requires the concept of an 'idealised agent', or what an actual agent would be if he had full cognitive and imaginative powers, and full information. An idealised agent can identify the actual agent's objective interests, which are what would be non-morally good for the agent. An action is morally right if people in general within a social group would impartially allow it, given full information about the action and its effect on the well-being of the individuals within the social group.

Cognitive powers can be used to reach moral truths, according to Railton. If we are cognitively able enough and have sufficient information, then we can recognise what is morally right. The question is how far we can really do this. My argument is that at least one set of experiments shows that Railton's naturalism faces problems that require Railton to amend it significantly. This is because Railton pays insufficient attention to the use of emotional powers in moral judgement-making. I take cognitive powers here to be powers of rationality. They do not include emotional powers. I explain why I do this in chapter two. First, it matches how Railton himself uses the term 'cognitive' in his writings. Second, if we call emotional powers cognitive from the outset, this gives him fewer ways to solve the problems I discuss for his naturalism at the end of the thesis.

In chapter three I discuss one set of experiments that one could argue affects Railton's naturalism, but actually fails to do so. These experiments relate to Jonathan Haidt's social intuitionism, which claims that we most often make moral judgements through emotional intuitions made on the spur of

---

[22] Greene (2007).

the moment, and afterwards we rationalise our judgements to convince ourselves and others that we have good reasons for our judgements. We find it hard to make moral judgements using our cognitive powers rather than our emotional instincts, and so Railton's naturalism does not actually reflect what happens in real life. I argue that Haidt's experiments do not justify these conclusions, as Railton has at least three responses. First, even if an act we think is morally wrong is actually harmless, we may still have an objective interest to stop the act. We may feel extreme emotional dismay at a particular action and then rationally judge that the action is morally wrong because of that dismay. Second, Railton can provide an account of intuitions that is congenial to his naturalism. Third, Haidt's experiments concentrate on moral situations which are so unusual and rare that we cannot draw any conclusions from them that would damage Railton's naturalism.

Afterwards, I take the thought that objective interests need not have a rational basis and apply it to our biological nature. I look at the hormone oxytocin, which promotes bonding between individuals. Mothers and children appear to have an objective interest to be close to each other, and this is at least partially due to non-rational, hormonal factors. Furthermore, oxytocin can give rise to objective interests that are morally right, since a lack of oxytocin in mothers can make them neglect their children, a moral wrong. Railton's naturalism can accommodate this by saying that objective interests can be generated non-rationally, and one can rationally judge that such interests can be moral as well.

This does, however, raise the question of how far one can rely on rationality for moral judgements. In chapter four, I consider psychological experiments on psychopaths and argue that psychopaths cannot make genuinely moral judgements, as they are unable to distinguish between moral transgressions (such as a school pupil hitting another pupil) and conventional transgressions (such as a pupil leaving the classroom without the teacher's permission). I then argue that the psychopath's inability to make moral judgements is unlikely to be because of the psychopath's lack of empathy or a failure of rationality. Instead, Psychopaths are unable to feel certain emotions (such as guilt and remorse). They are also unable to infer it in others, meaning that they cannot make moral judgements simply by copying what other agents do.

In chapter five, I argue that this this raises difficulties for Railton's naturalism. The idealised agent is supposed to be able to identify objective interests, which helps in turn to identify what is morally right and wrong. However, if the idealised agent is only cognitively (i.e., rationally) superior to the

actual agent, it would seem that idealised psychopaths cannot make moral judgements, because the psychopath's problem is not cognitive. Railton must decide whether a psychopath's idealised self can make moral judgements or not, but neither choice is unproblematic. If the idealised agent *can* make moral judgements, then it appears that there is no guarantee that he can identify the actual agent's objective interests, which Railton stipulates he must be able to do. If the idealised agent cannot make moral judgements, then Railton must accept that an agent cannot make correct moral judgements simply by virtue of being an idealised agent in his sense, and it also puts limitations on the powers that an idealised agent can have.

We start in chapter two with an outline of Railton's naturalism.

<u>Chapter 2</u>

In this chapter I will outline Railton's naturalism in enough detail to see how it can be affected by empirical research. I will not specifically address his naturalism's particular strengths and weaknesses, but I will discuss it in sufficient detail so we can see what sort of empirical research we can use to test it. I will show that it is a viable metaethical theory because it can defuse two major objections to moral naturalism in general. The first is G.E. Moore's Open Question Argument ('OQA').[1] The OQA purports to show that moral properties cannot be natural properties, but it fails to target naturalisms like Railton's. The second objection is Horgan and Timmons's Moral Twin Earth argument.[2] Whilst this argument does aim at naturalisms similar to Railton's, it is ultimately no more successful than the OQA. Having addressed the two objections, we will then be able to 'test' Railton's naturalism with empirical research in the next two chapters.

Peter Railton's naturalism claims that moral properties exist, moral judgements can be true, moral properties can be reduced to natural properties and that these properties can be empirically investigated.[3] He also writes early in his naturalism's development that moral properties are objective, but as we shall see, he later wavers on this.[4] There is already a lot to say about this naturalism before we get to the details, but I will only make three points. First, Railton openly invites empirical investigation by claiming that 'moral enquiry is of a piece with empirical enquiry'.[5] He declares himself to be a methodological and substantive naturalist, so he cannot complain if we use empirical research to investigate his theory.

Second, Railton admits that his naturalism contains a significant subjective element. He notes that realist theories are typically associated with the claim that moral facts exist independently of our opinion of them.[6] However, Railton links moral properties tightly with our psychologies. He states that 'no one kind of life is likely to be appropriate for all individuals and no one set of norms appropriate for

---

[1] Moore (1993).
[2] Horgan and Timmons (1992), (2000).
[3] The first two points help distinguish moral realism from expressivism. Moral realism holds that moral properties exist and therefore we can make statements about them that can be true or false. Expressivists claim that moral properties do not exist. Moral statements are expressions of attitudes. Even if such statements can be true, it does not depend on the existence of moral properties.
[4] Railton (1986a), p 165.
[5] Railton (1986a), p 165.
[6] Railton (1996), p 56-57.

all societies and all times'.[7] His realism is therefore not objective in the sense that it is not absolutist. An absolutist moral realism would hold that the exact same moral norms that apply to one agent apply to all agents without exception. Railton thinks differently; if our psychological makeup were different, moral properties would be as well. We cannot simply base moral properties on our opinions, though. One cannot declare that murder is morally wrong by personal fiat. Railton's realism is therefore carefully balanced. There are no categorical moral imperatives to do *x*, because our psychological states are contingent.[8] Moral properties can change (so different natural properties become moral properties) if our psychological and biological make-up changes, and no moral properties exist regardless of that make-up. However one can be morally right or wrong, because one can discover as a fact what natural properties are moral properties.[9] This depends on society's needs as a whole, and one cannot choose what one needs arbitrarily. It can be entirely outside one's control. This means that there can be moral progress; discovering which natural properties are moral can help us develop better ethical systems. A society can become morally better or worse than it previously was.[10]

Third, we need a definition of 'natural property'. It is easy to give examples of natural properties, such as having a particular mass. Mass is a feature of the external physical world and it can be measured empirically. Being wooden and being wet are other natural properties. Non-natural properties, if any exist, are just those properties that are not natural. But whilst we can point to many natural properties, defining 'natural property' is harder, and Railton does not attempt it. G.E. Moore suggests that a natural property is 'a property with which it is the business of the natural sciences or Psychology to deal, or which can be completely defined in terms of such'.[11] But what are the natural sciences supposed to deal with? Moore states that natural properties 'include all that has existed, does exist, or will exist in time', but provides no argument that the natural sciences and psychology deal with all such properties.[12]

We can interpret Moore's definition in various ways. Thomas Baldwin suggests that natural properties are causal.[13] Alexander Miller writes that natural properties are either causal or such that our senses can detect them.[14] It is a complex question whether all and only natural properties are causal, so the second criterion is important. It does not mean that radio waves are non-natural

---

[7] Railton (1986a), p 165.
[8] Railton (1986a), p 201.
[9] Railton (1986a), p 191.
[10] Railton (1986a), p 194.
[11] Moore (1993), p 13.
[12] Moore (1993), p 92.
[13] Baldwin (1993), p xxii.
[14] Miller (2013), p 10.

because we cannot sense them, so a clearer definition would be that natural properties are properties that are causal or that can in principle be detected through empirical means, resulting in a form that our senses can detect. We cannot sense radio waves, but we can detect them with equipment that gives visual or auditory evidence. The words 'in principle' are important. There are certainly properties in the universe that nobody will ever discover. These properties are natural if we could empirically detect them given the opportunity.

Another reason to prefer this definition of 'natural property' to Moore's definition is that it does not restrict natural properties to only those examined by the natural sciences or psychology. However, this is not too important because the empirical studies we will look at later are mainly psychological. Moore's definition, whilst imperfect, would still work. Neither does adopting Miller's definition allow us to escape the OQA, as we shall see, so we cannot bypass it by covertly cherry-picking our definition.

Moore thought that defining moral properties was impossible, and so left them exactly as they are. Railton's realism, though, is revisionist, claiming that moral properties can in some way be reduced to (and defined by) natural properties, which changes how we think about them in some way.[15] One may worry that this could end up revising moral properties out of existence, which Railton calls eliminative reductionism. Perhaps moral properties are so different from what we thought they were that we cannot really say that they exist. An example of eliminative reductionism is polywater, a form of water apparently discovered in the 1960's. By 1973 it was clear that polywater was only normal water contaminated with impurities. The investigation of polywater reduced it to impure water – 'polywater' did not exist.

Railton believes that his realism is not a form of eliminative reductionism, but of vindicative reductionism.[16] Though his realism describes moral properties in natural terms, we will still be justified in using moral terminology. Moral properties will remain 'undamaged' and not merely be explained away. In the same way, though we can define water as H_2O, reducing it to its chemical constituents, we do not think that water does not really exist.[17]

Any comparison of moral properties with natural properties appears to breakdown at an obvious point – natural terms such as 'water' do not carry any normative force. But Railton is not troubled. He

---

[15] Railton's naturalism is thus reductionist because it links moral properties with specific non-moral natural properties. Non-reductionist naturalists claim that moral properties are natural but cannot be reduced to non-moral properties.
[16] Railton (1989), p 161.
[17] Railton (1989), p 157.

uses as an example the property of seaworthiness, which he believes is a normative property.[18] The property of being seaworthy is not shown to be illusory just because we can point to a group of natural properties that combine to make a boat seaworthy. If you tell a seafaring community that a boat is not seaworthy, you are not only telling them about its physical condition but warning them not to use it. Admittedly this will only hold if the community wishes to use the boat for sailing. If they want to break it up for scrap its ability to sail well is unimportant to them, so the property has no normative force. Surely, though, we want moral properties to retain their normative force even when people have no interest in them. To overcome this problem, we must consider Railton's objectivity again. Railton's moral properties are not categorical and depend (at least in part) on our psychological dispositions. His realism cannot be absolutist, as moral properties can change along with our psychologies. How, then, can Railton retain objective moral properties? The answer is by limiting what is subjective.

'In a universe without subjectivity, there is not value either. But all actual subjective beings are at the same time objective beings. They have determinate properties that are not merely constituted by their conception of themselves, and these properties determine what sorts of things do, or can, matter to them. Their self-conception may be more or less objective, that is, may more or less accurately reflect what they are really like, how they are actually situated and so on… [T]hat which is genuinely valuable is constituted by what matters to subjective beings whose conception of themselves and their world is in this sense objective.'[19]

If we are to have values, it depends on having properties that actually matter to us. The values that we have are subjective to that extent. But we cannot choose on a whim which properties matter. We can be completely wrong about the values that we hold. We need accurate self-knowledge to discern our real values. In this sense Railton's naturalism is objective and mind-independent. It requires us to have values (which you need a mind to have) but we have particular values even if we do not think we have them.[20]

Railton is not only concerned with specifically moral properties and values. He starts his metaethical theory by first formulating a theory of non-moral good, before proceeding to moral rightness. I will take the same route to show how he constructs his realism.

---

[18] Railton (1989), pp. 163-166.
[19] Railton (1986b), p 56.
[20] Railton (1986a), p 172.

Call the group of an agent's conscious and unconscious desires and wants his subjective interests.[21] These interests might not be good for and may even harm him, and he might change them if he had more information. So instead of subjective interests we need objectified subjective interests. Imagine an agent, A, and an idealised agent, A+, who is what A would be if A had 'unqualified cognitive and imaginative powers, and full factual and nomological information about his physical and psychological constitution, capacities, circumstances, history and so on'.[22] A+ would use his knowledge of A's actual circumstances to work out what he (A+) would want if he were A. Take Lonnie, who feels ill and craves milk.[23] Lonnie does not know that he is dehydrated, and milk would only make it worse. Lonnie+, however, would know because of his perfect knowledge. Lonnie+ would reason that Lonnie should desire clear fluids instead. Lonnie thus has an objectified interest to drink clear fluids, which can be determined by factual information about his biological condition, the causes of dehydration, the properties of clear fluids and so on. These facts are the reduction basis of the interest, the basis upon which Lonnie+ would ascertain what desires Lonnie should have.

Consequently, what justify our values are facts which do not depend on our desires.[24] These facts affect our desires and interests, and this is how we defend our values; we do not justify our values to others just by appealing to our values. Lonnie+ would use facts about Lonnie's situation (and not Lonnie's desires) to ascertain what Lonnie's best interests are, to get well. Similarly, consider Railton's example of Beth, an accountant who wants to become a writer. Beth reasonably but incorrectly believes that she could have a successful writing career. However, a more informed Beth would know that she actually does not have the skill required. The more informed Beth would know more facts about Beth's psychology and situation. The facts explain why Beth would not be a successful writer and why a writer's life would not be good for her.[25] If Beth drops her desire to become a writer, she will do so because of these facts. The facts influence the desire.

Railton now introduces a criterion of non-moral goodness by using 'objective interests'. Lonnie has an objective interest to drink clear fluids because of various facts about his biology and psychology, dehydration, the properties of clear fluids and so on. His biology could have been such that clear fluids would not help, but since his biology is what it is, he cannot simply pick and choose what would be good for him. Given this, we can say the following:

---

[21] Railton (1986a), p 173.
[22] Railton (1986a), pp. 173-174.
[23] Railton (1986a), pp. 174-175.
[24] Railton (1986b), p 53.
[25] Railton (1986b), pp. 50-51, 63.

X is non-morally good for A iff X would satisfy an objective interest for A.[26]

This shows why Railton cannot be an absolutist. My non-moral good may not be yours because we have completely different objective interests. Railton even denies that there are particular objective interests that everyone necessarily has.[27] Maybe we all share some objective interests, but we could only discover that *a posteriori*. Nevertheless, we cannot arbitrarily decide what our objective interests are. Lonnie cannot make milk a cure for dehydration, no matter what he does. Also note that the criterion links non-moral good to motivation.[28] If we found out that X prevented us from satisfying our objective interests, we would lose the desire and motivation to obtain X.

Objective interests are the interests that A+ would want A to have. '[W]hat is intrinsically valuable for a person must have a connection with what he would find in some degree compelling or attractive, at least if he were rational and aware'.[29] But what values would A+ want A to have? Railton's examples all use the idea that we want our lives to go well. He occasionally confirms this explicitly.

'Generally, we can expect that what A+ would want to want were he in A's place will correlate well with what would permit A to experience physical or psychological well-being or to escape physical or psychological ill-being. Surely our well- or ill-being are among the things that matter to us most…'[30]

Most of us want our lives to go well. A+ knows that A's ultimate aim is (usually) to have well-being, something that can be a goal in its own right. Maybe some agents have objective interests that involve being miserable, but Railton has already conceded as much by rejecting absolutism. He therefore cannot insist on everyone having the same objective interests of well-being, even if most people do have it.

It is important at this stage to be clear about what cognitive powers are supposed to include. Railton writes that cognitive powers include 'thinking, reasoning, inference, and so on', but he does

---

[26] Railton (1986a), pp. 175-176.
[27] Railton (1986b), p 47-48.
[28] Railton (1986a), p 177.
[29] Railton (1986b), pp. 47 and 51.
[30] Railton (1986a), pp. 179-180.

not specifically define them.[31] One might take 'cognitive powers' to just mean whatever powers we employ in order to know things. 'Cognitive powers' may be left fairly vague because they are whatever mental powers an agent requires to discover what is an objective interest, and that is a factual matter. Cognitive powers can thus include rational powers and what I shall call emotional powers, by which I mean powers to respond emotionally. It might be – I shall not prejudge at this point – that we can know things using emotional powers, and so they would be cognitive according to the proposed definition. Rational powers certainly would be cognitive.

However, this analysis is slightly too broad. I will instead take cognitive powers to be rational powers for two reasons. First, the examples of judgements that Railton uses elsewhere (for example, his examples of Lonnie and Beth) indicate that Railton is thinking of powers of rationality. We can rationally make judgements about objective interests and moral problems like we rationally make judgements about logic, mathematics and scientific questions. When Railton talks of instrumental rationality, which instructs agents to take appropriate means to reach their goals, he talks of estimating probabilities, using deductive inferences and surveying alternative paths of action.[32] Furthermore, he recently wrote (regarding the use of empathy in making moral judgements) that '[p]sychopathy is thought by some to involve a… perhaps specifically emotional rather than cognitive or motor empathy'.[33] This indicates that he does not see emotional powers as cognitive powers.

