A Critique of *A Priori* Moral Knowledge

Ashley Alexander Lane

Submitted in fulfilment of the requirements for the
degree of Doctor of Philosophy

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

July 2018

Contents

<u>Abstract</u>

Many ethicists believe that if it is possible to know a true moral proposition, it is always possible to ascertain *a priori* the normative content of that proposition. I argue that this is wrong; the only way to ascertain the normative content of some moral propositions requires the use of *a posteriori* information.

I examine what I call determinate core moral propositions. I assume that some of these propositions are true and that actual agents are able to know them. Ethicists whom I call core-apriorists believe that it is always possible to ascertain *a priori* the normative content of such propositions. Core-aposteriorists believe that this is false, and that sometimes *a posteriori* information must be used to ascertain that normative content.

I develop what I call the *a posteriori* strategy to show that core-apriorists are likely to be wrong, and so core-aposteriorists are correct. The strategy examines the details of particular core-apriorist theories and then shows that the theories have one of two problems: either some of the knowable determinate core moral propositions in the theories are not knowable *a priori*, or some of the propositions are not determinate, so they cannot perform the epistemological work required of them. Therefore, some knowable determinate core moral propositions are only knowable with the aid of *a posteriori* information.

I apply the strategy to four different core-apriorist theories. The first is Henry Sidgwick's theory of self-evident moral axioms, as recently developed by Katarzyna de Lazari-Radek and Peter Singer. The second is Matthew Kramer's moral realism. I then examine Michael Smith's moral realism, and Frank Jackson and Philip Pettit's moral functionalism. The *a posteriori* strategy shows that there are serious difficulties with all four theories. I conclude that it provides good evidence that the core-apriorist is mistaken, and that the core-aposteriorist is right.

<u>Declaration</u>

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:

Acknowledgements

A huge amount of thanks must first go to my thesis supervisor, Hallvard Lillehammer, whose advice and criticisms have made this thesis much better than it otherwise would have been. Within the Birkbeck philosophy department, Cristian Constantinescu and Michael Garnett were also kind enough to discuss various sections of the thesis with me.

I would also like to thank the Birkbeck philosophy department in general for my entire time as a philosophy postgraduate student. I studied for both an MPhilStud and a PhD at Birkbeck, and during my MPhilStud I studied in the evenings whilst working full-time during the day. I also had to take time off from the MPhilStud course for medical reasons. The department was very supportive, which was not only essential for my MPhilStud work, but also for my progression to PhD study in the first place.

I have presented various drafts of sections of the thesis both inside and outside Birkbeck. I thank all the Birkbeck students and staff with whom I have been able to discuss my work. Outside Birkbeck, I have presented part of chapter 4 at the Spring 2018 London Graduate Philosophy Workshop, and part of chapter 5 at the VIII Braga Meetings on Ethics and Political Philosophy at the University of Minho. The participants at these conferences have given me invaluable help. Following the VIII Braga Meetings, the journal *Ethics, Politics and Society* has published my paper based on my presentation, 'Are Moral Functionalism's Moral *A Priori* Commitments Really *A Priori*?' (2018, volume 1).

Thanks to all my friends who, in many different ways, made sure I did not just bury myself under journal articles for three years without a break. I am afraid they are too numerous to name individually, but I have to give special thanks to Hazel Covill for proof-reading the thesis.

Major thanks must go to my parents and my family, who have provided all sorts of support from the very beginning which I could not have done without. I must mention my stepfather in particular. When I was fourteen, I was already trying to decide what to study at university. He suggested philosophy, and he gave me his copy of Bertrand Russell's *A History of Western Philosophy*. So he helped set the ball rolling; after that, he was at my BA graduation, he read my MPhilStud thesis, and he saw the start of my PhD work. Unfortunately, he died earlier this year. I dedicate this thesis to his memory.

<u>**Chapter 1 – Introduction**</u>

In this thesis, I want to argue for the following epistemological claim:

Central Claim: There is at least one determinate core moral proposition the normative

       content of which can only be ascertained as true by using *a posteriori* or

       mixed methods.

Many ethicists who endorse the possibility of agents having moral knowledge have held moral epistemological theories that would reject the claim.[1] Such theories imply that the normative content of determinate core moral propositions, if it is knowable, can always be known *a priori* (for example, through a process of *a priori* reasoning). I want to argue that this is mistaken, by demonstrating that theories that reject the Central Claim land themselves in serious difficulties.

 In this introduction, I first define the terms used in the Central Claim to show more clearly what it states. I will frame the debate as being between core-aposteriorists, who accept the Central Claim, and core-apriorists, who reject it. I will also give some reasons why rejecting the Central Claim may look plausible. In chapter 2, I will outline the strategy that I will use to defend the Central Claim. This strategy requires the examination of theories that reject the Central Claim, and I carry out this examination in chapters 3-5. Finally, I will briefly consider what the consequences of accepting the Central Claim are for both core-aposteriorists and core-apriorists.