Second, I am defining cognitive powers as rational powers for reasons that will become clear in chapter five. In chapter five I argue that empirical psychology presents a problem for Railton, and that he must choose one of two alternatives. Both have problems, but if we define emotional powers as cognitive from the beginning, then we force Railton to take arguably the worst option. I wish to give his naturalism as much breathing space as possible.

We should also be clear about the relationship between rationality and reasoning. In the research and arguments I examine, we will find that the writers I discuss, such as Jonathan Haidt and Shaun Nichols, tend to use 'rationality' and 'reasoning' more or less interchangeably, emphasising our ability to reason (as we do in making logical and scientific judgements).[34] Railton also seems to run reasoning and rationality together.[35] Strictly speaking, we should not conflate the two. Robert Audi points out that theoretical rationality (that is, the rationality we use to judge what to believe) has at

---

[31] Railton (1996), p 52.
[32] Railton (1986a), p 186.
[33] Railton (2014), p 844.
[34] Haidt (2001), Nichols (2002).
[35] See, for example, Railton (1986a), pp. 166-167.

least four sources: reason, perception, introspection and memory.[36] This means that experience, such as remembered experiences or experiences acquired through perception, can give rise to rational judgements just as much as reason can. With this in mind, however, I shall talk of 'reasoning' and 'rationality' to keep the discussion simple. The important distinction for my purposes is not between the different sources of rationality, but between rational powers and emotional powers. The arguments I consider in this thesis look at how emotional powers relate to moral judgement-making, in contrast to rational powers.

Having obtained a criterion for non-moral good, how can Railton adapt it to get a criterion of moral rightness? So far we have A+ working out what A's objective interests are. A's objective interests might entail severely hurting another agent, B, and currently we have nothing to stop him. B+ will not be too happy, though, as B likely has no objective interest to get hurt by A. This leads quickly to Railton's criterion for moral rightness.

Railton's aim is to extend the objective interests of an individual to the objective interests of both the individual and the society around him. Moral norms are generated by rationally deliberating on what is in society's objective interests from a 'social point of view'.[37]

> "'X is wrong" means "We the people (i.e., people in general, including the speaker) would disapprove of allowing X as part of our basic scheme of social cooperation were we to consider the question with full information from a standpoint that considers the well-being of all affected without partiality."'[38]

Alexander Miller helpfully converts this into a criterion for moral rightness:

> X is morally right iff X would be approved of by an ideally instrumentally rational and fully informed agent considering the question 'How best to maximise the amount of non-moral goodness?' from a social point of view, in which the interests of all potentially affected individuals were counted equally.[39]

---

[36] Audit (2002), pp. 14-16.
[37] Railton (1986a), p 196.
[38] Railton (1996), p 69.
[39] Miller (2013), p 196.

Let us call something that is morally right a *social objective interest*. Naturally, Railton realises that there is much to discuss about the criterion, such as the terms 'full information', 'social point of view', 'without partiality' and so on. It is not my aim to fully criticise or defend his realism, though. At present it is enough to show that his naturalism is recognisably realist and amenable to empirical investigation. Railton has a list of properties that he believes any realism (moral or otherwise) must have. We can now see that his naturalism has them all.[40]

Railton's naturalism first entails that moral statements are truth-evaluable. If I say 'stealing is morally right', I claim that from a social point of view stealing would maximise non-moral objective interests, which can in principle be checked. Second, at least some moral statements are true, as some acts will be in the best non-moral interests of society's citizens. If all moral statements were false there would be no moral properties, and moral realism would be false.

Third, his realism entails that there are real moral properties. Moral statements make truth claims about moral properties, and since some moral statements are correct, these properties must exist. His realism is also a vindicative reduction, claiming that moral properties can be reduced to non-moral natural properties without becoming redundant.

Fourth, Railton's realism is mind-independent where it matters. Obviously it is not entirely mind-independent. If we had no mental states there could be no objective interests and hence no moral properties. However, his realism is mind-independent in the sense that we cannot change what is morally good merely by wishing it, any more than one can change the dream one had last night. A society's objective interests may shift over time, meaning that moral properties will change, but this is unimportant. Moral rightness is always connected to our objective interests, whatever they happen to be. We cannot make it so that moral rightness no longer tracks what our objective interests are.

Railton's naturalism is what he wants it to be, a natural moral realism. But any moral naturalism has to overcome a major objection, Moore's Open Question Argument. We now have enough detail about Railton's naturalism to see how it fares.

The OQA purports to show that defining a moral property *G* in terms of a non-moral property *N* will always fail because one could learn everything about *N* without confirming that the things that are *N*

---

[40] Railton (1996), pp. 52-58.

are also *G*. Moore's own exposition of the OQA is not entirely clear, though. He later claimed that his exposition confuses three types of definition.[41]

a) The predicate *G* is identified with a particular predicate *N*

b) The predicate *G* is identified with an analysable predicate *N*

c) The predicate *G* is identified with a natural or metaphysical predicate *N*

Moore thought that (b) and (c) are the most important types of definition and that they are easily confused. The modern, 'typical' exposition of the OQA does not confuse them, but it does use both, which provides a clue as to how the naturalist can escape it. Whilst the naturalist is concerned with defending (c), the OQA relies more on rejecting (b). If the naturalist can sever the link between (b) and (c), the OQA fails.

Moore outlines the OQA in sections 12 and 13 of *Principia Ethica*.

1) Assume that the moral predicate *G* is analytically equivalent (is defined by) a natural predicate *N*.

2) Therefore, when one claims that *x* is *N*, they must also mean that *x* is *G*.

3) If one makes such a claim, knowing the meaning of '*N*', and then asks 'But is *x G*?', they would be guilty of a conceptual confusion. Suppose I tell you that 'a bachelor' is analytically equivalent to 'a never married man'. If I then tell you that David is a bachelor and you reply 'Alright, but has he ever been married?', that would show that you have not understand what I have said. It is logically impossible to be one and not the other.

4) However, it is always an open question whether something is *G*. If we say 'is good' means (for example) 'is desirable', there would nevertheless be no conceptual confusion if someone asked 'Yes, charity is desirable, but is it good?'

5) There will always be a similar open question for any predicate we slot into the naturalist definition of *G*. So there is no *N* that is analytically equivalent to *G*, and *G* cannot be a natural property.

---

[41] Moore (1993), pp. 16-17.

The OQA uses both types of definition (b) and (c) by trying to match a moral predicate with a natural predicate and make them analytically equivalent. However, the naturalist can deny that her naturalism does this. Her naturalism uses empirical *a posteriori* naturalist discoveries, not logical analyses. Consider the discovery that water is H_2O. This was an empirical discovery, not one made *a priori*. It was not discovered analytically that water is H_2O, or simply stipulated as such (like we stipulate a bachelor to be a never-married man). Similarly, the naturalist can hold that we have to discover empirically what natural properties moral properties are. Railton's naturalism, for example, states that we have to find out empirically what the correct social objective interests are, and that they can change depending on empirical facts about us.

Open questions are therefore not a problem for the naturalist. Moral properties are not analytically defined as natural properties. We must empirically investigate whether a particular property *N* is also a particular property *G*. We may have to revise our metaethical theories afterwards, but that is how empirical discovery works. Open questions only matter if one insists on analytical (or non-parsimonious) definitions; if the existence of open questions is uncontroversial in science and other areas of empirical investigation, there is no reason for it to be controversial in metaethical empirical investigation.

So the naturalist already has the tools necessary to defuse the OQA. She already says that we depend on empirical research to help analyse moral properties as natural properties, and the OQA only affects analytical definitions.[42] Railton's naturalism can therefore avoid it. We need to discover empirically which moral properties are which natural properties. It depends on our desires and needs, our psychologies, the shape of society around us, and this is an empirical matter. Railton has no need of analytical definitions. This also means that he need not say that any particular natural property is necessarily any particular moral property. We find out empirically that natural property X is moral property Y, but it might not have been if our objective interests were different.

One could wonder whether Railton has not just pushed the problem back a stage. He says that no particular natural property is analytically identical with a moral property. But could one argue *a priori* that the property of being a social objective interest is analytically identical with the property of being morally right? If so, then Railton's naturalism falls prey to the OQA after all, and Railton does think that social objective interests are what is morally right. However, I do not think he involves analytical

---

[42] The analytical naturalist might try objecting that the OQA wrongly assumes that analytical equivalences must be obvious. See Joyce (2007), pp. 146-152.

identity here. He writes that he wants to 'provide an analysis of… ethics that permits us to see how the central evaluative functions of this domain could be carried out within existing (or prospective) empirical theories'.[43] This indicates that it is just an empirical fact that social objective interests are what is morally right, given how we actually think about morality. He further notes that he has no 'proof' to debunk moral scepticism, which seems to be an admission that he is not making any *a priori* or analytical claims about his naturalism.[44] So it makes sense to think that Railton only considers empirical analyses of moral properties.

This does leave a residual worry, though. Even if there is no necessary identification of water with H_2O, no informed person would seriously ask 'Is water H_2O?' We have too much evidence that it is. The question might not be conceptually confused, but it still looks confused in some way. In metaethics, though, we have no similarly confused questions. There are no identifications of natural properties with moral properties yet that we can accept without controversy, so perhaps we should assume that no such identifications work. No candidate identifications appear to work and it would take much work to show that they can.

The naturalist can reply that modern naturalist metaethics is a relatively recent development. It was only in the 1950s, some fifty years after the immensely influential *Principia Ethica* appeared, that naturalism began creeping back into metaethical debates.[45] Modern forms of naturalism developed even later. Railton's naturalism first appeared in the 1980s. The Cornell realists, who argue that moral properties are irreducibly natural, appeared around the same time.[46] Only in the last few years have philosophers started seriously using empirical research in ethical study. One cannot complain about only having preliminary results if it is too early for anything else. Admittedly this response will not hold good forever – if the naturalist has no uncontroversial results yet, she had better see about getting some. But at present this objection is not powerful enough to greatly damage naturalism.

Terence Horgan and Mark Timmons take another route to target forms of empirical naturalism, using Hilary Putnam's Twin Earth thought experiment. The experiment concerns Earth and twin Earth, two planets that are identical in every way except one. Whatever is H_2O on earth is actually XYZ on Twin Earth. Aside from being different chemical compositions, the liquids are identical in all other respects.

---

[43] Railton (1986a), p. 164.
[44] Railton (1986a), p 163.
[45] Darwall, Gibbard and Railton (1992), p 121.
[46] For example, Sturgeon (1985).

Twin-Earthlings even call XYZ 'water'. Putnam asks whether Susan on Earth means the same thing when she uses the term 'water' as Twin-Susan on Twin Earth does when she uses it. Putnam thinks not. Even if Susan and Twin-Susan are unaware of the chemical composition of the two liquids, when Susan talks of water she means H_2O, but Twin-Susan means XYZ. Putnam points out that Susan and Twin-Susan have identical mental states even though they mean different things. Therefore the meaning of 'water' does not depend on our psychological states because it can change even when our mental states do not.[47]

Consequently Susan and Twin-Susan cannot really argue over whether a certain liquid is water because they mean different things by the word. It is as if I talked of river banks and you talked of investment banks. Horgan and Timmons claim that if moral naturalism were correct, the same thing would happen with ethical arguments.[48] Suppose Earthlings and Twin-Earthlings used the same moral terms, such as 'good' and 'right'. However, on Earth the natural properties identified as moral are identified as such by following a consequentialist moral theory. For example, Earthlings may think that lying is morally wrong because it produces negative social consequences. On Twin Earth, the natural properties identified as moral are identified as such by following a deontological moral theory. For example, Twin-Earthlings may think that lying is morally wrong because lying violates the rights of those whom one lies to, and it would whatever the social consequences were.[49] Usually Earthlings and Twin-Earthlings would still agree on moral rules, but there would be disagreements. Or rather we want to say that there would be disagreements, but we cannot. Just as 'water' means something different when used by Earthlings and Twin-Earthlings, so would moral terms, as they rely for meaning on different natural properties on the different planets. Any moral argument would just involve talking past each other, as the Earthlings and Twin-Earthlings would mean different things.

Horgan and Timmons claim that whatever natural properties you use to get moral properties, they can generate a Twin Earth where different natural properties are used, so one must either accept that moral properties cannot be natural (as if they were, moral disagreement could not exist), admit that there is no moral disagreement (in which case an unacceptable relativism beckons) or make one's naturalism so vague that it cannot link moral properties to particular natural properties, which renders it useless.

---

[47] Putnam (1975), p 223 onwards.
[48] Horgan and Timmons (1992).
[49] Horgan and Timmons actually talk of moral properties supervening on natural properties rather than of moral properties being natural, but as they note it makes no practical difference ((1992), p 230).

The naturalist could argue that 'water' refers to both liquids in Putnam's thought experiment, and that anyone who identifies water as both H_2O and XYZ would be right. Railton's own response can be different, though. He can accept that Earthlings and Twin-Earthlings use moral terms in ways that relate to different natural properties; as he says, moral properties can change over time as societies change. Horgan and Timmons might object that this just makes his naturalism relativist, but this ignores an important part of Railton's naturalism. Consider why Earthlings and Twin-Earthlings have different ethical theories. Horgan and Timmons write that:

> 'The differences… are due at least in part to species-wide differences between psychological temperament that distinguish Twin Earthlings from Earthlings.'[50]

This is all that Railton, or any naturalist who relies on our psychological states, needs. As Earthlings and Twin-Earthlings are psychologically different, they will have different mental states and objective interests. Earthling-idealised agents and Twin-Earthling-idealised agents will reach different conclusions. Metaethical questions are settled using the same methods on both planets, but because our objective interests are different, moral properties will be identified with different natural properties. Furthermore, this is not relativist in Railton's sense. Earthlings and Twin-Earthlings use the same method to make moral judgements and cannot arbitrarily choose what their objective interests are. We can have moral arguments because an Earthling can correctly argue that something is not in the objective interests of a Twin-Earth society whilst Twin-Earthlings disagree. If the Twin-Earthlings started eating babies, the Earthling can argue that this is morally wrong because it fails the criteria for moral rightness in both the Earth and the Twin-Earth society.

We can check that moral terms are used legitimately on the two Earths by comparing their use to the use of 'water' on the Earths. Susan can use 'water' to refer to whatever is liquid H_2O, even though she does not know what 'H_2O' means. It is not necessary for her to know the full meaning of the word 'water' for her to refer to it successfully or for her to talk about it intelligibly. If she were to say sincerely 'water is transparent', that is good evidence that she knows that water is transparent.

Similarly, an agent can refer to moral properties without knowing anything about the content of moral properties. It would be perfectly intelligible for an alien studying Earth to report back to his

---

[50] Horgan and Timmons (1992), p 245.

planet 'Humans are aware of moral properties, though I do not know what these things are'. However, one needs to know something about the meaning of moral terms (though not everything) to make correct moral judgements. For Railton, one has to know how a judgement would affect social objective interests in order to reliably evaluate the judgement's moral rightness. This is the same for Susan and water, though. If Susan knew nothing about water she might be able to refer to it, but she could not make reliable judgements about it.

Consequently, Railton can freely choose whether to agree with Putnam that meanings are not given solely by psychological states. Two people can refer to different things called *x* even if they would give the same answer to the question 'what is *x*?' Susan and Twin-Susan are incompletely aware of the meanings of 'water' on Earth and Twin Earth, but they can still use the terms correctly in most situations. They can also use moral terms correctly in most situations even if they are incompletely aware of the different meanings of moral terms on the respective planets.

Horgan and Timmons ultimately have a problem. They cannot give up the psychological difference between Earthlings and Twin-Earthlings because they need it to make the moral differences they rely upon. However, this allows Railton to avoid their argument. The Moral Twin Earth argument is no more effective against Railton than the OQA.

Now that we have set out Railton's naturalism, we can see how empirical investigation can help support or damage it.

1) It is *a posteriori*. Nowhere does it depend on *a priori* analyses. Railton thinks that what is morally good will depend on what we (in an ideal world) rationally determine is in our collective objective interests, and that we generally associate our objective interests with our well-being.

2) These claims that be empirically tested. Most obviously we can examine a society to find out what its objective interests are and discuss if this is really what moral rightness requires.

3) Railton opens his realism up to psychological testing in at least two ways. First, we can discover whether we really do pursue social objective interests (so far as we can) with a specifically moral aim in mind. For one could do so simply because keeping society

functioning is best from a personal, egocentric, non-moral point of view. Second, we can discover whether we can make such judgements solely on rational grounds.

I shall focus on the second option. In the following two chapters I will consider two sets of experiments that purport to show that we do not or cannot make moral judgements as Railton describes, or at least that he has left a gap in his explanation of how to make moral judgements. Chapter three looks at Jonathan Haidt's social intuitionism, which claims that we typically make moral judgments using emotion, and not by using reason. I shall argue that this fails; Haidt's experimental evidence is not strong enough to back up his claims. Chapter four uses psychological research to argue that psychopathic agents are unable to make moral judgements, and that this is because of a defect in their emotional powers, not a defect of rationality. Chapter five argues that this means that Railton has to amend his naturalism to accommodate psychopathic agents. We will see where the gaps in his naturalism are, and how we can fill them in.