<u>1 – Normative content and the *a posteriori*</u>

Moral propositions are propositions that express moral norms. They therefore have normative content. Assume that the moral proposition 'Drink-driving is morally impermissible' is true. To know that it is true, we may need to know that drink-driving causes fatal accidents. It appears, though, that accident statistics are just descriptive, non-normative information. Consequently, one could coherently accept that drink-driving causes fatal accidents but hold that it is

---

[1] E.g. Kant (1996); Ross (2002); Shafer-Landau (2003); Kramer (2009); Parfit (2011).

nevertheless morally permissible to drink-drive. To accept the proposition as true, we need normative premises that entail (for example) that causing fatal accidents is morally impermissible. These premises justify the normative content of the proposition.

It is very popular to think that the normative content of a moral proposition can be ascertained as true *a priori*. However, there has been a significant push over the last few years for the claim that the normative content of moral propositions is, or is only, ascertainable *a posteriori*. The philosophers defending this idea often claim that we can use evidence from psychology, anthropology and other such disciplines to attain *a posteriori* knowledge of such normative content (From now on, I will use *a priori/a posteriori* moral knowledge' to mean '*a priori/a posteriori* knowledge of the normative content of true moral propositions'.).[2]

Whilst I will not be too concerned with the exact definitions of '*a priori*' and '*a posteriori*', I do need to say a little about them. Some philosophers explicitly claim that they are identical to the concepts 'non-empirical' and 'empirical' respectively, but there is no clear agreement on this.[3] For example, Immanuel Kant claims that *a priori* knowledge is knowledge attained completely independently of experience.[4] Philip Kitcher points out that if Kant is right, this appears to imply that only innate knowledge can be *a priori*. This is problematic, because there is no reason why the endorsement of *a priori* knowledge should entail endorsing innate knowledge.[5] Kitcher's own analysis entails that 'if a person knows a priori that *p* then she could know that *p* whatever sufficiently rich experience she had had' in her life.[6] Whatever life the agent lived, there would be a similar process to that in her actual life that would allow her to know that *p*.[7] This separates *a priori* knowledge from empirical knowledge, since in cases of empirical knowledge, there may be possible worlds in which the agent lives and has a sufficiently rich experience, but where there are no similar processes that allow her to know empirically the propositions she knows in her actual life.

Another analysis, given by Philip Stratton-Lake, holds that a proposition is *a priori* knowable if it is knowable 'solely on the basis of an understanding of [the proposition]'.[8] Again, on this analysis, empirical knowledge is not *a priori*. One cannot know that one is looking at something

---

[2] E.g. Boyd (1988); Nichols (2004); Prinz (2007); Flanagan, Sarkissian and Wong (2008).
[3] E.g. Kramer (2009), p. 47.
[4] Kant (1998), A3/B7-A6/B10.
[5] Kitcher (1984), p. 21.
[6] Kitcher (1984), p. 24.
[7] Since I am only using Kitcher's analysis of the *a priori* as an example, I will not examine exactly what 'sufficiently rich experience' means.
[8] Stratton-Lake (2002), p. xliv.

blue simply by understanding the proposition 'I am looking at something blue'. One must use perception as well. I am not going to subscribe wholesale either to Kitcher's analysis or Stratton-Lake's analysis, but I do want to retain the idea that empirical knowledge is not *a priori* knowledge. This implies that empirical knowledge is a type of *a posteriori* knowledge, even if these two types of knowledge are not identical. Given that core-aposteriorists frequently hold that one can ascertain the normative content of at least some moral propositions only by using empirical knowledge, this is all they need to distinguish their position from the core-apriorist's. According to such core-aposteriorists, it is not always possible to ascertain the normative content of a determinate core moral proposition without the use of empirical information (and consequently, it cannot be known without the use of an *a posteriori* or mixed method).

A priori and *a posteriori* propositions are propositions which, if they are true, can be known *a priori* or *a posteriori* respectively. We should note that it may be possible to know the same proposition both *a priori* and *a posteriori*. Calling a proposition 'an *a priori* proposition' merely means that it is knowable *a priori*. It says nothing about whether it is also knowable *a posteriori*. The same is true, *mutatis mutandis*, for *a posteriori* propositions.

A method is simply a mode of access by which one can ascertain the truth of the normative content of a moral proposition. I will place no restrictions here on what this mode of access might be; it could be a process of practical reasoning, some form of empirical discovery, recognition of the truth as self-evident in some way, or something else.