When we make moral judgements, do we rationally assess facts or rely on non-rational factors such as emotions? Moral realists can plump for any number of answers here – they may pick one option only, go for a blend of rational and non-rational factors or even reject both options and choose another method entirely. However, naturalists have commonly relied on rationality. If moral properties are like other natural properties, and moral statements can be true or false, it is tempting to treat moral problems like, say, scientific problems. Psychologists have also traditionally placed emphasis on rationality's role in moral judgement-making.[1] This suits Railton's naturalism. If all idealised agents have access to the same information and can reason perfectly, it becomes more likely that they can resolve moral problems.

Jonathan Haidt, though, claims that whilst we can make moral judgements rationally, typically we do not. We have emotionally-driven intuitions, such as that incest is horrible, and we rationalise the intuitions afterwards to justify them to ourselves and others (e.g. 'Incest is psychologically damaging'). However, since reasoning did not generate the judgement, we still want to follow our intuitions even if the reason is debunked. Therefore we should stop thinking of reasoning as providing a basis for morality.[2]

I will first look at why psychologists have traditionally endorsed moral rationalism. Then I will outline Haidt's claims and argue that they are not strong enough to refute the rationalist. I will then turn to neurobiological evidence that non-rational processes contribute to at least some moral judgement-making. Railton's naturalism can accommodate this, but we will end with the question of exactly how rational his naturalism can be.

Making moral judgements can be easy. We can judge as easily as our idealised selves can that, say, setting fire to cats for fun is wrong; our idealised selves here do not possess any cognitive powers or information needed to make that judgement that we do not possess ourselves. Many times individuals' objective interests can coincide and become a social objective interest, as when students have a collective interest in getting a good education. Getting sufficient information about a particular situation can be simple when the facts are few, obvious or readily available. And when answers are

---

[1] Kohlberg (1973) was particularly influential.
[2] Haidt treats 'rationality' and 'reason' as interchangeable (e.g. (2001), p. 815). I shall continue using the terms as I outlined in chapter 2.

difficult to find, Railton thinks that moral progress can occur when there is sufficient 'social rationality' to recognise the interests of neglected groups.[3] Examples include votes for women, anti-slavery laws and the legalisation of homosexual activity.

Not everyone makes moral judgements rationally, but this does not cause any direct problems for Railton. He does not think that actual agents never make moral judgements non-rationally, even though they should be made rationally. However, this can only go so far. For Railton's naturalism to be plausible, he has to allow that much of the time we make correct moral judgements rationally. Consider the abolition of slavery. For Railton, slavery's abolition was moral progress because people considered the objective interests of everyone involved and realised that slavery unjustifiably ignored the interests of the oppressed. If they judged that slavery was wrong because they tossed a coin then it would not be a moral judgement or genuine moral progress. It would be a happy accident.

So Railton must hold that most of the time we use reasoning in moral judgement-making, though we do not always do so. But whilst Railton speaks a lot about reasoning – as do other naturalists – he has given little attention to emotional and other non-rational processes that could contribute to moral judgement-making. He never says that there are no such processes, but he has only recently explicitly considered them, and it is not clear how they connect to his naturalism.[4] He is not alone; moral rationalism is common in both philosophy and psychology.

Lawrence Kohlberg, a psychologist who argued that 'individuals prefer the highest stage of reasoning they comprehend', might be called an 'arch-rationalist' who ignores non-rational processes entirely.[5] He influentially claimed that there are six stages of moral development, each stage building on, and therefore being morally better than, the previous stage. The first stage, reached by small children, holds that an action is morally good or bad depending on the direct physical consequences to oneself. An individual then goes up the scale as they become more rational. Stages five and six are reached by agents able to separate morality from the demands of the social group. Stage five has 'generally… utilitarian overtones… [depending on] standards that have been critically examined and agreed upon by the whole society'.[6] Stage six is more Kantian. The agent recognises abstract, consistent and universal moral principles. Kohlberg claims that 'both [these stages] aim at determining moral judgements… on which all rational men… can ideally agree'. Stage six, however, is rationally

---

[3] Railton (1986a), p 193.
[4] Railton (2014).
[5] Kohlberg (1973), p 633.
[6] Kohlberg (1973), p 632.

superior to stage five.[7] Kohlberg relates these stages to how we perceive rights and duties. Stage six requires all rights to generate corresponding duties, with the implication that agents who deny this do not reason as well as agents who accept it.

Railton's idealised agents are perfectly rational, so according to Kohlberg they should be at stage six. But Kohlberg explicitly assumes that moral judgement-making is almost entirely a matter of reasoning. Reason well enough and you can reach stage six too. Kohlberg only considers rationality, and Haidt notes that his successors have not done much better.[8] This produces some puzzling results; for example, Kohlberg interviewed three philosophers, two of whom were at stage five. The third was at stage six because his thinking was 'more… integrated than that of the other two philosophers'.[9] Since 'individuals prefer the highest stage of reasoning they comprehend', this means that the stage five philosophers could not reason as well as the stage six philosopher, since they could not really 'comprehend' him. This sounds bizarre – is Kohlberg claiming that, say, act-utilitarians do not really understand Kantian arguments? Is he correct that all rights have corresponding duties, or that rights even exist? If a stage five morality is the best morality, then a stage six morality cannot be ethically superior, even if it is more rational. And maybe it is not more rational; Joshua Greene has used Haidt's research to claim that consequentialists are more likely to make moral judgements rationally than Kantians.[10]

What are the alternatives to Kohlberg's philosophical claims? The first is that morality is based on rationality but he has not correctly described how moral development happens. I suspect that Railton would go for this option. The second is that people do not typically make moral judgements rationally, and we cannot automatically assume that this is a bad thing. I will now use Jonathan Haidt's research on moral judgement-making to see how resilient rationalism is. Before starting, though, we must remember that very few rationalists hold the extreme view that moral judgement-making requires reason and only reason. My aim is to see how far reason can be a useful moral tool at all.

Haidt endorses social intuitionism, which claims that although we can rationally reason our way to moral judgements, usually we do not. When confronted with a moral problem we form a non-rational intuition, which may simply be a knee-jerk reaction. Afterwards we use reason to justify our response.

---

[7] Kohlberg (1973), p 635.
[8] Haidt (2012), p 13.
[9] Kohlberg (1973), p 640.
[10] Greene (2007).

We do not reason our way to an answer, even if we think we do.[11] However, morality is supposed to be rationally justifiable. One cannot say 'Using the Union Jack to clean your toilet is morally wrong' without providing rational evidence. As Haidt found, we feel the need to justify our responses, which we think we can do best if we can give rational reasons.[12] By 'rational reasons', I mean reasons for our judgements that are generated using our cognitive powers (The relationship between cognitive powers and rationality remains the same as I discussed in chapter two.).

Haidt tested moral rationality by using two 'harmless, but disgusting' stories designed to draw out a conflict between reason and intuition.[13]

Incest: Julie and Mark, who are brother and sister are travelling together in France. They are both on vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but decide not to do it again. They keep that night as a special secret between them, which makes them feel even closer to each other. So what do you think about this? Was it wrong for them to have sex?

Cannibalism: Jennifer works in a medical school pathology lab as a research assistant. The lab prepares human cadavers that are used to teach medical students about anatomy. The cadavers come from people who have donated their bodies to science for research. One night Jennifer is leaving the lab when she sees a body that is going to be discarded the next day. Jennifer was a vegetarian for moral reasons. She thought it was wrong to kill animals for food. But then, when she saw a body about to be cremated, she thought it was irrational to waste perfectly edible meat. So she cut off a piece of flesh, took it home and cooked it. The person had recently died of a heart attack, and she cooked the meat thoroughly, so there was no risk of disease. Is there anything wrong with what she did?

For each story the test subjects gave an initial judgement, whereupon the experimenter would try to change their mind. The results were:

---

[11] Haidt (2001), pp. 820-822.
[12] Haidt (2012), pp. 27-30.
[13] Haidt et al (2000), p 7.

	Incest	*Cannibalism*
Initial judgement (yes/ok)	20%	13%
Final judgement (yes/ok)	32%	28%
Percentage who changed	17%	17%

The figures did not change much, so one would assume that the subjects had good rational reasons for their initial judgements. But this is not what often happened. 23 out of 30 subjects were explicitly 'morally dumbfounded' with *Incest* and 11 were with *Cannibalism*, declaring that the actions were wrong but 'unable to find the words to explain themselves'.[14] This does not show that we cannot make moral judgements rationally, since some subjects did change their minds after discussion. Haidt's point is that we usually do not, which fits other psychological experiments concerning how often we do not make non-moral judgements rationally.[15] If we rely heavily on non-rational processes when making non-moral judgements, we should not be surprised when it happens with our moral judgement-making.

One could argue that moral judgement-making is entirely unlike non-moral judgement-making, or that Haidt has not shown that moral dumbfounding is very widespread in society. If it is not, then it does not pose any real threat to rationalism. Let us grant him these, though, for the sake of argument. What does this mean for Railton? Just because we usually make moral judgements in a certain way, it does not mean that we *should* make them that way. But suppose an entire society unanimously objected to *Cannibalism*. A hard-core rationalist could insist that everyone was wrong because Jennifer's actions were harmless, but that would take considerable argument. Many people would be morally dumbfounded. Railton has suggested that intuitions are the result of unconscious reasoning, and moral dumbfounding results when one cannot consciously articulate those reasons.[16] I will examine this response a little later; for now, I want to look at an alternative response that he could combine with his actual response. The alternative response claims that it does not matter why people are ethically revolted by *Cannibalism*. The fact is that they do, so it is in the social objective interest to condemn it. Objective interests themselves need not be rational. Our hugely negative emotional

---

[14] Haidt et al (2000), p 12.
[15] E.g. Albright et al on how beauty influences our judgement (1988), Nisbett and Wilson on unconscious influences (1977) and Rozin et al on 'sympathetic magic' (1986).
[16] Railton (2014).

reactions might be enough to override any interest Jennifer has to eat human flesh. It is immaterial that the reason is emotional; intentional emotional damage should be avoided, so why not unintentional emotional damage? Most people would feel profoundly uncomfortable about living in a society where *Cannibalism* was acceptable.

So far this response favours the majority over Jennifer simply because it is the majority. Suppose most people were emotionally repulsed by the thought of homosexual behaviour. That is not a good reason to ban it. Fortunately Railton's idealised agents include those who would be severely hurt by a ban, so there would be no agreement amongst them that a ban would be morally correct. The idealised agents must still use reason to ensure that the disadvantaged do not lose out when the majority is emotionally weighted against them.

The problem goes deeper than this, though. Assume that Haidt is right. He also claims that philosophers more often use reason to make moral judgements than laypeople.[17] This means that whilst both Mr Smith the butcher and Mr Jones the philosopher think that rape is wrong, Mr Jones is far more rationally reliable. He has a reasoned ethical theory that demonstrates the immorality of rape, whilst Mr Smith 'only' thinks that rape is immoral because of his emotional reaction when he thinks of traumatised rape victims. Railton's idealised agents may make reasoned moral judgements, but according to Haidt most of us find this very difficult. Consequently we will not recognise moral properties as our idealised selves would. One could argue that this means that most people do not 'really' make moral judgements because only rational 'moral experts' can, but surely that goes too far. We need very good reasons to suppose that Mr Smith cannot really see that rape is immoral.

This does not mean that moral experts cannot exist. The danger is that we may write off the moral judgements of non-experts because their judgements rely more on non-rational processes. Mr Smith calls rape immoral because of his sorrow for the victims and his contempt for the rapists, and these emotional reactions look perfectly morally valid. We might need moral experts for difficult ethical problems, but Mr Smith can make at least some correct, genuinely moral judgements.

Emotions can be manipulated, though, and this is where reason can come in. Where reason and emotions violently conflict, Railton can claim that one should follow reason. A violent conflict would occur where the emotional response goes against Railton's moral rightness criterion. Nobody is dreadfully harmed if we condemn *Cannibalism* as nobody's important objective interests are

---

[17] Haidt (2001), pp. 819-820.

threatened, but they would be if we legalised rape. It does not follow, though, that one cannot use one's emotional reactions as a rough guide to identifying moral properties. In many cases they may be good enough. Consider again that an individual's objective interests can be virtually anything given the right circumstances. Suppose Albert is offered £10. It is at the end of a corridor. Along the wall is chained a pack of feral dogs who would attack if they could, but Albert knows definitely that the chains will hold. Albert has a phobia of dogs, though, and is terrified. Does he have an objective interest to accept the challenge?

Maybe Albert is so frightened that £10 would not be sufficient compensation. Maybe his objective interest is not to torment himself, and Railton would be quite happy with this. Since Albert+, the idealised agent, has 'full factual and nomological information about [Albert's] physical and psychological constitution', Albert's phobia already matters.[18] It is irrelevant that it is irrational. The emotional cost is higher than if he had no phobia, and Railton does not think that objective interests must be rational or that we should ignore non-rational effects on an agent when evaluating interests. Over time Albert+ may advise Albert to overcome his phobia, but right now such advice is too late.

Both individuals and groups can have non-rational goals. In the Mexican horror film *At Midnight I'll Take Your Soul* the protagonist is a profoundly vicious man. English-speaking audiences are usually puzzled by a scene in which he deliberately eats chicken during a Good Friday procession. Contemporary Mexican audiences would have understood immediately; by religious convention one should not eat meat on Fridays, much less Good Friday. The protagonist is deliberately being as offensive as possible, making people deeply uncomfortable for fun. He is trying to act immorally. There may be no rational reason for the prohibition of eating meat on Fridays, but many of us would think that the protagonist was acting unethically. This fits with Haidt's experiments, but Railton's naturalism can also accommodate it. Idealised agents can take account of objective interests even if the interests are not rational, whether for an individual or a group, with the caveat that such interests are not trumped by more rational interests.

Even if this is right, though, it does not quite address the central problem – Railton's naturalism still depends a lot on rationality, and Haidt claims that we do not typically use rationality. So naturalists must attack Haidt's central claim, and it turns out to be very shaky. Haidt thinks that we must use non-

---

[18] Railton (1986a), pp. 173-174.

rational means to reach moral conclusions, and this is how we *should* do it. But why accept this?[19] Morality should not be too difficult, but it should not be too easy either. Just because we may usually be bad at making moral judgements rationally, that does not automatically mean that a good morality will have little to do with rational thought. Emotions and intuitions can be fantastically bad tools for moral and non-moral judgement-making. Consider Rozin's experiments in which subjects were told to put labels reading 'sodium cyanide' on bottles of water.[20] After doing so, few subjects wanted to drink from the bottle that they themselves had labelled, even though they knew that the bottles contained water. This could be because we have to constantly make judgements in a multitude of situations every day, and we do not often have the luxury of consciously reasoning our way through all of them. We instead make rough helpful rules – for example, to avoid feral dogs, cannibalism, etc. Most of the time the rules work and we can rely on intuitions because they are based on reason, but occasionally the process breaks down, as Haidt and Rozin show.

The naturalist makes two claims here. First, we need reason to make moral and non-moral judgements, whatever Haidt claims, because some judgements are better justified than others, and those are the judgements we should follow. Second, Haidt ignores that both rational and non-rational processes can play a part in moral judgement-making. One way of developing this idea comes from Horgan and Timmons, who call their position morphological rationalism. They argue that moral judgements are generated by reason through following pre-existing moral principles. However, the process is often automatic and unconscious, contrary to Haidt's assumption that reasoning must be a conscious process. As morphological rationalism is an empirical hypothesis, we should be able to test it against Haidt's intuitionism.[21]

Horgan and Timmons start by talking of 'moral principles'. Moral principles tend to be general rather than specific (e.g. 'Cruelty is wrong' is more likely to be a moral principle than 'Cruelty towards Mr Smith is wrong'.). Some moral principles have no exceptions whilst others can be overridden, although Horgan and Timmons do not say by what. Neither do they define moral principles, although they give a shortlist of examples.[22] This unfortunately tells us little about how they are generated, or about which principles can be overridden and which cannot. For now, though, let us go on with their

---

[19] I am not the only one with such concerns. See Jost (2012).
[20] Rozin et al (1986).
[21] Horgan and Timmons (2007), p 280.
[22] Horgan and Timmons (2007), pp. 283-284.

argument. Horgan and Timmons claim that for a certain input to trigger an individual's moral judgement:

a) The trigger must non-accidentally and systematically conform to the principle held.

b) The conformity must result from the individual's psychology.

c) Typically the individual does not need a token representation of the principle to reach the judgement. He does not need a specific input situation and the relevant moral principle in his conscious thought at the time of making the judgement (i.e. he need not consciously think '*X* is happening and *X* conflicts with moral principle *Y*, so I will condemn *X*.').[23]

Is this plausible? Horgan and Timmons point out that morphological rationalism does not only apply to moral judgements. A professional violinist, when giving a performance, does not need to consciously remember how to play all the notes on her violin. The technical procedure of playing the violin has become 'thoroughly internalised', even if she consciously considers other aspects of her performance such as mood and tempo.[24] She can correctly say afterwards that she played using the technical skills she consciously learnt when starting to play the instrument. She would not be rationalising after the event the technical judgements she took. So if it is plausible that morphological rationalism holds for the violinist and other non-moral judgements, why can it not apply to moral ones?

Horgan and Timmons thus have an alternative to Haidt's intuitionism. People have moral intuitions that are based on unconscious reasoning and not simply emotion. Actually, Horgan and Timmons are happy for emotions to play a role in moral judgement-making processes so long as reason is also involved.[25] Another advantage of using reason rather than emotion, they claim, is that rational responses are better than emotional responses at reacting to moral situations. If our responses were just emotional, how could we ever know that we were acting appropriately? Suppose you have got a brilliant new job, but when you call a friend to share the good news you find her upset over marital problems. Morphological rationalism claims that you would unconsciously adjust your behaviour to act appropriately, based on what you know about your friend and similar situations. On the social

---

[23] Horgan and Timmons (2007), pp. 285-286.
[24] Horgan and Timmons (2007), p 286.
[25] Horgan and Timmons (2007), pp. 287-288.

intuitionist model, though, why favour one emotional response over another? Reason is better at providing the best answers.[26]

This is rather speculative. Much work must be done to show the morphological rationalism is true; even in this brief description I have noted some problems. However, morphological rationalism shows that there are live competitors to social intuitionism that should be taken seriously. The number of plausible competitors shows that Haidt's arguments are not as conclusive as he believes. We do not even have to reject Haidt's arguments completely. Jesse Prinz, for example, thinks that we have emotional responses to moral situations, but we need reason to categorise the situations correctly. We can rationally disagree over whether praying in school discriminates against atheists, but if we agree on the meaning of 'discrimination' we cannot have a rational argument over whether discrimination is morally permissible or not. Prinz's criticism of Haidt is not about the importance of emotional responses, but that Haidt only pays lip-service to Prinz's idea of rational moral disagreement.[27] Joshua Greene actually uses Haidt's experiments to argue that consequentialists are more likely to make rational moral judgements than deontologists. Deontology is (says Greene) more rooted in emotional responses, meaning that deontological defences of moral judgements need more 'fancy philosophizing' (meaning 'rationalisation') than consequentialist defences.[28]

The point is not that Greene (or Prinz or morphological rationalism or…) is correct, but that such arguments can be reasonably made. There is also empirical evidence against Haidt's philosophical conclusions. Using only Haidt's scenarios, such as *Incest*, Feinberg et al tested subjects' initial responses to the scenarios and subsequent moral judgements. They found that subjects who were more inclined to reassess their initial responses appeared to do so rationally. They took longer to reach a conclusion and were less likely to judge cases like *Incest* to be immoral. One subject said, regarding a story about a man who had sex with a dead chicken he had bought:

'At first I was overcome with disgust and horror. But as I read the question and considered, I started to see that it was the man's choice… The horror and disgust started to fade away.'[29]

---

[26] Horgan and Timmons (2007), pp. 290-292.
[27] Prinz ((2007), pp. 29-32, 124-125.
[28] Greene (2007), pp. 36-41, 55-58.
[29] Feinberg et al (2012), p 791.

Haidt admits that one can use reason to make moral judgements, but the more that people use reason to deliberate, the less evidence there is to think that our morality must be more closely tied to our emotions than to our reason. Morphological rationalism can explain the subject's change of mind by claiming that the subject's initial unconscious reasoning got it wrong. The situation was too novel and difficult for unconscious reasoning to work, as the subject had never considered it before. At this point conscious reflection took place, like an experienced violinist trying a new, difficult technique for the first time. Prinz would claim that whilst the subject's final reaction was emotional like his first was, it was based on rational consideration of the circumstances. And even if the subject's reasoning was not consequentialist, Greene can take heart in the fact that people may reassess moral judgements rationally.

In fact, the rationalist may say that we should all take heart in it. As Walter Sinnott-Armstrong points out, non-rational intuitions are too weak to build moral systems on.[30] Sinnott-Armstrong tested the Trolley Problem that I mentioned in chapter one on a group of undergraduates, with surprising results. The Problem asked whether it is morally better to let five people die or to perform a positive act and save the five at the expense of another person's life. 35% of subjects thought it was wrong to kill the one. However, when the same situation was described in more vivid language, that percentage jumped to 61%.[31] If Haidt claims that our moral judgements are justified when based on non-rational intuitions, the basis for those judgements looks far too flimsy to support the weight. Would it not at least be better to, say, use our instincts as an initial guide and then refine them when contrary information appears, or when we have more time to judge? Following Feinberg et al, this seems to happen more than Haidt thinks.

There is also something curious about the scenarios used in Haidt's experiments. They were engineered to provoke reactions of moral disgust even though they could not cause any harm. Haidt tried to remove all the rational arguments against them; in *Incest* both siblings were old enough to consent, they had no emotional problems, two forms of birth control were used and so on. Just how likely would this be in real life? How often could something like *Cannibalism* occur? Is it ever mentally healthy to knowingly eat human flesh?

Daniel Jacobsen raises the point that it is impossible to understand the motivations of Haidt's protagonists. The scenarios do not give sufficient information for the protagonists to be plausible to

---

[30] Sinnott-Armstrong (2006), p 344.
[31] Sinnott-Armstrong (2006), p 349.

us. What on earth, for example, would make Jennifer think that a human corpse would be a good meal? Jacobsen objects that her motivation could not psychologically spring out of nowhere – how could a moral vegetarian cheerfully eat human flesh?[32] I think, however, that Jacobsen does not go far enough. Even if Jennifer's character was psychologically plausible, that does not mean we could ever imagine sharing those motivations. Even if we had every reason to predict that Jennifer was going to eat a corpse, our emotional reactions and intuitions may render it practically impossible to empathise with her – and it is much better that we cannot, for a society that condones her behaviour is not a place we want to live. This is both for clearly rational reasons (How much confidence could we have in a medical/educational system that allowed *Cannibalism*?) and emotional reasons. Jacobsen himself explicitly agrees that, as I argued above, emotional reactions themselves can be grounds for moral condemnation.[33] Furthermore, we have no reason to believe that *Cannibalism* could ever be psychologically healthy, as Haidt gives no argument that it could. Haidt's scenarios may be psychologically impossible, making them irrelevant to questions of moral judgement-making.

Even if we assume that *Cannibalism* could occur, though, it would be very, very rare. Cannibalism is most often wrong because it usually involves murder, it mentally harms those involved, and so on. Our negative knee-jerk reaction to *Cannibalism* is therefore not surprising (Morphological rationalism would call this the result of unconscious reasoning about 'typical' results of cannibalism.). If this reaction is misguided, it is no wonder; Haidt's scenario is highly artificial. He has taken a situation that is typically ethically charged for good reason and then tried to remove all rational reasons to object. Since we have for years condemned cannibalism and have emotional reactions about it, it is no surprise that we cannot turn our intuitions and emotions off.

Nor does Haidt consider situations when reason tells us that something is morally wrong and emotion does not. Consider Karen's boss, a petty, spiteful, vindictive man. He bullies his staff, expresses grossly bigoted views and evades taxes. One day Karen sees him crossing the road, oblivious to an oncoming car. Karen knows that the car will hit him, causing minor injuries. Her boss does not know she is there. She can, without any risk to herself, pull him to safety, but she lets him be hit by the car instead, feeling only a pleasant *schadenfreude* afterwards. People may not be too emotionally anguished by Karen's behaviour, but we cannot assume that they would find it morally

---

[32] Jacobsen (2012), p 307.
[33] Jacobsen (2012), pp. 306-307, emphasis in the original.

acceptable. We are emotionally tempted to do many things, but we realises that there are good moral or non-moral reasons not to do them.

Haidt cannot object that my scenario is too artificial, as his were as well. Furthermore, even if Haidt is right that we mainly produce reasons for our moral judgements to rationalise them and convince others, the fact is that we feel compelled to.[34] Haidt was 'surprised' to find out that many subjects 'invented victims' in their justifications (for example, that a family should not eat their dead dog because of the health risks, even though the scenario stated that there were none).[35] But without good justifications for our moral judgements there is little chance of developing a moral system we can all take part in.

I have mentioned that Railton recently commented on psychological research, specifically Haidt's. Railton seems favourable towards it, but Haidt would not agree with his interpretation because Railton disagrees with Haidt over what an intuition is, giving it a quasi-morphological rationalist reading from the beginning. This colours Railton's use of Haidt's scenarios in ways that Haidt would not accept.

Haidt sees that intuitions can lead to moral dumbfounding and argues that if we cannot consciously give a rational reason for our reactions, there is *no* rational reason. Railton thinks that intuitions are the result of learning. Even if we cannot consciously articulate our reasons for rejecting *Incest*, we do have such reasons. This sounds like morphological rationalism, though it may not be identical. Morphological rationalism requires learning at the beginning of the process, but Railton's learning is looser and need not even be conscious. His main example of a trial lawyer almost completely involves unconscious learning. Even though the lawyer does it 'without knowing', she unconsciously knows how to read the mood of a jury. She knows when the jury is against her based on emotional reactions produced through unconscious learning and reasoning, even though she only has a feeling that something is wrong. Her intuitions can also provide a solution, telling her that doing *X* will swing the jury even though she does not consciously know why.[36]

One might think that Railton is too optimistic about basing intuitions on reason. He claims that with *Incest*:

---

[34] Haidt (2012), p 52.
[35] Haidt (2012), p 12.
[36] Railton (2014), section V.

'[Committing incest] was a highly risky, poorly motivated idea… Julie and Mark [were] insensitive to, and insufficiently motivated by, the lasting harm they might have caused to each other…'[37]

This is a good reason to reject *Incest*. It did not cause harm, but Julie and Mark had good reason to think that it would. It was just luck that it did not go wrong. But Haidt had a specific intention behind his scenarios to strip out any rational reason to reject them. All we need is *Incest 2*, in which Mark and Julie are experienced psychologists who have spent years researching the effects of incest and have knowingly minimised the risks. Railton's objection disappears, but we cannot assume that the intuitions against *Incest 2* do. Haidt might play this game forever, stripping out any objections that Railton comes up with whilst still retaining a scenario that people morally object to. However, Railton has a variety of responses. The first is that with each change the scenario comes closer to being psychologically impossible. Alternatively, we know from Feinberg et al that we do change our minds when presented with good evidence. A more interesting response, though, points out that most of the time incest is damaging, putting people at risk of immense harm. We feel this so strongly that we feel it even about highly exceptional cases like *Incest* where no harm is involved. These cases are extremely rare and difficult to identify, meaning that making exceptions to the rule would be very difficult and benefit very few people. Our intuitions are based on good, rational reasons for the vast majority of incest cases, and we are so used to them that the intuitions arise even when the individual case under consideration does not fit any objection. But it may cost so much for such little benefit to exempt it that idealised agents would agree that allowing the case is far too expensive. The *Incest* cases, even if they are harmless, are so rare and difficult to recognise when (or if) they occur that idealised agents may judge against them as a social objective interest and as morally permissible, even though *Incest* might be in Mark's and Julie's individual objective interests.

Ultimately, Haidt's philosophical position that our moral system should depend primarily on our emotional responses rather than rational judgement-making is not supported by the empirical evidence. He does not fully define what an intuition is, which allows Railton to argue that intuitions are the result of unconscious reasoning. His test scenarios of *Incest* and *Cannibalism* are too rare in real life for us to draw many conclusions from how his test subjects reacted to them. Finally, Railton can

---

[37] Railton (2014), p 848. Jacobsen (2012) makes similar points (pp. 301, 308).

also argue that emotions can be part of non-moral objective interests, and hence there is nothing in principal preventing them from becoming social objective interests.

As reason remains in the picture we can keep Railton's idealised agents with their perfect reasoning and full information. However, Haidt does have a point, that we should not necessarily exclude emotion from moral judgement-making. The emotional pull of our intuitions may tell us what the best course of action is, even if we are consciously unaware of the reasons behind them. We must also allow non-rational objective interests to play a part in moral judgement-making. In fact, as I shall now discuss, it seems that we cannot possibly ignore some of them.

Some of our objective interests appear to be biologically hardwired, which means that unless some special circumstance occurs we cannot get rid of them. For example, most people cannot avoid feeling an aversion to pain. The biological reason for this is that pain indicates injury, something to be avoided. However, when I have a headache my first reaction is not to wonder what the damage is. I just feel bad. The pain does not switch off when I realise that something is wrong or that I cannot alleviate it. I do not rationally choose to dislike pain; it just makes me feel awful. Other judgements are more rational. One does not sensibly become an entrepreneur just because one enjoys it, which may be an emotional reaction that one has no control over. One must also be good at it, have the means to succeed at it, and so on.

What about a society's collective interests? The main example I shall use of a biological, non-rational objective interest that is also morally correct is the effect of certain hormones on our interactions on others. It is a gross simplification to say that any one hormone or hormone-producing gene is solely responsible for a particular character trait or disposition towards particular behaviour. Nevertheless, they can influence our moral judgements. I noted in chapter one that the inability to clear serotonin from one's synapses has been linked to increased aggression, but I will concentrate on another hormone called oxytocin.

Oxytocin has a profound effect on humans' and other mammals' caring behaviour, most notably between mothers and their offspring. Oxytocin has various physical influences, such as contributing towards the mother's ability to breastfeed, but it also makes the mother more inclined to care for her child by feeding it, protecting it from harm, and so on. Why would the mother do this? As Patricia Churchland points out, '[n]ormally… tending to the infant is rewarding; it feels good. By contrast,

anxiety levels rise when the infant is crying, taken away or suffering, and this feels very bad'.[38] This effect also runs in the opposite direction, as the infant experiences a comforting release of oxytocin when cared for. 'Being together feels good. Humans know what this feels like even if we do not know anything about oxytocin…'[39]

Some people produce less oxytocin than others, such as people suffering from borderline personality disorder. They show very little trust or unconditional affection, which has devastating effects on their children. The psychologist Simon Baron-Cohen described one sufferer he met:

> 'If her children don't do what she says, she screams and swears at them [and makes suicide threats]… Minutes later she will drive to one of her friends and spend the evening having fun, leaving her children reeling with the impact of her hurtful words… [H]er own needs are paramount and her children's needs (or anyone else's for that matter) never even feature on her radar.'[40]

The lack of oxytocin can have a massive effect on how we see and treat people around us, whatever goals we have. It is simple enough to point out rational reasons why a parent should care for her children, but if a parent accepted those reasons without actually having any affection for the children, we would suspect that something had gone badly wrong. As David Hume recognised, we do not care for children just because we think it is a moral duty.[41] Oxytocin has existed for thousands of years, long before we could consciously think of morality. The mother's goal is the child's goal, and the mother does not reach this conclusion solely through rational deliberation. There is a non-rational, biological process going on. The compulsion to care for one's children is not based solely on unconscious reasoning.

If we have a particular goal that is not (or usually not) influenced by reason, then we may need very strong persuasion to abandon it. Imagine saying to a mother 'Your child is another mouth to feed. He drains time, effort and money from you. It would be best if you gave him up for adoption'. Unless the advantages for the child were massive, it is tough imagining any loving mother agreeing with this. It is even harder if you only point out the material advantages that the mother would gain. Children

---

[38] Churchland (2011), p 34.
[39] Churchland (2011), p 40.
[40] Baron-Cohen (2011), pp. 31-32.
[41] Hume (2000), section 3.2.1.5.

can be a constant source of anxiety and worry, but it does not seem strange to use that parents are willing to cope with it.

Railton's idealised agents can cope well with individual objective interests based on biological factors such as oxytocin, even when it is not in the agent's rational interest. Railton never claims that objective interests must be rationally justifiable or egocentric. Such conditions would sit uncomfortably with his claim that different objective interests need have no common traits. One can have a non-moral objective interest to satisfy another agent's object interests without getting anything in return. Oxytocin contributes non-rationally to the creation of a mother's objective interest to look after her child, even at great personal cost, and idealised agents must take this into account when assessing objective interests. The interest can be overridden, but we cannot assume that it will be. One's objective interests can be quite disadvantageous to oneself, but to a degree one can say 'So what? I am willing to suffer.'

This is important because so far we have been looking at moral intuitions which, on Railton's viewpoint, are based on reasons that could in principle be uncovered (such as the psychological risks Julie and Mark unjustifiably ignored when committing incest). Ultimately, however, such reasons could themselves at least partly depend on non-rational factors. It is not automatically moral to care for one's children just because one is biologically compelled to, but oxytocin contributes to a moral society by helping form non-moral objective interests which are themselves candidates for being moral objective interests. At first glance this seems similar to Lonnie in chapter two, whose biological makeup contributed to his interest in drinking clear fluids. However, neuroscience appears to uncover more ways in which non-rational objective interests can be found, including those that can become moral objective interests. Note also that Railton has left it open for the idealised agent to discover the actual agent's objective interests non-rationally. As they share the same psychology, they would be psychologically pulled in the same direction in the same situation. The idealised agent might realise that the psychological pull would actually be bad for the actual agent, but she would nevertheless identify the pull based on non-rational means. It must be remembered that the idealised agent is not meant to be someone who, like a psychologist, only has rational knowledge of the actual agent, but *is* the actual agent, just with superior cognitive powers.