If the normative content of a proposition can be ascertained as true using an *a priori* method, it is possible to ascertain that truth *a priori*. This implies that the normative content has an *a priori* status, even if it can also be known *a posteriori*. Similarly, if the normative content can be known by using an *a posteriori* method, its truth can be ascertained *a posteriori* and it has an *a posteriori* status.[9]

It is important to remember that *a priori* and *a posteriori* methods are not the only types of methods available. One may be able to discover a truth *a priori*, purely from the armchair, or *a posteriori*, for example through scientific examination of the physical world. However, knowledge of certain propositions may require a combination of these methods. For example, suppose that we can only discover that proposition A is true via an argument that requires

---

[9] The *a priori*/*a posteriori* status of a particular claim is therefore determined by whether it can be known via an *a priori* method or an *a posteriori* method respectively.

premises B and C. Premise B can only be discovered *a priori*, and premise C can only be discovered *a posteriori*. We cannot discover that A is true simply via an *a priori* method, because of C, or via an *a posteriori* method, because of B. Nevertheless, since there is an argument for A, there is a method to attain knowledge of A. This method is a mixed method, as it uses a combination of *a priori* and *a posteriori* methods.

A mixed method to ascertain the truth of the normative content of a moral proposition would therefore be a method that uses a combination of *a priori* and *a posteriori* methods to ascertain the truth. Whilst I will frequently talk of *a posteriori* methods in this thesis rather than *a posteriori* and mixed methods, this is merely for convenience. My defence of the Central Claim does not imply that mixed methods are impossible or even unlikely, and I will return to this point in the conclusion. The defence does not entail that we can or should remove mention of mixed methods from the Central Claim.

Furthermore, also for convenience, I will usually say that there are *a priori*, *a posteriori* and mixed methods to know a moral proposition, rather than say that the methods are methods to ascertain the truth of the normative content of a moral proposition.

2 – Core moral propositions

The Central Claim is about core moral propositions. The term 'core moral proposition' is something of a placeholder, because different moral theories will flesh out in different ways what the exact content of a core moral proposition is. However, core moral propositions comply with the Core Moral Proposition Criterion ('CMPC'):

CMPC: A proposition is a core moral proposition iff it is part of the set of propositions that is the smallest possible set of determinately true propositions that can justify on its own the determinately true normative content of every moral proposition. Whilst no individual core moral proposition necessarily justifies the determinately true normative content of every moral proposition, or is necessarily part of the justification of that normative content, there are

necessarily core moral propositions within that set that do, either individually or collectively.

Whilst 'core moral proposition' is my own term, the idea frequently occurs amongst ethicists. For example, Henry Sidgwick has self-evident axioms, Michael Smith has *a priori* moral platitudes, and Matthew Kramer has basic moral principles.[10] All of these are core moral propositions. This is why it is difficult to specify precisely what the exact content of a core moral proposition is, because philosophers such as Sidgwick, Smith and Kramer all make different claims about what that content is.

A good way of thinking about core moral propositions is that we can use them as normative premises in arguments for other moral propositions, and that they provide *fundamental* reasons, explanations or grounds for the truth of the normative content of non-core moral propositions.[11] We can give a non-moral example here of a fundamental reason: suppose I know that my train is not running because the train operator has cancelled it. I know a reason why the train is not running, but this might not be the fundamental reason. The fundamental reason may be that the train's engine has broken down, forcing the train operator to cancel it. Similarly, core moral propositions provide us with the fundamental reason why the normative content of certain non-core moral propositions are true. They justify the non-core moral propositions, in the sense that there is a good argument from the set of the core moral propositions to the conclusion that the normative content of the non-core moral propositions is true.[12] Suppose that 'Physical attacks on sentient beings for revenge are morally impermissible' is a core moral proposition. If I want to know the reason why I may not beat someone up because they have cheated me financially, this proposition may provide the fundamental reason: people are sentient beings, and beating someone up is a form of physical attack. It is therefore morally impermissible.

One reason why many ethicists are core-apriorists, who favour *a priori* methods over *a posteriori* and mixed methods, and who therefore reject the Central Claim, relates to a long-standing metaethical problem: how we can attain normative knowledge at all. The thought is that we can use *a posteriori* information to attain knowledge in many situations, but all such

---

[10] I will discuss all these ideas later in the thesis.
[11] For the purposes of this thesis, I treat 'reasons', 'explanations' and 'grounds' as interchangeable.
[12] I discuss justification and knowledge more in section 4 of this chapter. A good argument here is an argument (deductive or otherwise) that provides good evidence for a true conclusion, even if it does not guarantee that the conclusion is true.

knowledge appears to be descriptive, not normative. We can discover *a posteriori* how humans act, what they believe and how they form moral judgements, but that will not tell us anything about how they *should* act, what they *should* believe, and how to form *correct* moral judgements. *A posteriori* evidence can only contain descriptive information that yields descriptive conclusions, not normative conclusions (or at least not normative conclusions that are moral). So, the thought continues, we can only attain moral knowledge *a priori*, since it is impossible to do so *a posteriori*. This line of argument can be fleshed out in different ways. For example, Derek Parfit claims that certain normative claims cannot be identical to empirical causal and psychological claims because they cannot have the same meaning.[13] Evidence for the empirical claims thus cannot be identical to the evidence for the normative claims. Another core-apriorist, Patrick Clipsham, argues that disagreements about moral claims are intelligible even when there is no way to resolve them empirically.[14] Empirical evidence alone, being descriptive and non-normative evidence, will not let us know if the claims are true or false.