Our objective interests can be influenced by many non-rational factors outside our control. We may have good reasons not to give such interests up, and they may even become social objective

interests (that is, moral interests) despite the lack of a rational base. It is, in a sense, just the way we are built. Recent research thus helps demonstrate more about how Railton's naturalism can work in practice, it provides more reason for idealised agents to share psychologies with actual agents, and it emphasises how non-rational interests can contribute to some forms of naturalism (including Railton's).

I should warn here that I have kept my discussion simple. Our behaviour, desires and needs are influenced by many factors, both biological and non-biological. A single gene will not cause a certain behaviour by itself, and genes combine in multiple ways to have multiple psychological and physical effects. Biological properties also have limits. Prairie voles given an overdose of oxytocin developed a weaker attachment to other voles, and oxytocin can make one trust even transparently dishonest people.[42] We should not blindly go down any route our neurobiology pushes us towards. We may have too little oxytocin, too much, or we may simply fail to see reason. At some stage we have to ignore what we are biologically inclined to do and examine our judgements more closely. However, unless we dig out the oxytocin receptors in our brains, oxytocin is here to stay. To a certain extent, for better or worse, oxytocin influences what we morally care about. The fact that the influence is non-rational is immaterial.

Even though we have uncovered a promising field of study of non-rational objective interests, it still appears that reason reigns supreme when it comes to making moral judgements. Non-rational interests may contribute, but reason judges if we accept them or not morally. Even emotionally-loaded moral intuitions rest, it appears, on unconscious reasoning. But could purely non-rational processes be *essential* for moral judgement-making? Haidt was unable to show that they could, but in the next chapter I shall argue that moral judgement-making is impossible with emotional input.

---

[42] Churchland (2011), pp. 78-779.

<u>Chapter 4</u>

In the last chapter I argued that Railton's naturalism needs to take into account both rational and non-rational factors. But could we entirely exclude non-rational factors? Can we recognise any moral properties solely through reasoning? I do not think that we can, and this has important implications for Railton's naturalism. In this chapter, I will use an argument from John McDowell as a basis for arguing that psychopaths cannot make moral judgements because they lack the power to feel and recognise certain moral emotions. Consequently, recognising moral properties requires a non-rational, emotional element, which Railton does not currently include in his naturalism.

What is important about McDowell's argument is not its conclusion, but its roots in Wittgenstein's arguments about rule-following.[1] The argument purports to show that various forms of non-cognitivism (such as quasi-realism) cannot adequately explain how we can correctly follow moral rules. McDowell claims that we can only recognise moral properties correctly if we are part of a moral community that recognises those properties, but non-cognitivism implies that we can do so independently of the community. I will not examine whether the argument works, but rather McDowell's idea of an agent's relationship with a moral community. Let us call those agents who have no connection with a particular moral community 'moral outsiders'. I shall argue that psychopaths are a type of moral outsider and therefore cannot make genuinely moral judgements, and that the deficiency in their moral thinking is not one of rationality, but is instead emotional.

I shall first describe in more detail McDowell's idea of moral outsiders. Using psychological research, principally experiments conducted by R.J.R. Blair, I shall then argue that psychopaths are moral outsiders because they cannot tell the difference between moral wrongs (such as attacking other people for fun) and conventional wrongs (such as parking on double yellow lines). I will then examine Shaun Nichols's argument that the psychopath's disability is not a rational one. Unfortunately, the argument is not as strong as it could be because Nichols does not provide an alternative type of deficiency that the psychopath could suffer from. One possibility is the psychopath's lack of empathy for other people, but I shall argue that this is not enough to make psychopaths moral outsiders. Instead, they are unable to feel various emotions that are commonly

---
[1] McDowell (1998), pp. 198-218.

involved in moral situations, and they cannot infer these emotions in other people either. This bars them from making genuinely moral judgements and so makes them moral outsiders.

Imagine that a certain community has a particular moral value (A community is simply a group of agents with a shared viewpoint on some matter.). An outsider to that community plays no part in it and does not see the value from the viewpoint of the community. McDowell asks whether the outsider could correctly ascertain the extension of the term that denotes the value whilst remaining independent of the community. If she can, she can use the term correctly in any novel situation even though she has never seen the term through the community's eyes. This does not mean sharing the value, but understand why the community holds it, putting oneself briefly in the community's shoes. McDowell objects that moral outsiders cannot correctly use moral terms because to determine the terms' extensions one must engage with the community's moral outlook.

McDowell discusses two possible ways of determining a term's extension without participating within the community using the term.[2] One might identify the rules of determining the extension by codifying the practice or by grasping a particular universal (for example, one can only follow the practice of identifying red objects by grasping the universal of redness). McDowell thinks we usually grasp universals when a practice resists codification, but both methods presume that there is something 'independently there' (a universal or a rule) that we can use to get the correct results.[3] Following Wittgenstein's rule-following considerations, McDowell argues that this is mistaken.

Consider the practice of adding 2. It seems that we can codify the practice to form a universally applicable rule about adding 2, which can be applied to any number. We can prove that each correct answer is right. However, we can never be certain that the rule is properly understood, because we can only check a finite number of examples from an infinite number of possibilities. An agent can continue up to 996, 998, 1000 and then continue with 1004, 1008, 1012… There are two possible problems that one can discuss here. The first is that the agent fails to understand why his actions are wrong. How can we prevent him – or even *ourselves* – from making unwitting errors? We may follow many rules incorrectly because we honestly believe that the rules tell us otherwise. How can one ensure that they always follow a rule correctly? One might get a rule wrong because it has not been clearly explained, but eventually the explanations must stop. Eventually there will be no new helpful

---

[2] The term need not be moral.
[3] McDowell (1998), pp. 203-204.

way to explain the rule, and even then there is no guarantee that the rule will be correctly understood. Even if you can give rules about how to follow other rules, there cannot be an infinite regress of rules about how to follow rules about how to follow rules about… No set of rules guarantees that other rules will be followed correctly.

The second possible problem is that one cannot be certain that one is following a rule at all, either correctly or incorrectly. If we cannot know whether we are following a purported rule, then how can we call anything a rule? We will not know whether we can follow any purported rule, since there are no definite criteria to decide if we are. Rules therefore become meaningless. This is McDowell's own interpretation, but his solution applies to both problems. He claims that one cannot follow a rule 'independently of all the human activities and reactions'.[4] We must instead share 'routes of interest and feeling'.[5] Mathematical calculations are not subjective, but learning mathematical rules will not guarantee correctness unless one has been 'trained' in how society uses mathematics. We must participate in the community that generates the rule. We do not recognise that $13^2 = 169$ because mathematical platonism is true and $13^2 = 169$, claims McDowell; we are instead trained to find 'such-and-such calculations compelling'.[6] We cannot point to a particular set of rules to ensure that other rules will be followed correctly, and this applies to moral and evaluative practices as much as to mathematics.

McDowell concludes that if a particular community uses a moral term in a certain way, an agent can only ensure that she uses it correctly by engaging with the community's viewpoint, however temporarily. He then argues that non-cognitivists deny this, making their position implausible. Whilst I will not discuss the rest of his argument here, it is worth noting that non-cognitivists can argue against McDowell even if they accept the existence of moral outsiders.[7] My use of moral outsiders does not presuppose the truth or falsity of non-cognitivism, though I am examining how it will affect a form of moral realism.

What makes an agent a moral outsider? There can be many reasons: outsiders may not know the moral community's viewpoint, have no interest in appreciating it, rationally reject it and so on. Psychopaths, I believe, are one type of moral outsider because of some deficiency in their mental makeup. An obvious explanation of the deficiency is that they cannot rationally understand moral

---

[4] McDowell (1998), p 207.
[5] McDowell (1998), p 206.
[6] McDowell (1998), p 208.
[7] For example, see Miller (2013), pp. 240-252 and Blackburn (1998), p 93.

properties, but as McDowell's use of moral outsiders hints, this is not the key problem. The psychopath's deficiency is not cognitive.

Formerly, Railton might not have agreed that psychopaths are moral outsiders.[8] He considers a situation where an ethically competent person suffers an injury that deprives him of the ability to sympathise with others.[9] Railton claims that the sufferer can make moral judgements by following his past behaviour and what he would have done before the accident. One could analogously claim that a man who suddenly becomes blind can still use visual terms as a sighted person would. Though he cannot see the sky, he remembers what it looks like. Railton further claims that:

'A person [who has lost his ability to sympathise]… who nonetheless was concerned to make genuinely moral judgments would perhaps try to situate himself like 'people in general' to compensate for his peculiar incapacity.'[10]

However, this runs into difficulty with McDowell's moral outsiders. Remember McDowell's questions: can an outsider correctly ascertain the extension of a term that denotes a particular value whilst remaining independent of the community? Can the outsider correctly use the term in novel situations? If the person who has lost the ability to sympathise is a moral outsider, then past performance is not a guide to future results. He cannot depend on his past behaviour to cope with novel situations. Furthermore, Railton's suggestion that he copies other people to compensate presupposes that he is already a moral insider, because copying other people correctly would mean seeing moral judgements from the community's viewpoint. But Railton gives no argument that such a person can do this. In fact, we shall see that we have good reason to believe that psychopaths, who famously have no true sympathy for others, do not. Between McDowell and psychological research on psychopaths, it appears that Railton is mistaken.

What is a psychopath? A person is psychopathic if they display a persistent disregard for others that involves three or more of the following traits:[11]

---

[8] His position on this may have recently shifted, as I will discuss in chapter five.
[9] Railton (1996), pp. 73-74.
[10] Railton (1996), p 74.
[11] American Psychiatric Association (2013), section 301.7.

1) Failure to conform to social norms of lawfulness:

   - Performing acts that are arrestable offences

2) Deceitfulness:

   - Repeated lying

   - Use of aliases

3) Impulsivity or failure to plan ahead

4) Irritability and aggression:

   - Physical fights and assaults

5) Reckless disregard for the safety of oneself or others

6) Consistent irresponsibility:

   - Repeated failure to sustain work commitments

   - Repeated failure to honour financial obligations

7) Lack of remorse:

   - Indifference to having hurt, mistreated or stolen from someone

   - Rationalising having hurt, mistreated or stolen from someone

Psychopaths have trouble relating to others, but it is not simply that they cannot imagine others' reactions. Deceitfulness requires a certain awareness of how others think, and it has been recognised since at least 1941 that psychopaths display 'superficial charm', which requires social awareness.[12]

The criteria do not explicitly mention any rational moral defects. Psychopaths may have reasoning problems regarding, say, their inability to plan ahead, but this is not specifically a moral failing. One can be moral yet unable to recognise many moral and non-moral consequences of one's actions. The difference between the moral agent who lacks foresight and the psychopath is that if their actions injure others, the moral agent will feel remorse and the psychopath will not. Maybe, however, it is implicit in the criteria that psychopaths suffer from a rational defect. Current psychological research, though, does not point in this direction. R.J.R. Blair has performed a number of experiments with psychopaths showing that they can identify that an action is wrong, but not that it is *morally* wrong. These experiments form the basis of a good argument that whatever deficiency psychopaths have in making moral judgements, it is not one of rationality.

---

[12] Baron-Cohen (2011), p 46.

Blair's experiments focus on the moral/conventional distinction.[13] A wrongful act can be either a moral or conventional transgression, which Blair defines as follows:

'[M]oral transgressions have been defined by their consequences for the rights and welfare of others, and social conventional transgressions have been defined as violations of the behavioural uniformities that structure social interactions within social systems.'[14]

We should not take these definitions as gospel. One might argue that rights do not exist, that considerations of welfare can be trumped by other considerations, that social interactions often have moral implications and so on. Even Blair himself wanders a little from the definitions, as he concentrates more on welfare than rights. Another problem with defining the two types of transgression is that a particular definition may give the impression that one can judge whether a transgression is moral or conventional using purely rational methods. Since, as we shall see, psychopaths cannot recognise the distinction, such a definition would presuppose that the psychopath's failure is a rational failure. Equally, another definition might presuppose that the failure is non-rational, which is what I am trying to show.

This problem is important, and I shall return to it when I discuss the psychopath's emotional powers. For the moment, though, we can put it to one side because however we define the types of transgression, there appears to be a real difference between them. Most of us recognise that the rule not to kick babies for fun is qualitatively different from the rule to include eleven people in a professional football team, for example. Though it might be difficult for us to tell if a particular transgression is moral or not, much of the time we can do so. Furthermore, the examples of transgressions that Blair uses are fairly unproblematic and need not hinge on accepting the definitions that he uses.

I shall concentrate on two of the hypotheses that he tested:

1) That psychopaths will not make a distinction between moral and conventional rules.

2) That psychopaths will be less likely to make reference to the pain and discomfort of others than non-psychopaths.[15]

---

[13] Blair (1995).
[14] Blair (1995), p 5.

The test subjects were ten psychopaths and ten non-psychopaths, who were the control group. All participants were in jail. Each was presented with four moral and non-moral stories set within a school. The moral stories involved a child hitting another child, pulling another child's hair, smashing a piano and breaking a playground swing. The conventional stories involved a boy wearing a skirt, children talking in class, a child leaving the classroom without permission and a child paying no attention to the teacher. The subjects were asked whether acts would still be wrong if the teacher allowed them.

The results showed that the psychopathic subjects and the non-psychopathic subjects could recognise that the acts were wrong. However, the psychopaths were much worse at making the moral/conventional distinction. Eight non-psychopaths made a clear distinction and only two psychopaths did. Of those two, one only seemed to recognise the distinction when it involved physical injury. Psychopaths were also far less likely to claim that moral transgressions were wrong based on the welfare of others. They were more likely to refer to rules (for example, 'you should not do that because it is not allowed'). This implies that they did not consider the rights of others either.

This fits with McDowell's idea of moral outsiders. Just giving psychopaths rules about morality is no guarantee that they will be able to make the moral/conventional distinction. They need something more to help them look at morality as the community does, to care for other people's welfare and rights. As they seem unable to determine what makes a transgression moral, this implies that they are isolated from the communities that can make the distinction, and thus that they are a type of moral outsider.

What makes psychopaths outsiders? At first sight it may look like a failure of reasoning. If one justifies a moral judgement by saying that a rule forbids or commands something, it is not much of an explanation. We want to know why the rule exists. However, it is a better explanation for conventional transgressions. It informs the listener that society does not tolerate the transgression, even if the underlying reason for the prohibition is a bad one. So the psychopath does not fully understand what is being asked about moral transgressions, and this is a failure of reasoning about what makes something morally wrong. However, this rationalist explanation of the psychopath's problem is incorrect. We can introduce here a theory attacked by Shaun Nichols called empirical rationalism.

---

[15] Blair (1995), p 13.

Empirical rationalism claims that 'it is an empirical fact that moral judgment in humans is a kind of rational judgment, i.e., that our moral judgments derive from our rational faculties or capacities'.[16] Nichols uses the terms 'reason' and 'rationality' more or less interchangeably, so I will treat empirical rationalism as claiming that we can make moral judgements using the same rational powers we have to make scientific, mathematical and logical judgements.[17] If empirical rationalism is true and psychopaths cannot make genuinely moral judgements, then the failure can be empirically shown to be rational. Nichols criticises empirical rationalism on the grounds that non-rational faculties are actually essential to making moral judgements. Psychopaths are moral outsiders because they lack these faculties, and so empirical rationalism is unable to explain what the psychopath's moral problem is. Nichols lists three possibilities, arguing that none of them works:

1) The psychopath lacks rational perspective-taking abilities.
2) The psychopath lacks general rational abilities.
3) The psychopath suffers from intellectual arrogance.[18]

The third option seems the most unlikely. It claims that intellectual arrogance leads the psychopath to claim that his own views are superior to everyone else's without being able to explain why. Nichols notes that this does not explain the psychopath's inability to recognise the moral/conventional distinction. Since psychopaths can acknowledge that their actions are wrong, how can intellectual arrogance prevent them from recognising that those actions are specifically *morally* wrong? I would add further that though intellectual arrogance exists, 'arrogance' implies that the psychopath chooses such behaviour wilfully and deliberately. This seems unlikely. First, psychopathy is extremely difficult to treat, meaning that it is unlikely to be 'arrogance' that someone can be reasoned out of. I might be arrogant about my running skills, but if I fail dismally in a hundred-metre race with other competitors it might make me reassess my abilities and become less arrogant. There is not yet any way to demonstrate to psychopaths that they are not morally superior to others in a way that would cause them to reassess their attitudes. Second, the psychopathic brain is different from the neurotypical brain. For example, the sections of the brain related to empathy show less activity in psychopaths and their orbito-frontal cortex is much smaller than the average. The orbito-frontal cortex is active when,

---

[16] Nichols (2002), p 290.
[17] Nichols (2002), p 292.
[18] Nichols (2002), pp. 269-299.

say, one sees a needle going into another person's hand, and appears to be linked to empathy.[19] The

psychopath's mental deficiency thus seems linked to physical properties of the brain, which is not

something that the psychopath can choose. So intellectual arrogance is not the solution.