The idea that moral propositions are *a priori* goes back a considerable distance. Possibly the most famous example comes from Immanuel Kant in 1785. Kant tries to base his moral theory on the idea that we can ascertain what is morally right through a process of practical reasoning. He talks of imperatives, which express a command of reason that the will must obey – that is, the will of a practically rational agent must conform to the commands, otherwise the agent is not practically rational. Categorical imperatives command an agent to avoid actions that are intrinsically wrong, no matter what the agent (or anyone else) personally desires to be the case.

In fact, Kant only allows for one categorical imperative:[15]

CI: Act only in accordance with that maxim through which you can at the same time will that it become a universal law.[16]

The practically rational agent can ascertain the categorical imperative *a priori*, and then use it to guide her behaviour. For example, consider a man who borrows money, having promised to repay it later even though he knows that he cannot. He makes a promise that he plans to break

---

[13] Parfit (2011), pp. 325-327 (see also pp. 332-341).
[14] Clipsham (2014).
[15] Kant (1996), sections 4.415-4.421. Kant gives more than one formulation of the categorical imperative, but claims that they are all equivalent (section 4.436). It is open to debate whether they really are.
[16] Kant (1996), section 4.421.

because he desperately needs the money, According to Kant, this breaches the categorical imperative, because rational agents cannot coherently will that the maxim 'Break your promises when it is necessary for your own welfare' be a universal law. The practice of promising presupposes that promises will be reasonably believed. If the maxim were a universal law, then no promises could be reasonably believed, and so the practice of promising would not exist. This entails that the maxim is asking us to do the impossible: one cannot break a promise if there are no such things as promises in the first place. Such a universal law would contradict itself, and therefore borrowing money by making promises that you will break because of desperate need will contravene the categorical imperative.[17] As we can discover the contradiction without appealing to any *a posteriori* evidence, we can work out that making false promises when one needs to for one's welfare is morally wrong *a priori*.

All moral propositions must comply with the categorical imperative. The categorical imperative therefore helps to justify or refute the normative content of a moral proposition *a priori*. A Kantian may thus endorse a proposition such as 'You should always act in accordance with CI' as an *a priori* core moral proposition. The core moral proposition implies that moral agents cannot accept as true moral propositions that endorse maxims that cannot be willed as universal laws.[18]

In the 1930s, W.D. Ross developed another famous moral theory that endorses *a priori* moral knowledge. He claims that it is a mistake to examine *a posteriori*

'a variety of types of act that are commonly called right [in order] to find, or argue that they have some characteristics in common, e.g. that of being comparatively highly evolved, and then to assume that this is what 'right' or 'obligatory' means.'[19]

Even if everyone acts as though a certain property is a moral property because of evolutionary influence on our moral views, that does not mean that the moral property is identical with that property, or that we can correctly identify moral properties by examining evolutionary history.[20]

---

[17] Kant (1996), section 4.422.
[18] I should note that I use the categorical imperative as an example here, and I will not argue against Kantian ethics as an instance of a core-apriorist theory in this thesis. Given the large amount of literature on Kantian ethics, it would be difficult to do it justice here for reasons of space. My defence of the Central Claim will instead focus on examining more contemporary examples of core-apriorist theories (in chapters 3-5). However, if my defence is successful, then it should be successful against Kantian ethics as a core-apriorist theory as well.
[19] Ross (1939), p. 12.

Ross thought that we have what he calls self-evident *prima facie* duties, which give us insight into the 'broad principles of morality'.[21] This insight will not tell us what is morally required on every occasion, but it will tell us how we should behave in general. 'Self-evident' means 'evident without any need of proof, or of evidence beyond itself'.[22] To discover the truth of a self-evident proposition, all one needs to do is understand what it claims, and then one will know without further argument that it is true. Self-evident truths include those relating to various moral duties, logical validity and mathematical axioms. So, for example, it is a self-evident principle that there is generally a duty to keep promises, and we can know the principle *a priori*. Therefore the proposition 'One should generally keep one's promises' is an example of an *a priori* core moral proposition.

More recently, we find that moral theories that endorse *a priori* moral knowledge are still going strong. Derek Parfit writes that we have 'an a priori guarantee' that some things are intrinsically good, and cannot be explained as good by reference to something else.[23] Parfit compares moral truths to mathematical truths, with the implication that the two types of truth work in the same way. For example, it is true that 2 is greater than 1, but mathematical truths do not exist 'in an ontological sense'. They do not actually or even possibly exist. When we say that no number exists that is both odd and even, what we are doing is saying that such a concept is inconsistent or contradictory. We are not making an ontological claim about numbers.[24]

Since there is no ontological claim, one can argue *a priori* for the existence of mathematical truths, because 'existence' here carries no ontological implications. To argue that something ontologically exists, one may have to discover it *a posteriori*, for example in the laboratory. But since mathematical truths do not ontologically exist, there is no reason to think that we can or must discover them *a posteriori*.[25]

Parfit claims that even in an empty universe in which nothing ontologically existed (stars, planets, people, etc.), some irreducibly normative truths would still exist, but in a non-ontological sense, otherwise the universe would not be empty.[26] Moral propositions about such moral truths are propositions about intrinsic value, not instrumental value. It is difficult to see how anything

---

[20] Ross (1939), p. 14.
[21] Ross (2002), p. 14.
[22] Ross (2002), p. 29. This does not mean that self-evident propositions are obvious or easy to know.
[23] Parfit (2011), pp. 430-431.
[24] Parfit (2011), pp. 479-481.
[25] Parfit (2011), p. 485.
[26] Parfit (2011), p. 486.

could be instrumentally good in an empty universe, because there is nothing that it could be good for. Such propositions about intrinsic value are good candidates for being *a priori* core moral propositions.

Parfit also compares morality to mathematics when writing of necessary truths. He claims that mathematical truths are necessary truths, and we do not discover these truths by using abilities such as perception. We can discover these truths merely by thinking about them.[27] Similarly, we can do so with normative truths. We have reason to believe that 2+2=4 whatever our world is like, and to care about certain things whatever our world is like. Parfit comments that '[s]ince our fundamental normative beliefs are not contingent features of the world, we do not need to have empirical evidence for their truth'.[28]

There are many other examples of contemporary metaethical theories that endorse *a priori* core moral propositions. However, we must remember that some metaethical theories do not endorse *any* core moral propositions, or even allow that moral knowledge is possible. For example, emotivist theories reject the possibility of moral knowledge entirely. Moral propositions do not have truth values, so there are no methods to ascertain the truth of their normative content. I want to focus exclusively on the debate between core-apriorists and core-aposteriorists, so I will make two assumptions in this thesis: at least some moral propositions are true, and it is not exceptionally difficult for actual agents to know them. Most core-apriorists and core-aposteriorists share these assumptions, and in this thesis I will only consider those who do.

One possible problem with claims that core moral propositions are *a priori* concerns analyticity. One could argue that *a priori* propositions are analytic, and that this entails that core moral propositions cannot be *a priori*.

We can know whether an analytic proposition is true or false by examining the meaning of the terms in the proposition. The proposition 'If Bert is a bachelor, then Bert is unmarried' is analytic, because we can check its truth by unpacking the definition of 'bachelor'. Synthetic propositions are not analytic. The proposition 'If Ernie is a bachelor, then Ernie is unhappy' is synthetic, as the definition of 'bachelor' does not include the criterion that bachelors are

---

[27] Parfit (2011), p. 489.
[28] Parfit (2011), pp. 517-518.

unhappy. The distinction between the analytic and the synthetic is not entirely uncontroversial, but we need not investigate the matter here.[29]

The trouble is that many analytic propositions are what I will call 'vacuous'. Suppose that 'Cruelty is bad' is an *a priori* core moral proposition. If we include in the definition of cruelty that it is bad, then the proposition says that something bad is bad. This gives no new information about the badness of cruelty, so it is difficult to see how the proposition can justify any further moral claim. Since it just says that something bad is bad, we can do without it.

Consequently, it appears that the core-apriorist must hold that *a priori* core moral propositions are synthetic.[30] However, Frank Jackson and Philip Pettit have made a rescue attempt for *a priori* propositions that I will call 'non-vacuously analytic'. Propositions such as 'The property of fairness is whatever plays the role of having characteristics X, Y and Z' (where we define X, Y and Z in a particular way) are analytic *a priori*. But they are not vacuous, because they help identify what natural property in the world is the property of fairness.[31] Since such propositions are analytic, but can provide new information, they are non-vacuously analytic.

Later, I will argue that the propositions that Jackson and Pettit identify as non-vacuously analytic are not *a priori*, so the core-apriorist cannot use their metaethical theory to refute the Central Claim. For now, though, I will grant for the sake of argument that there may be *a priori* non-vacuously analytic propositions. I will not object to core-apriorist theories on the basis of analyticity.