Could there be a problem with the psychopath's perspective-taking abilities? Perspective-taking

abilities allow us to take the point of view of another person. Nichols points out that there is good

empirical evidence that psychopaths do have such abilities. Psychopaths can be charming and

manipulative, two traits that one cannot have unless one can predict to some extent how other people

will behave. Richell et al (2003) tested psychopaths to see if they could correctly predict the emotions

of another person when they could only see the eye region of the person's face. The psychopathic

subjects generally did no worse than the non-psychopathic control group, indicating that they were

just as good at determining the state of mind of others. Psychopaths also know when they are hurting

others. The problem is not that they do not notice, but that they do not care. They may hurt others

without remorse for trivial reasons, such as killing someone who has 'disrespected' them, but the

injury is intentional and deliberate.[20] Psychopaths can generally tell what others are thinking and

feeling. As we shall see, it turns out that this ability is not perfect; there are some situations in which

psychopaths cannot predict what others are feeling. However, this failure does not seem to be one of

rationality. I will discuss this later in the chapter when looking at the psychopath's emotional powers.

Nichols moves over the psychopath's perspective-taking abilities rather quickly and does not note

another experiment from Blair that strengthens his argument. Autism is a disorder that hinders

sufferers from understanding other people's perspectives. Some sufferers are impaired to the point

that social interaction is practically impossible, whilst others have varying degrees of difficulty. Baron-

Cohen describes one sufferer who was bewildered by everyday conversation because he 'hadn't a

clue' how to cope with 'humour or sarcasm or metaphor, or – even worse – body language'.[21] The

sufferer did not know how to interpret the social behaviour of those around him.

Baron-Cohen et al (1985) investigated whether autistic children could make inferences about what

other people believe, using the Sally-Ann test. Test subjects saw a doll, Sally, place a marble in a

basket and leave the room. Another doll, Ann, came in and moved the marble to a box. Sally then

came back in and the subjects were asked where she would look for the marble. 80% of the autistic

children pointed at the box, not realising that Sally would not know that the marble was there. Since

---

[19] Baron-Cohen (2011), pp. 22, 54-55.
[20] Baron-Cohen (2011), pp. 43-45.
[21] Baron-Cohen (2011), p 67.

the autistic children performed worse than the subjects with Down's syndrome, the failure was not due to low IQ or 'the general effects of mental retardation'.[22]

Nichols fails to note that this experiment provides evidence, along with another of Blair's experiments, that rational perspective-taking is not essential to making moral judgements.[23] In Blair's experiment neurotypical ('normal') children, autistic children and children with mild learning disabilities were asked questions about various scenarios to see if they could make the moral/conventional distinction. The autistic children were grouped separately depending on whether they could pass tests such as the Sally-Ann test, and thus whether they could understand other people's viewpoints. The result was that autistic children could generally make the distinction, and their ability to do so had no correlation with their ability to pass the Sally-Ann test. Autism sufferers can therefore do what psychopaths cannot, even when they lack perspective-taking abilities.

Maybe, then, psychopaths have a more general rational deficiency. Nichols notes that for this to work one must explain what the deficiency is. He concedes that it is possible, but that the onus of proof is on the empirical rationalist.[24] However, the rationalist could reply that the best Nichols can hope for here is a stalemate. It seems uncontroversial to many people that good moral judgements require reasoning. If someone admits to making moral judgements by non-cognitive means, such as being emotionally affected by the situation, we tend to be suspicious of the judgement. The empirical rationalist can claim that Nichols has not shown that a non-cognitive explanation of the psychopath's moral deficit is better than cognitive explanations, and until he does the empirical rationalist can shrug off the onus of proof. This links back to my worry about attempts to concretely define moral and conventional transgressions, as some definitions risk being entirely rational. If, say, we define moral transgressions as transgressions that negatively affect X's welfare under conditions Y, where X is a strictly defined group of individuals and Y is a strictly defined set of conditions, then we have a rational way to judge whether a particular transgression is moral. Nichols's arguments would fail, as one could make the moral/conventional distinction purely by using a rational judgement-making process. Therefore the psychopath's inability to make the distinction would be rational as well.

Nichols has two options. He could provide a more plausible alternative to rational deficits, which I will explore a little later. He could also show that empirical rationalism is so implausible that we should reject it anyway. We have already seen that autistic children and children with low IQs can make the

---

[22] Baron-Cohen et al (1985), p 44.
[23] Blair (1996).
[24] Nichols (2002), pp. 297-298.

distinction, but we can go further. Very young children seem to recognise the distinction, which raises questions of how much rationality one needs to do so.

Judith Smetana tested preschool children on their ability to make the moral/conventional distinction and found that they generally could:

'The findings suggest that by age 2½, children's judgments of transgressions depend on the nature of the act and that moral events are more consistently distinguished than social-conventional events. Preschool children evaluated all moral transgressions as more serious and more deserving of punishment than all conventional transgressions. Nearly all moral events were evaluated as wrong in the absence of rules, evaluations of conventional events as rule contingent approached but did not always reach [statistical] significance.'[25]

So two-year-olds can do what psychopaths cannot! Smetana also tested preschool children to ascertain their reason for making the distinction.[26] She found that children consistently thought that moral transgressions were wrong because they impacted on other people's welfare, whereas conventional transgressions were wrong because they created disorder. This happened even when the children did not know what the specific transgression was. For example, they were told that Sally did something that made Jessica cry, but not what it was. Blair's experiments showed that psychopaths were bad at considering an act's impact on people's welfare, so this shows another difference between psychopaths and pre-school children. Furthermore, Nichols has pointed to experiments that show children of one and two years old altruistically trying to comfort others whom they believed to be in distress.[27] These experiments appear to indicate that preschool children are more aware of moral properties than psychopaths are. If the ability to recognise the moral/conventional distinction is rational, then, it seems that it is a rational ability shared by very young children, autistic children and children with low IQs, but not psychopaths. As Nichols concedes, the ability might indeed be rational, but with each example like this it looks more unlikely.

We also have more evidence that psychopaths are a type of McDowell's moral outsiders. It seems that most agents, from an early age, are able to recognise when a transgression is moral or not. This is not simply because we rely on authority to tell us; consider Smetana's discovery that preschool

---

[25] Smetana (1981), pp. 1335-1336.
[26] Smetana (1985).
[27] Nichols (2001), p 437

children can identify a transgression as moral or conventional even when they do not know what the transgression is. But psychopaths are unable to distinguish between moral and conventional transgressions. They can tell that a moral wrong is a wrong, but not specifically a *moral* wrong. It thus becomes puzzling how they could genuinely label a transgression as moral when they cannot separate moral and conventional transgressions. It is more plausible that they cannot.

We also have some reason to doubt that the psychopath's deficit is rational, but it would be better if we could find a suitable non-rational explanation of the deficit. Is there one?

Even if one rejects empirical rationalism, there remains something attractive about the idea that psychopaths cannot see another individual's point of view. A fairly common view today, to the extent that writers such as Steven Pinker complain about it, is that morality is primarily a matter of empathy.[28] Simon Baron-Cohen is a good example of an empathy theorist, as he has tried to replace 'the unscientific term "evil" with the scientific term "empathy"'.[29] He believes that what makes the psychopath 'evil' is their inability to see people as people rather than as objects, because they famously have no empathy. He writes that '[w]hen our empathy is switched off, we are solely in the 'I' mode. In such a state we relate only to things, or to people *as if they were just things*'.[30] However, this does not mean simply being unable to see another person's point of view. Baron-Cohen's definition of empathy is 'our ability to identify what someone else is thinking or feeling, and to respond to their thoughts and feelings with an appropriate emotion'.[31] Psychopaths can normally identify what others feel, but they cannot respond appropriately. An 'appropriate response' is presumably one that treats others as human beings. Baron-Cohen uses this definition to try to construct an empathy-based account of morality.

If an empathy-based account of morality is true, than I have been looking in the wrong direction. The correct test for whether an agent can make genuinely moral judgements would be whether they have sufficient empathy, in which case the moral/conventional test I have been using becomes redundant. One can check if an agent has empathy without using the moral/conventional test. The test even conflicts with some empathy-based accounts of morality, as autism sufferers lack empathy but can identify the oral/conventional distinction. A crude empathy based account would conclude that

---

[28] Pinker (2011), pp. 689-691.
[29] Baron-Cohen (2011), p xi.
[30] Baron-Cohen (2011), p 5, emphasis in the original.
[31] Baron-Cohen (2011), p 11.

autism sufferers cannot make genuinely moral judgements, despite what moral/conventional tests indicate. As we should not call people moral outsiders unless we really have to, one could argue that this only shows a problem with empathy-based theories; that autism sufferers can make moral judgements, and, since they can, empathy is therefore not necessary for moral judgements. This does not mean, though, that every empathy-based theory has this problem. Baron-Cohen's own theory specifically allows autism sufferers the ability to make genuinely moral judgements.

Empathy-based accounts can also avoid other obvious difficulties. Jesse Prinz mentions that empathy is typically bad at motivating, it can be biased (for example, when someone shows more empathy towards cute animals than ugly ones) and that it can be manipulated.[32] More sophisticated empathy-based accounts can avoid these pitfalls, though, as they can claim that empathy needs to be appropriately directed and that it is essential but not sufficient for good moral judgement-making. Your empathy cannot be manipulated if you have sufficient reasoning power and all the relevant facts, whilst an agent with empathy for all animals may have more appropriate empathy than an agent who only has empathy for cute animals. The first true problem for empathy-based accounts of morality is the potential number of exceptions to the rule. We can see this in Baron-Cohen's own account. He acknowledges that his goal of replacing 'evil' with 'lack of empathy' requires substantial caveats, as '[h]aving empathy is not the sole route to developing a moral code and moral conscience'.[33] He has to say this because he thinks that autism sufferers can make moral judgements.[34] How can he do this?

Autism sufferers are typically very good at systematizing. They can easily identify patterns in the world that others do not notice and they become disturbed by situations with no patterns, as such situations are unpredictable. Autism sufferers greatly prefer routine lives with no surprises. Baron-Cohen argues that:

'[Autism sufferers] have developed their moral code through systematizing. They have a strong desire to live by rules and expect others to do the same, for reasons of *fairness*… [They] are often the first to leap to the defence of someone who is being treated unfairly, because it violates the moral system they have constructed through brute logic alone.'[35]

---

[32] Prinz (2011), pp. 225-227.
[33] Baron-Cohen (2011), p 65.
[34] Baron-Cohen (2011), chapter 4.
[35] Baron-Cohen (2011), p 84, emphasis in the original.

However, this makes it more mysterious how psychopaths can be moral outsiders. Baron-Cohen claims that 'brute logic' is enough to be moral, so we do not actually need empathy. It is only one route to having a moral conscience. Therefore the psychopath actually has two problems: lack of empathy and the inability to systematize to construct a moral system, which appears to be a rational failure. So empirical rationalism can re-enter the picture if it can explain this failure. Furthermore, if neurotypical agents reach moral judgements through empathy and autism sufferers do so through reason, Baron-Cohen needs to explain how two such different processes can yield the same answers for much of the time.

Baron-Cohen's explanation of the psychopath's moral problem in terms of empathy is at best only half a solution. The autism sufferer is an important exception to the explanation. Another important exception is moral judgements that seem to hold no place for empathy. Jesse Prinz claims that empathy is not enough to construct a moral system because some moral transgressions are victimless or involve no salient victim, so empathy cannot apply.[36] Unfortunately he does not say much about this objection, but we can find cases where the victims are at least nebulous enough for empathy to seem unnecessary. Take the example of stealing food from a supermarket. It is difficult to point to a particular person who will be especially aggrieved by the loss. Should it be the supermarket board of directors? The shareholders? The employees? Maybe it should be the company, but companies do not have feelings and one cannot empathise with them. Perhaps nobody cares very much, which shows another problem with empathy-based accounts of morality. If the immorality of a certain transgression depends on how the victim feels (or would feel if he were aware of it), then we face a 'happy slave' problem. No matter how injurious the action is, it is not unethical if the victim would not care. So an empathy-based account would have to add various conditions that, say, the victim would care if he had the correct attitude about the transgression, and the account would grow more complex as we added more conditions. Maybe it is better to drop empathy-based accounts of morality altogether.

If we do drop them, we need another way of using non-rational processes to make moral judgements. Fortunately there is one, and it allows autism sufferers to make genuinely moral judgements whilst keeping psychopaths as moral outsiders. Baron-Cohen hints at it when he says that autism sufferers care about fairness. But why should they? Basing a code of behaviour on

---

[36] Prinz (2011), p 214.

fairness is not the only way to construct a logical, consistent, predictable code, so why take fairness as a basis? The reason is that autism sufferers can feel certain emotions that psychopaths cannot.

Prinz suggests that we use particular emotions rather than just empathy to make moral judgements, such as guilt.

'[W]hen I judge that something is wrong, that judgment token derives from a sentiment, and consists in the appropriate emotional response. If I judge that I was wrong to eat the last cookie, my judgment consists in a feeling of guilt about my action.'[37]

Prinz further suggests that empathy comes later in moral development when, if it were essential to morality, it should come earlier.[38] For example, Nancy Eisenberg-Berg has demonstrated that in some moral scenarios children are more likely to respond empathetically the older they are, indicating that empathy is not necessary for early moral development.[39] Instead, Prinz claims, a psychopath's lack of empathy is the result of an emotional deficiency rather than the cause of a moral deficiency. The emotional deficiency causes both lack of empathy and the moral deficiency. It is the inability of psychopaths to recognise particular emotions such as guilt or remorse to feel those emotions deeply, a condition known as 'shallow effect' which is often displayed by psychopaths. This means that psychopaths cannot feel moral emotions such as guilt and consequently cannot make genuinely moral judgements. Neither can they feel empathy. We do not have to agree here with Prinz that moral judgements only consist of feelings. All we need to do is demonstrate that particular emotions are necessary for moral judgements and that psychopaths cannot share these emotions. Psychological research gives good reasons for thinking that psychopaths really do have an important emotional deficit.

Psychopaths can generally tell what others think or feel, but they cannot do so perfectly. There is some evidence that psychopaths cannot successfully predict the emotions of other people in certain situations because they themselves lack the ability to feel such emotions. Blair tested what emotions neurotypical children and children with psychopathic tendencies attributed to protagonists in various

---

[37] Prinz (2011), p 215.
[38] Prinz (2011), p 217.
[39] Eisenberg-Berg (1979).

stories.[40] The stories covered the emotions of happiness, sadness, embarrassment, fear and guilt. Examples included a protagonist finding a bee in their car, a protagonist unintentionally causing pain to another person and a protagonist finding a very small crying child. The children with psychopathic tendencies could easily attribute happiness, embarrassment and fear to story protagonists, but they were far less likely to attribute guilt, sympathy or sadness than the other children were. This matches results from Blair et al (1995), which indicated that adult psychopaths were unable to attribute 'moral' feelings (such as guilt and remorse) to other people that they themselves could not feel. This may be a problem with perspective-taking, but in a specific way. The psychopath is bad at feeling specific emotions. They can recognise that an act is wrong (if not morally wrong), but whereas ordinary people can respond to moral wrongs with specifically moral feelings, the psychopath cannot. This affects the psychopath's ability to infer such feelings in other people. It is a perspective-taking problem concerning particular affective, emotional powers rather than reasoning powers.

Nichols has pointed to further experiments by Blair to support the view that affective responses matter. Blair found that psychopaths and children with psychopathic tendencies showed considerably less response to distress cues (such as pictures of upset people) than others did, whereas there were no statistically significant differences when they were shown neutral images such as picture of hairdryers.[41] Autistic children showed a far greater response to distress cues, such as pictures of upset people, than to neutral images, with two children even covering their eyes and refusing to look at the cues.[42] Even if autistic children are bad at inferring the beliefs of others they may emotionally react more to other people's suffering than psychopaths do. They can apparently share moral emotions that neurotypical agent shave, whilst psychopaths cannot. Autistic sufferers' ability to systematise may help them construct moral systems, but those systems also require appropriate emotional input which is denied to the psychopath.

This lends more weight to the idea that psychopaths are moral outsiders and we now have a good idea why they are. Psychopaths are not members of the moral community because they cannot emotionally react to moral situations as the community does, and autism sufferers can be insiders if they do react appropriately. We can also see why it is no good giving psychopaths lists of rules to help them copy moral insiders. They cannot 'piggy-back' on the actions of moral insiders to follow

---

[40] Blair (1997). The children with psychopathic tendencies were significantly less likely to identify the moral/conventional distinction.
[41] Blair (1999b).
[42] Blair (1999a), p 483.

moral rules correctly, because their inability to feel moral emotions bars them from predicting when moral insiders feel such emotions. They are caught in a catch-22 situation; to copy moral insiders correctly they need to predict the insiders' emotional states correctly, but if they could do that they would not be moral outsiders in the first place.