3 – Determinate moral propositions and determinate truth values

The Central Claim relates to determinate propositions, and the CMPC relates to determinately true propositions. Importantly, these types of determinacy are not the same. Determinately true propositions have determinate truth values; that is, they have one unique truth value. For example, Fermat's Last Theorem is either true or false. We may never have found its truth value, but its truth value would still be determinate. Propositions such as 'He is tall' can have indeterminate truth values, since whilst some men are clearly tall and others are clearly not,

---

[29] E.g. Quine (1980).
[30] For example, Kant claims that the categorical imperative is synthetic *a priori* ((1996), sections 4.388, 4.420).
[31] Jackson and Pettit (1995).

there is no definite cut-off point where all men over X cm in height are tall and all men under it are not. Similarly, there may be moral propositions with indeterminate truth values because of moral vagueness. It is not that we cannot know whether a particular action is morally permissible or not, but rather that it is indeterminate whether it is. Furthermore, there may also be incommensurable values, which cannot be compared against each other. If values 1 and 2 cannot be compared, the propositions

a)   Value 1 is morally better than value 2

b)   Value 2 is morally better than value 1

c)   Value 1 and value 2 are equal in moral worth

are all false, even though they appear at first to exhaust all the possibilities when one attempts to compare the values. Not every philosopher accepts that incommensurable moral values exist, but for simplicity I will not take a stand on the issue.[32]

The determinacy of a proposition is not the same as the determinacy of a proposition's truth value. It may be quite possible for there to be determinate propositions with indeterminate truth values. In order to avoid confusion, from now on I will use the term 'determinate' to refer to the determinacy of propositions unless otherwise stated.

The determinacy of a proposition is based on the content of the terms within it. For example:

1)   Humans should never drink-drive

is a fairly determinate moral proposition. It is clear to whom the proposition applies (human beings), and 'drink-drive' is a term with fairly precise application conditions; that is, it is fairly clear in which circumstances the term can be applied. Agents using it know what drink-driving is, the conditions under which someone can be described as drink-driving, and so on.

By contrast, the proposition

2)   Humans should generally be nice to other humans

---

[32] There has been much debate on moral incommensurability; for example, Nagel (1979), chapter 5; Williams (1979); Richardson (1994), chapter 5; Dworkin (2004), chapter 4.

is less determinate than (1), because the term 'nice' has far less precise application conditions than 'drink-drive'. Many agents would flesh the term out in different ways. Does being nice entail indulging an agent's every desire? Does it mean never upsetting them? Simply telling someone that they should be nice may leave it completely unclear what they should do, but telling them not to drink-drive is a direct appeal to stay away from alcohol if they are going to drive. It is not just that agents disagree over what the precise application conditions of 'nice' are, but that the application conditions themselves are imprecise.

The more precise that the application conditions of the terms in a proposition are, the more informative the proposition becomes, and hence the more determinate it becomes. (2) is not very determinate, as it is not very informative. It tells us very little about what niceness consists of, what humans have to do in order to be nice, and so on. The use of the term 'generally' also has an effect, as (2) does not inform us exactly when the exceptions occur when it is permissible not to be nice. As for (1), 'drink-drive' has more precise application conditions than 'nice', and so it is clearer what humans have to do according to (1). Furthermore, the term 'never' allows for no exceptions, so we know that humans should always refrain from drink-driving. Consequently, (1) is more determinate than (2) because it is more informative.

I will use the phrase 'a determinate proposition' as shorthand for 'a proposition that is more determinate than most other propositions' and the phrase 'a non-determinate proposition' as shorthand for 'a proposition that is less determinate than most other propositions'. 'Determinate' is a comparative term, like 'tall', and so there is no firm division between determinate and non-determinate propositions. The determinacy of a proposition depends on the precision of the application conditions of the terms in the proposition, and there is no clear tipping point where the precision of the application conditions becomes great enough for a proposition to turn from being non-determinate to being determinate. We should expect a grey area of propositions that are not particularly determinate or non-determinate. For convenience, I will ignore this grey area as I wish to concentrate on propositions that do not fall into it.

Calling a proposition 'determinate' does not mean that other propositions cannot be more determinate. Compare (1) to

3) Humans should never operate any car for 48 hours after consuming any alcohol

In some cases, it may be unclear when drink-driving occurs. If a driver has a sip of wine thirty minutes before driving, is that really drink-driving? (3) eliminates more borderline cases such as these than (1) does because it is more informative and tells us more about how humans should behave.

It may be possible for determinate propositions to have indeterminate truth values. Here is an example: assume that a thief is going to jail. Given the facts of the case, it is morally permissible to jail him for six months, but not for sixty years. At some point, the punishment will be long enough to become morally impermissible. But when will this happen? It may be that there is no exact time at which point the punishment is so long that it becomes morally impermissible. Assume that there is an exact time, T. This implies that jailing the thief for T-minus-two-seconds is morally permissible, and jailing him for T-plus-two-seconds is not. But on what basis could a difference of only four seconds make such a large moral difference? We may answer the question by saying that there is no basis, so there cannot be a time T. There is a period of time where it is indeterminate whether it is morally permissible or impermissible.[33]

Suppose that a punishment lasting exactly five years sits inside this grey area. The proposition 'Jailing the thief for exactly five years is morally permissible' thus has an indeterminate truth value. However, it seems to be a determinate moral proposition. It applies to a certain thief, and we know exactly what the punishment is. It is therefore a determinate moral proposition with an indeterminate truth value.

Similarly, non-determinate propositions may have determinate truth values. 'Nice people are nice' clearly has a determinate truth value, but the proposition is not determinate because it gives very little information. It does not tell us much about how to identify nice people, what makes them nice, what specific actions nice people perform, and so on.

The determinacy of a proposition depends on the terms within the proposition. Can terms have application conditions independently of their occurrence in a proposition? Do they only have application conditions when they are included in a proposition? Our answers to these questions depend on our philosophy of language, and I will remain neutral on this issue here. All

---

[33] This example is based on Matthew Kramer's ((2009), pp. 109-113).

we need is that terms have application conditions that affect the determinacy of the proposition in which they appear.

A proposition may be non-determinate even if some of the terms in it have very precise application conditions, because the proposition also includes other terms with less precise application conditions. Take

4) Humans should generally keep promises to other humans

We might think that 'promise' is a term with fairly precise application conditions. Most of the time, we can identify when a promise has been made, and we know the conditions under which a promise is kept or broken. However, (4) is non-determinate, because it includes the term 'generally'. What does 'generally' mean here? Should we keep promises unless it is inconvenient? Unless keeping promises conflicts with other moral norms? Does it mean that in most possible situations humans should keep promises, or in most actual situations? (4) gives no information about when the exceptions occur, and there may be lot of exceptions. So (4) is non-determinate, despite the fact that the term 'promise' has fairly precise application conditions. We can swap the term 'generally' with 'always' to get

5) Humans should always keep promises to other humans

This proposition is more determinate, as it allows no possible exceptions to humans keeping promises to other humans.

A proposition's determinacy is not necessarily related to its status as an *a priori* or an *a posteriori* proposition. The *a priori* proposition 'An agent is nice if they are nice' and the *a posteriori* proposition 'Animals consume things' are both non-determinate. Similarly, the *a priori* proposition 'Bachelors are unmarried' and the *a posteriori* proposition 'In 2000, the Queen of England was the Queen of Canada' are both more determinate.

Core moral propositions can be determinate or non-determinate. Theories that endorse *a priori* core moral propositions are more attractive if at least some of the propositions are determinate, because the normative content of non-core moral propositions are justified by core

moral propositions. If the core moral propositions are non-determinate, it may be unclear what exactly they can justify. Suppose that there is only one core moral proposition: 'Being nice to others is morally obligatory'. 'Nice' has quite imprecise application conditions, so it is unclear which other moral propositions it can justify that can help to guide our actual moral behaviour. We can give the term 'nice' more precise application conditions, but that will make the proposition more determinate. In fact, a theory that only contains non-determinate core moral propositions will be much less capable of justifying the normative content of various non-core moral propositions. I will examine this point in greater detail in section 3 of chapter 2, as it is essential for the core-aposteriorist's defence of the Central Claim.

One can interpret Kant's, Ross's and Parfit's moral theories as endorsing the existence of *a priori* determinate core moral propositions. We already know that they endorse *a priori* core moral propositions, so it is not a great leap to say that some of these propositions are determinate. Take Kant and drink-driving. Could it be morally permissible to get drunk and drive a car? No, because no rational agent could will a maxim that allowed such a theory to be a universal law. The Kantian's endorsement of the categorical imperative can be expressed with the determinate core moral proposition

6) Humans should always act in accordance with the categorical imperative

The Kantian can claim that we can discover *a priori* both what the categorical imperative is and that humans should always act in accordance with it, so (6) is an *a priori* determinate core moral proposition. It can be used in the following argument:

6) Humans should always act in accordance with the categorical imperative
7) Drink-driving is an act that always contravenes the categorical imperative
   Therefore
8) Humans should never drink-drive

(7) is a descriptive proposition, and (6), an *a priori* core moral proposition, justifies the normative content of (8), a non-core moral proposition.

One might object that (6) cannot be determinate because it is too abstract a proposition. Propositions can be more or less abstract than other propositions, and the determinacy of a proposition is necessarily linked to how abstract it is. I use Matthew Kramer's criterion for abstractness here. Moral principle A is more abstract (and less concrete) than moral principle B if and only if the falsity of A would affect the truth value of a wider range of moral principles than the falsity of B would.[34] The objection is that the more abstract a proposition is, the less determinate it must be. The categorical imperative is very abstract – it says nothing explicitly about particular moral duties we have (to develop the skills we have, to support those in need, to keep promises even when it does not benefit us, etc.) or particular situations that actual agents may find themselves in. For this reason, the objection claims, (6) cannot be a determinate proposition. Conversely, the proposition 'Mary should give £2 to the person collecting for Cancer Research UK that she just passed on Tottenham Court Road' is a concrete moral proposition, and it is very determinate.