We can now also respond to the worry I raised when I considered Blair's definition of moral and social transgressions. If the definitions of 'moral transgression' and 'social transgression' were sufficiently simple, it would be easy to rationally distinguish between the two. All one would need to do is ask, say 'does this transgression negatively affect someone's welfare?' If it does, it is a moral transgression; if not, it is a social transgression. I have already pointed out that the definitions Blair uses raise a lot of questions, and that neat and simple definitions require a lot of argument to be plausible. We can now see another objection to definitions that are exclusively rational. Though psychopaths may have rational defects, what really separates them from the moral community is their emotional deficiency. You can show them that a certain act affects someone's welfare, but they will not respond with moral emotions to it or regard it as morally important. They will not respond with moral emotions to any harm they cause, and they find it difficult to imagine anyone else doing so either. They therefore cannot fully understand a moral community's outlook because they are unable to share or recognise the appropriate emotional, non-cognitive traits of the community. When we consider that autism sufferers, people with low IQs and very young children can respond with appropriate emotions and recognise the moral/conventional distinction, it is hard to see what the psychopath's problem specifically is if it is one of rationality. As Nichols noted, it is still possible that empirical rationalism is correct, but the available psychological evidence is against it. Whatever the correct definition of 'moral transgression' is, it is unlikely to only include properties that we can ascertain through purely rational means.

I will not explore precisely what this affective response is or how it is generated. For my purposes it is enough to have shown that it helps produce genuinely moral responses to situations, and since psychopaths lack it they cannot make genuinely moral judgements. They really are a type of McDowell's moral outsiders. However, rationality still plays an important role in making moral judgements, as one's emotional response to a certain situation may be inappropriate. If I see someone crying, I may feel an emotional pull to comfort him. If I then discover that he is crying because he has been rightly convicted of defrauding a charity, my emotional impulse might still be

there but I would resist it because I would think that he does not deserve comforting. Consider also very young children trying to comfort others in distress. As they do not understand what others need they often offer ineffective means of comfort, thinking that what comforts them, such as a security blanket, will comfort others.[43] It is thus worth highlighting an important caveat. Affective, non-rational responses are needed to make genuinely moral judgements and to recognise moral properties, but we cannot ignore the role of reason in moral judgements and we cannot have a stable moral system without important rational input.

If psychopaths cannot make genuinely moral judgements because of an emotional deficiency, what implications does this have for Railton's naturalism? We have already seen that his claim that an agent who has lost his sense of sympathy with others can make moral judgements is wrong, at least if such an agent is psychopathic. Railton's naturalism, though, never claims that psychopaths *have* to be able to make moral judgements. All Railton has to do, it seems, is say that psychopaths cannot discover what is morally good like idealised agents can. But what of a psychopath's own idealised self? Consider John, a psychopath. His idealised agent, John+, is perfectly rational and has all the cognitive powers that other idealised agents have. But if he is in all other respects like John, then he will be psychopathic as well, for psychopathy does not depend on rational deficits. John+ will be unable to distinguish between conventional and moral wrongs, something that other idealised agents can do. The 'neurotypical' idealised agents can recognise that an action is wrong based on the welfare of individual actual agents and groups of actual agents, which will lead them to declare certain actions morally obligatory, permissible or neither. The psychopathic idealised agent will still be able to agree on what other agents want and what would create a stable society, but he would not agree with these things from any genuinely moral motive. At best, a group of psychopathic idealised agents would create a functioning society based solely on self-interest. In the next chapter we will see just how large this problem is and what options Railton has to resolve it.

---

[43] Nichols (2001), p 442.

In the previous two chapters I discussed two challenges to Railton's naturalism: Haidt's objection that we do not typically use reason for moral judgement-making and the objection that reason alone will not generate a genuinely moral judgement. In this chapter I will show that the second objection forces us to make important alterations to Railton's naturalism, and consider some wider implications for metaethics.

Haidt's experiments pose challenge for Railton. Our objective interests can be non-rational, so non-rational interests can be morally right. Idealised agents can rationally judge that our emotional, non-rational reactions against, say, *Cannibalism* outweigh anyone's interest in eating a corpse. This does not mean that such reactions always or even typically take priority. Idealised agents can judge that although an action is emotionally repulsive it is so beneficial that it is morally permissible. We can also argue that our emotional intuitions are based on unconscious reasoning, such as morphological rationalism, and that Haidt's scenarios are too artificial to be useful. Our moral judgement-making is hardly infallible so we should not be surprised if our emotions or unconscious reasoning, after condemning incest for years, register a false positive result with *Incest*. Even then, Feinberg et al shows that we can use reason to change our minds. It may also be so difficult to construct a moral system that allows rare cases like *Incest* but condemns general incest that for sheer practical reasons we can ignore them. Furthermore, without rational justifications it becomes puzzling how we can build consistent, coherent ethical systems that society can agree upon. Emotional reactions alone are too unstable for the task. Nevertheless, Haidt does show that emotions can be used in moral judgement-making. Other non-rational factors can also play a part, such as biological properties. We saw, for example, that the hormone oxytocin can affect how we care for others.

We can say a little more here about Railton's naturalism. First, it dealt with Haidt's experiments by taking account of the empirical evidence concerning intuitions. Railton's naturalism relies on intuitions having an underlying rational basis, which is an empirical claim. If psychological research later revealed that moral intuitions were made purely non-rationally, Railton could not use them to defend his naturalism. At present, though, it is a plausible response to Haidt. Empirical psychology supports Railton's naturalism here.

Second, we can say more about non-rational objective interests. If the idealised agent Beth+ has the psychology of the actual agent Beth, then Beth+ can potentially use non-rational or emotional powers to ascertain Beth's objective interests. For example, Beth+ would understand Beth's compulsion to care for her children, whether Beth's reasons were based on rational or non-rational processes. Beth+ would even share that compulsion were she in Beth's place because she would share Beth's emotions, and so Beth+'s understanding of the compulsion does not simply rely on the factual information Beth+ has. Beth+ would also emotionally understand Beth's desire to become a writer. However, because Beth+ rationally knows that Beth would not be successful, she would not think that Beth has an objective interest to become a writer. Such reasoning carries over to Beth's moral judgement-making, as a non-moral objective interest is automatically a candidate for being morally permissible (or obligatory) or not, regardless of whether it is formed rationally or non-rationally. Idealised agents would take our non-rational motivations seriously when deciding our objective interests, because that is just the sort of creature we are. We are biologically and psychologically inclined towards various interests that we are unable to give up, as we saw with the hormone oxytocin in chapter three. Even if it were more advantageous to ignore such interests it would be too distressing.

Why does Beth+ need to have Beth's psychology? Could she not just factually know about it, like a psychologist could know about a patient's personality? Railton's naturalism tells us to

'[g]ive to [Beth] unqualified cognitive and imaginative powers, and full factual and nomological information about [her] physical and psychological constitution, capacities, circumstances, history and so on. [Beth] will have become [Beth+], who has complete and vivid knowledge of [herself] and [her] environment, and whose instrumental rationality is in no way defective.'[1]

So Beth+ is not just a general idealised agent, one who could be an idealised agent for anyone. She is 'super-Beth', sharing all Beth's emotional powers and psychological attributes. Beth+ does not know about Beth's psychology simply because she examines a group of facts, but because she largely shares Beth's psychology.

---

[1] Railton (1986a), pp. 173-174.

This is fair enough when considering non-psychopathic actual agents, but things get trickier when we consider the psychopath. Can a psychopath's idealised self make genuinely moral judgements? In chapter four we saw that Railton seemed to think so. He discussed a formerly morally competent agent who no longer has the ability to sympathise with others. Railton thought that the agent could make moral judgements by following his past behaviour and copying other morally competent agents. Unfortunately this conflicted with McDowell's idea of a moral community. Moral outsiders cannot see the community's viewpoint or correctly use moral terms in novel situations. Whilst Railton's sympathy-impaired agent may be able to follow his past behaviour in familiar circumstances, it will not help when he faces an unfamiliar situation. Neither can he be sure of copying moral insiders' behaviour correctly. Assuming that the agent has become a moral outsider, he has lost the ability to make genuinely moral judgements. We also have a good reason why this is; he lacks the emotional powers to do so.

Consider again the definition of what is morally wrong according to Railton's naturalism:

"'X is [morally] wrong" means "We the people (i.e., people in general, including the speaker) would disapprove of allowing X as part of our basic scheme of social cooperation were we to consider the question with full information from a standpoint that considers the well-being of all affected without partiality.'"[2]

Given such a definition, it appears that sufficiently rational psychopaths can make genuinely moral judgements. However, a psychopath may recognise whether something was socially beneficial without being able to identify any moral property – as opposed to conventional property – about the answer. A society of perfectly rational psychopathic idealised agents might survive because of its members' self-interest in keeping it alive, and its mutually beneficial judgements would therefore be moral. This does not sound convincing; even if such a society lasts, the psychopaths' judgements do not look moral because everyone is in it for themselves. Clear-sighted selfishness might lead to mutual benefit, but that is no moral recommendation. One could object that such a society is impossible, since it would need stronger social bonds than perfectly rational psychopaths could have.

---

[2] Railton (1996), p 69.

However, this needs argument, as psychopaths can be quite successful in normal societies and can often cooperate with others, albeit for selfish motives.

One can get past this problem by focusing on the use of disapproval in the definition. To identify X as morally wrong, one not only needs to know the consequences of allowing X, but one must also disapprove of allowing X. What does 'disapprove' mean? Psychopaths may be able to work out the consequences of allowing X, but they would not respond emotionally in the correct way. Their disapproval would carry no emotional content (such as guilt, remorse and so on) that we would recognise as contributing to the creation of moral judgements. This means that they cannot disapprove of X in such a way as to fit Railton's definition of 'X is morally wrong', given that one needs various emotional powers to identify something as moral.

Taking this line means that we can reply to another objection that a hard-headed rationalist might make. The problem with psychopaths, claims the rationalist, is that they *can* make moral judgements, but they cannot recognise if a particular judgement is moral or not. Just because they are unaware that they are making a moral judgement, it does not mean that the judgement ceases to be moral. Psychopaths can rationally make moral judgements without identifying anything about their moral nature. Making moral judgements therefore does not depend on an agent's emotional powers. Emotional powers are only needed, if at all, to recognise that the judgement is moral. This objection rests on a distinction between moral judgements and moral agents. It states that one need not be a moral agent who can identify the moral features of a moral judgement in order to make that judgement, since the moral features of that judgement do not rely on the moral awareness of the agent. A moral judgement remains moral even if the agent making it cannot identify it as such. Psychopaths can therefore make moral judgements without being moral agents. However, given our proposed use of disapproval in Railton's analysis of moral rightness, the rationalist's objection is blocked. X is morally wrong, according to the speaker, if people in general *including the speaker* would disapprove of allowing X. If the speaker is a psychopath, then he is incapable of disapproving of X in a way that would make his judgement moral. Therefore he cannot truly judge anything to be morally wrong.

Railton has not specifically argued on these lines, but recent comments suggest that he now thinks that psychopaths cannot make moral judgements. For example, he states that '[p]sychopathy is thought by some to involve a… perhaps specifically emotional rather than cognitive or motor

empathy'.[3] He thinks that empathy is important for moral development and appears to concede that the psychopath's problem may be emotional. He has specifically discussed using affective processes (such as intuitions and empathy) to reach moral judgements, which are denied to the psychopath. So it appears that Railton now thinks that the psychopath cannot make moral judgements, at least partially due to non-cognitive deficits.

Railton has not linked this to his naturalism, though. We do not know how he would deal with the psychopath's idealised agent. There are two options:

1) The psychopath cannot make genuinely moral judgements, but his idealised self can.
2) Neither the psychopath nor his idealised agent can make genuinely moral judgements.

Option (1) looks untenable at first, as the idealised agent is only cognitively superior to the actual agent. In all other respects they are psychologically identical. Since the psychopath's moral deficit is not cognitive, the idealised self will also suffer from it and thus cannot make genuinely moral judgements. Railton does allow idealised agents to have 'unqualified… imaginative powers', but he does not discuss what he takes such powers to include, and we cannot assume that he would include emotional powers.[4] Furthermore, though the psychopath does have an impaired imagination, as he cannot always accurately determine how others feel, this ultimately appears to rest on an emotional deficit rather than an imaginative one. The reason for the imaginative failure is that the psychopath cannot feel certain emotions, and so cannot infer them in others.

This means, however, that Railton can go for option (1) by explicitly allowing idealised agents to have emotional powers. All he need say is that cognitive powers *do* include emotional powers, as the idealised agent has the ability to feel moral emotions and can therefore make moral judgements. This separates the idealised agent further from the actual agent psychologically (as the idealised agent will have extra emotional powers that the psychopath cannot have), but the idealised agent can still recognise what the psychopath's non-moral objective interests are. The psychopath may not know what his idealised self's moral judgements are, but that is his problem and not his idealised self's.

---

[3] Railton (2014), p 844.
[4] Railton (1986a), pp. 173-174.

However, this move is more problematic than it appears at first sight. In order to show why, I will first look at an objection from Connie S. Rosati. Her objection fails, but we can use the psychopath to create a related objection that works.

Rosati argues that actual agents and their idealised selves are so different that an idealised agent may not be someone who can truly identify the actual agent's objective interests.[5] Our abilities to appreciate certain facts depend on what we are like, and what our abilities and experiences have been. The idealised agent can appreciate every relevant fact to ascertain the actual agent's objective interests, including those that the actual agent cannot appreciate. If the actual agent could appreciate every relevant fact, this would be because he would have abilities and experiences that he currently has not had. However, if the actual agent had those extra abilities and experiences, he might become a completely different person. Some experiences can be life-changing. If one's personality changes enough, then presumably one's objective interests will change as well, as one will have different personal traits, such as different goals and abilities.[6] Idealised agents may be so different from actual agents because of these experiences that what they advise for the actual agent is not really based on the actual agent's personal traits and hence will not be his objective interests. Idealised agents are not really actual agents psychologically, because if actual agents became fully informed like idealised agents, their personalities and objective interests would change significantly.

Unfortunately this objection only works against accounts of ideal agents defined in terms of an agent's actual or near-actual responses. Suppose there was a theory that stated that what is tasty for an agent was what is agreeable to the idealised agent's taste. We can object to this theory by using a similar criticism to Rosati's. Idealisation would change the actual agent's personality so much that it could change what he would find tasty. The actual agent and his idealised self could have completely different senses of taste. Such an objection works here, but Railton's account is not of this type. Railton could reply that certainly the idealised agent needs some psychological connection with the actual agent to properly identify the actual agent's objective interests, but this psychological connection need not be as strong as Rosati assumes.

What is the psychological connection between Lonnie today (call him Lonnie-t) and Lonnie yesterday (call him Lonnie-y)? It is the shared history, shared memories, and so on. Lonnie-t need not be psychologically identical to Lonnie-y to identify Lonnie-y's objective interests. Lonnie-t need not

---

[5] Rosati (11995), p 307.
[6] Rosati (1995), p 308.

share Lonnie-y's objective interests either. Our objective interests can change over time. Now suppose that Lonnie-t woke up this morning and discovered that he had miraculously become an idealised agent. He still has the history and memories as Lonnie-y did, which means that he can understand Lonnie-y better than any other idealised agent. Only now Lonnie-t also has extra information. His objective interests might be very different to Lonnie-y's, but he can understand Lonnie-y's psychology and circumstances. Importantly, he has the same types of emotional powers that Lonnie-y has, even if they are now better than Lonnie-y's. There is nothing problematic about this, as we can see when we compare their rational powers. Both Lonnie-y and Lonnie-t may have the rational power to make mathematical calculations. Lonnie-t can do mathematics better because his rational power here is superior to Lonnie-y's, but is not a different *type* of rational power. The same goes for emotional powers; perhaps Lonnie-t would be happy more often than Lonnie-y, but Lonnie-y does possess the power to feel happiness if the opportunity presented itself. Lonnie-t can use these powers to imagine correctly and so fully understand what it is like to be Lonnie-y, which is essential for discovering Lonnie-y's objective interests. Lonnie-t has all the psychological connections to Lonnie-y that he needs in order to identify Lonnie-y's objective interests. Rosati has to do much more to show that adding extra powers and information will make Lonnie-t unable to recognise Lonnie-y's needs and objective interests.

Rosati's objection fails, but the psychopath provides a related problem which is more troubling. Suppose that John-y, a psychopath, goes to bed and wakes up as John-t. He discovers that he has miraculously become an idealised agent and so can make moral judgements. He looks back on what John-y was like, and shudders. 'What was I thinking?!' he cries out, appalled by his former callousness. This is not the same case as Lonnie's. The psychological connection between Lonnie-y- and Lonnie-t involved Lonnie-t being able to understand Lonnie-y's psychology because he could imagine what it would be like to share it. Lonnie-y had no type of emotional power that Lonnie-t did not – it just happens that Lonnie-t's powers are superior. John-t and John-y, though, have different types of emotional powers. John-t cannot fully understand or imagine what it is like to be a psychopath. He can predict what psychopaths will do or what is going through their heads, but this is very different from participating, even momentarily, *as* a psychopath. John-t cannot fully imagine what John-y is like, and if that is the case, there is no guarantee that he will be able to identify John-y's objective interests. Remember that idealised agents have full information about actual agents and so

can recognise the actual agent's objective interests at any given time. Normal people may be able to predict a psychopath's behaviour or needs some of the time by trying to ignore all moral considerations, but if we try to do this continuously we will at some point slip up because it is incredibly easy to unconsciously smuggle in moral concepts (This is partly why a psychopath's behaviour can be so shocking even if we know he is a psychopath.). A good analogy here is children and their parents. Parents have superior cognitive and emotional powers to toddlers and can often predict their toddler's behaviour and needs. They typically know as much as anyone about the child. Yet their predictions can go wrong because they cannot fully grasp what it is like to be their toddler. It might even be impossible for some reason to turn a particular power off once we attain it. Idealised agents end up having powers so greatly superior to the psychopath's that they cannot truly look through the psychopath's eyes, and hence cannot always identify the psychopath's objective interests.

Railton's idealised agents *must* be able to identify objective interests, though. This means that John-t must have the same emotional powers as John-y, and option (1) fails. If Railton says that a psychopath's idealised self can make moral judgements, then there is no guarantee, contrary to Railton's account of idealised agents, that the idealised self will be able to identify the actual agent's objective interests.

So let us go for option (2), then. The psychopath's idealised self is also psychopathic, and cannot make moral judgements. Railton must give up the idea that a group of idealised agents can always judge what is morally right. A society of psychopathic idealised agents may be able to survive together for reasons of self-interest, but any transgressions they recognise will be conventional. As I pointed out above, given Railton's analysis of 'X is morally wrong', psychopaths cannot make moral judgements. They may identify which actions are socially beneficial, but they cannot disapprove or approve of those actions in the appropriate way to make their judgements moral. They can understand social benefits in natural terms, but that will not help them understand or describe the benefits in moral terms, which echoes what John McDowell argued concerning moral outsiders in chapter four. Following on from this, we could say that in order for an idealised agent to identify moral properties, they must have certain emotional powers that allow them to do so and the actual agent *must have these types of powers too.* So Lonnie+, for example, has exactly the same types of emotional powers that Lonnie has. If these types of powers allow Lonnie to make moral judgements,

then Lonnie+ can do so as well, but if Lonnie cannot, then Lonnie+ cannot either. This avoids the difficulties encountered with option (1), at the cost of implying that idealised agents cannot make moral judgements if their actual selves cannot.

Railton may choose to live with that, but there is another sting in the tail. We can now come back to my discussion regarding cognitive powers in chapter two, when I wrote that I was going to work with the idea that cognitive powers do not include emotional powers. We can now see why this is essential if we go for option (2). If cognitive powers include emotional powers, then an idealised agent (who has unqualified cognitive powers) will be able to make moral judgements even if the actual agent cannot, and we wind up with option (1) again. The only way to go for option (2) is to deny that cognitive powers include emotional powers. Once we go for option (2), cognitive powers appear to be better defined as rational (non-emotional) powers. If we can know things via our emotional powers, this means that cognitive powers cannot be whatever powers we employ in order to know things. Consequently, Railton has to be careful about what he takes 'full information' about the actual agent to be, because there may be some information only accessible via emotional powers that the idealised agent lacks the emotional powers to understand or appreciate.

Nevertheless, though I do not have the space to probe these considerations more deeply, I prefer option (2) to option (1). Option (2) involves alterations to Railton's naturalism, but they look more achievable than the task of showing that a morally aware idealised agent can ascertain and understand a psychopathic actual agent's objective interests. In either case, though, Railton has significant gaps in his naturalism that he needs to fill.

As far as Railton's naturalism is concerned, we can now answer the central question in chapter one: Railton may be able to use empirical psychological research, but he must make changes because at the moment he takes insufficient notice of our emotional powers and the role they play in moral judgement-making. Empirical studies have an important effect on how we can identify Railton's moral properties. Note also exactly how we reached this conclusion. Haidt's social intuitionism states that moral judgements are largely a matter of emotional reaction rather than rational reflection, but it could not affect Railton. This was because the empirical evidence that Haidt used was not strong enough to back up his claims either that we typically make moral judgement purely emotionally, or that we should make moral judgements purely emotionally. The experiments in chapter four, starting with

Blair's 1995 experiments with psychopaths, succeeded tin showing that moral judgement-making has to involve emotional processes and that empirical rationalism was likely to be wrong because it could not adequately explain the evidence presented. Empirical psychology can therefore at least influence methodological moral naturalism, though one has to be careful about how one interprets the evidence obtained from experiments.

But could empirical psychology affect other metaethical theories? Although I used Railton's naturalism as a 'test subject', I could have used another theory. Take Cornell realism, a non-reductive naturalism. Like Railton, it holds that moral properties are natural; unlike him, it holds that we cannot reduce moral properties to non-moral natural properties. This does not stop empirical research potentially playing a role, because Cornell realists agree that changing non-moral natural properties can affect the moral properties connected with them. Non-reductive naturalisms can come in analytical and non-analytical flavours, and we saw in chapter 2 that analytical naturalisms have considerable problems with Moore's Open Question Argument, but that non-analytical naturalisms can avoid it. So the Cornell realist has good reason to go for a non-analytical naturalism, in which case it is hard to see why empirical considerations should not affect his position. Many Cornell realists seem to recognise this. Nicholas Sturgeon states that moral reasoning – 'like reasoning in the sciences' – does not rely on analytical truths or *a priori* foundations. Furthermore, moral explanations can be refined 'in the light of both empirical evidence and theoretical criticism'.[7] Geoffrey Sayre-McCord argues against the claim that 'moral theory is totally insulated from observational consequences'.[8] Richard Boyd's realism is explicitly amenable to empirical testing, as we saw in chapter one. Outside Cornell realism, the relativist realist Jesse Prinz has already discussed empirical psychology. He uses Millgram, Joshua Greene, Haidt, Kohlberg, Blair and others to provide grounding for his metaethical theory.[9] Non-naturalists can also potentially make use of empirical evidence if they endorse some sort of supervenience (as discussed in chapter 1), and many of them do endorse supervenience..

Outside of realism, non-cognitivists can try using empirical evidence to provide support for their theories, which would mean that moral properties do not really exist. The expressivist Allan Gibbard says that the difference between himself and Railton 'may well be empirical'.[10] There is nothing

---

[7] Sturgeon (1985), p 242.
[8] Sayre-McCord (1988a), p 256.
[9] Prinz (2007), pp. 21-22, 24-25, 29-32, 32-35, 43-46.
[10] Gibbard (1990), p 122, fn 15.

strange about accepting methodological naturalism (e.g. studying morality through empirical means) whilst rejecting substantive naturalism (e.g. denying that moral properties exist). Maybe a more detailed analysis of moral judgement-making would show that moral properties are highly unlikely to exist. This means that moral error-theorists, who hold that moral discourse tries but fails to be factually correct (because there are no moral properties) can also consider empirical research. Richard Joyce, for example, uses psychological and evolutionary research to argue that our moral development gives us no reason to believe that moral properties exist.[11]

In this thesis I have concentrated specifically on Peter Railton's moral realism, and psychological research on moral judgements. However, I have shown that empirical evidence can be used to affect and alter a particular metaethical area, and there is no reason to think we cannot use empirical research more broadly to examine the strengths and weaknesses of other metaethical theories.

---

[11] Joyce (2006).

Bibliography

Albright, Linda, Kenny, David and Malloy, Thomas, (1988), 'Consensus in Personality Judgments at Zero Acquaintance', *Journal of Personality and Social Psychology* 55(3), pp. 387-395

American Psychiatric Association, (2013), *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition, Washington D.C.: American Psychiatric Publishing

Audi, Robert, (2002), *The Architecture of Reason: The Structure and Substance of Rationality*, Oxford: Oxford University Press

Ayer, A.J., (1936), *Language, Truth and Logic*, London: Penguin

Baldwin, Thomas, (1993), Editor's Introduction to Moore (1993)

Baron-Cohen, Simon, (2011), *Zero Degrees of Empathy*, London: Allen Lane

Baron-Cohen, S., Leslie, A.M and Fritch, U., (1985), 'Does the autistic child have a theory of mind?', *Cognition* 57, pp. 27-36

Blackburn, Simon, (1998), *Ruling Passions*, Oxford: Oxford University Press

Blair, R.J.R., (1995), 'A Cognitive Developmental Approach to Morality: Investigating the Psychopath', *Cognition* 57, pp. 1-27

Blair, R.J.R., (1996), 'Brief Report: Morality in the Autistic Child', *Journal of Autism and Developmental Disorders* 26(5), pp. 571-579

Blair, R.J.R., (1997), 'Moral Reasoning and the Child with Psychopathic Tendencies', *Personality and Individual Differences* 22(5), pp. 731-739

Blair, R.J.R., (1999a), 'Psychophysiological responsiveness to the distress of others in children with autism', *Personality and Individual Differences* 26(3), pp. 477-485

Blair, R.J.R., (1999b), 'Responsiveness to distress cues in the child with psychopathic tendencies'', *Personality and Individual Differences* 27(1), pp. 135-145

Blair, R.J.R, Sellars, C., Strickland, I., Clair, F., Williams, A.O., Smith, M. and Jones, L., (1995), 'Emotion Attributions in the Psychopath', *Personality and Individual Differences* 19(4), pp. 431-437

Boniolo, Giovanni and De Anna, Gabriele (eds.), (2006), *Evolutionary Ethics and Contemporary Biology*, Cambridge: Cambridge University Press

Boyd, Richard, (1988), 'How to Be a Moral Realist', in Sayre-McCord (1988b)

Brink, David, (1989), *Moral Realism and the Foundation of Ethics*, Cambridge: Cambridge University Press

Churchland, Patricia, (2011), *Braintrust*, Princeton: Princeton University Press

Copp, David, (1990), 'Explanation and Justification in Ethics', *Ethics* 100(2), pp. 237-258

Darwall, Stephen, Gibbard, Allan and Railton, Peter, (1992), 'Towards *Fin de siècle* Ethics: Some Trends', *Philosophical Review* 101(1), pp. 115-189

Dawkins, Richard, (2006), *The Selfish Gene*, 20th Anniversary Edition, Oxford: Oxford University Press

Eisenberg-Berg, Nancy, (1979), 'Development of Children's Prosocial Moral Judgment', *Developmental Psychology* 15(2), pp. 128-137

Gibbard, Allan, (1990), *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, Massachusetts: Harvard University Press

Greene, Joshua, (2007), 'The Secret Joke of Kant's Soul' in Sinnott-Armstrong, Walter (ed.), *Moral Psychology Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders and Development*, Cambridge, Massachusetts: MIT Press

Haidt, Jonathan, (2001), 'The Emotional Dog and the Rational Tail: A Social Intuitionist Response to Moral Judgment', *Psychological Review* 108(4), pp. 814-834

Haidt, Jonathan, (2012), *The Righteous Mind*, New York: Penguin

Haidt, Jonathan, Björklund, Fredrik and Murphy, Scott, (2000), 'Moral Dumbfounding: When Intuition Finds No Reason', unpublished manuscript

Horgan, Terence and Timmons, Mark, (1992), 'Troubles on Moral twin Earth: Moral Queerness Revived', *Synthese* 92(2), pp. 221-260

Horgan, Terence and Timmons, Mark, (2000), 'Copping out on Moral Twin Earth', *Synthese* 124(1), pp. 139-152

Horgan, Terence and Timmons, Mark, (2007), 'Morphological Rationalism and the Psychology of Moral Judgment', *Ethical Theory and Moral Practice* 10(3), pp. 279-295

Hume, David, (2000 [1739-1740]), *A Treatise of Human Natures*, eds. David Fate Norton and Mary J Norton, Oxford: Oxford University Press

Jacobsen, Daniel, (2012), 'Moral Dumbfounding and Moral Stupefaction' in Timmons, Mark (ed.), *Oxford Studies in Normative Ethics Volume 2*, Oxford: Oxford University Press

Jarvis Thomson, Judith, (1985), 'The Trolley Problem', *Yale Law Journal* 94(6), pp. 1395-1415

Jost, John, (2012), 'Left and Right, Right and Wrong: The Politics of Morality', http://301a32ba31149ad50392-db76e8c8d94f3afdf40f06644f9ca222.r22.cf1.rackcdn.com/wp-content/uploads/2012/09/Jost-book-review-of-Haidt-for-the-Monkey-Cage.pdf, accessed on 14 November 2013

Joyce, Richard, (2007), *The Evolution of Morality*, Cambridge, Massachusetts: MIT Press

Kohlberg, Lawrence, (1973), 'The Claim to Moral Adequacy of a Highest Stage of Moral Judgment', *Journal of Philosophy* 70(18), pp. 630-646

McDowell, John, (1998), *Mind, Value and Reality*, Cambridge, Massachusetts: Harvard University Press

McPherson, Tristram, (2012), 'Ethical Non-Naturalism and the Metaphysics of Supervenience', in Shafer-Landau, Russ (ed.), *Oxford Studies in Metaethics, Volume 7*, Oxford: Oxford University Press

Miller, Alexander, (2013), *Contemporary Metaethics: An Introduction*, second edition, Cambridge: Polity Press

Millgram, Stanley, (1963), 'Behavioral Study of Obedience', *Journal of Abnormal and Social Psychology* 67(4), pp. 371-378

Moore, G.E., (1993 [1903]), *Principia Ethica*, revised edition, Cambridge: Cambridge University Press

Nichols, Shaun, (2001), 'Mindreading and Cognitive Architecture Underlying Altruistic Motivation', *Mind and Language* 16(4), pp. 425-455

Nichols, Shaun, (2002), 'How Psychopaths Threaten Moral Rationalism: Is it Irrational to Be Amoral?', *The Monist* 85(2), pp. 285-303

Nisbett, Richard and Wilson, Timothy, (1977), 'Telling More Than We Can Know: Verbal Reports on Mental Processes', *Psychological Review* 84(3), pp. 630-646

Pinker, Steven, (2011), *The Better Angels of Our Nature*, New York: Penguin

Prinz, Jesse, (2007), *The Emotional Construction of Morals*, Oxford: Oxford University Press

Prinz, Jesse, (2011), 'Is Empathy Necessary for Morality?', in Coplan, Amy and Goldie, Peter (eds.), *Empathy: Philosophical and Psychological Perspectives*, Oxford: Oxford University Press

Putnam, Hilary, (1975), 'The Meaning of "Meaning"', in his *Mind, Language and Reality: Philosophical Papers Volume 2*, Cambridge: Cambridge University Press

Railton, Peter, (1986a), 'Moral Realism', *Philosophical Review* 95(2), pp. 163-207

Railton, Peter, (1986b), 'Facts and Values', reprinted in his *Facts, Values and Norms*, (2003), New York: Cambridge University Press

Railton, Peter, (1989), 'Naturalism and Prescriptivity', *Social Philosophy and Policy* 7(1), pp. 151-174

Railton, Peter, (1996), 'Moral Realism: Prospects and Problems', in Sinnott-Armstrong, Walter and Timmons, Mark (eds.), *Moral Knowledge? New Readings in Moral Epistemology*, New York: Oxford University Press

Railton, Peter, (2014), 'The Affective Dog and Its Rational Tale: Intuition and Attunement', *Ethics* 124(4), pp. 813-859

Richell, R.A., Mitchell, D.G.V., Newman, C., Leonard, A., Baron-Cohen, S. and Blair, R.J.R., (2003), 'Theory of mind and psychopathy: can psychopathic individuals read the 'language of the eyes'?', *Neuropsychologia* 41, pp. 523-536

Rosati, Connie S., (1995), 'Persons, Perspectives and Full Information Accounts of the Good', *Ethics* 100(2), pp. 296-325

Rozin, Paul, Millman, Linda and Nemeroff, Carol, (1986), 'Operation of the Laws of Sympathetic Magic in Disgust and Other Domains', *Journal of Personality and Social Psychology* 50(4), pp. 703-712

Ruse, Michael, (1988), 'Evolutionary Ethics: Healthy Prospect or Last Infirmity?', *Canadian Journal of Philosophy*, supplementary volume 14, pp. 27-73

Sayre-McCord, Geoffrey, (1988a), 'Moral Theory and Explanatory Impotence', in Sayre-McCord (1988b)

Sayre-McCord, Geoffrey (ed.), (1988b), *Essays on Moral Realism*, Ithaca: Cornell University Press

Shafer-Landau, Russ, (2003, *Moral Realism: A Defence*, Oxford: Oxford University Press

Sinnott-Armstrong, Walter, (2006), 'Moral Intuitionism Meets Empirical Psychology' in Horgan, Terry and Timmons, Mark (eds.), *Metaethics After Moore*, Oxford: Oxford University Press

Smetana, Judith, (1981), 'Preschool Children's Conceptions of Moral and Social Rules', *Child Development* 52, pp. 1333-1336

Smetana, Judith, (1985), 'Preschool Children's Conceptions of Transgression: Effects of Varying Moral and Conventional Domain-Related Attributes', *Developmental Psychology* 21(1), pp. 18-29

Street, Sharon, (2006), 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127(1), pp. 109-166

Sturgeon, Nicholas (1985), 'Moral Explanations', reprinted in Sayre-McCord (1988b)

Trivers, Robert, (1971), 'The Evolution of Reciprocal Altruism', *Quarterly Review of Biology* 46(1), pp. 35-57