This objection fails, though, because the distinction between determinate and non-determinate propositions has no necessary link with the distinction between abstract and concrete propositions. First, abstract propositions can be just as determinate as concrete propositions. Take:

White: It is always morally permissible to tell white lies

Social: It is morally permissible to tell white lies to Mr Jones, the complete stranger whom you have just met at tonight's dinner party (and expect never to meet again) to save some slight personal embarrassment

White implies *Social*, but *Social* does not imply *White*. In fact, *Social* appears to apply to only one situation: *this* particular dinner party, *this* particular stranger, *this* particular motive for lying .The truth value of *White* affects the truth value of a wider range of moral principles than the truth value of *Social*, and so *White* is more abstract (and less concrete) than *Social*. However, if we are able to determine just what a white lie is, and under which conditions one may be correctly said to be telling a white lie, then *White* may be just as determinate as *Social*. This

---

[34] Kramer (2009), p. 10. Kramer characterises his metaethics partly with the use of the distinction between abstract and concrete moral principles, and it is important to remember this point in chapter 4, where I examine his moral realism.

does not mean that it is as concrete as *Social*, because it is still the case that the truth value of *White* affects the truth value of a wider range of moral principles than the truth value of *Social* does. Therefore, if a proposition is abstract, that does not imply that it is non-determinate.

Furthermore, we cannot assume that a concrete moral proposition is determinate. Take

Friend: It is morally obligatory for Jack to be particularly nice to Jill today because she is under a lot of stress from her upcoming exams right now

This is a concrete moral proposition, as it applies to a particular agent, Jack, and a particular situation. But it is a non-determinate proposition, since we may have no idea what exactly Jack must do to be particularly nice to Jill.

Returning to (6), we now see that (6) can be determinate even though it is more abstract than many other moral propositions. It may not tell you whether many particular moral propositions are true or false, but this is because (6) is abstract, not because it is non-determinate.

These points also apply to another distinction, that between general laws and particular cases. *White*, if it is true, is a general law regarding the moral permissibility of telling white lies. *Social* and *Friend* are propositions about particular cases. However, the fact that *White* is a general law does not imply that it is non-determinate, and the fact that *Friend* concerns a particular case does not imply that it is determinate.

I am aware that I have used a number of terms in the Central Claim that do not appear in the literature, and one might think that it would be better instead to defend a well-known claim from the literature that is similar to the Central Claim. As we see in this chapter, there are clearly existing disagreements in the literature between ethicists who believe that various normative moral claims can only be defended with *a posteriori* evidence, and ethicists who reject this idea. However, I have concentrated on the Central Claim for two reasons. The first is that whilst there are clearly disagreements between the two groups of ethicists, we need to articulate precisely what the disagreements are about. The Central Claim presents a way of formulating one such disagreement. I do not wish to defend any specific claims made by any particular core-aposteriorist theory, as such claims may not be made by other core-aposteriorist theories. All

core-aposteriorist theories imply the Central Claim, though, and my defence of the Central Claim does not rely on any particular core-aposteriorist theory being correct. The second reason is that the Central Claim is actually a rather narrow claim about the place of the *a posteriori* in the attainment of moral knowledge. It does not imply that (for example) every true moral proposition can be justified *a posteriori*, and it does not deny that *a priori* evidence is often important or essential for ascertaining the truth of the normative content of various particular moral propositions. What I want to show in this thesis is that *a posteriori* evidence is necessary for ascertaining the truth of the normative content of at least some moral propositions. Defending the Central Claim is a good way to do this.

4 – Knowledge and justification

The Central Claim is a specific claim concerning moral epistemology, and therefore I will not be concerned with other metaethical issues, such as the existence of moral properties. However, I will say a little about knowledge, justification and the relationship between them.

Gettier cases show that it is not sufficient simply to have a justified true belief that p in order to know that p.[35] Suppose that I know that Jones has ten coins in his pocket, and that I have good justification to believe that he got a coveted job yesterday. I form the justified belief that a man with ten coins in his pocket got the job. Actually, Smith got the job. Unbeknownst to me, Smith also has ten coins in his pocket. My belief that a man with ten coins in his pocket got the job is justified and true, but I did not *know* that a man with ten coins in his pocket got the job. The truth of my belief depends largely on luck. A lucky belief cannot be knowledge.

Perhaps knowledge is justified true belief plus something else that excludes luck. Whilst this strategy tries to eliminate Gettier cases, it still defines knowledge in terms of justification and belief. Consequently, knowledge is not a basic term. However, this idea has been increasingly challenged. For example, Linda Zagzebski argues that no matter what you add to justified true belief to get knowledge, there is always a way to adapt Gettier cases to ensure that it is not knowledge.[36] Timothy Williamson has famously defended the claim that knowledge is a basic term. He believes that knowing that p is a mental state of an agent, just like believing that p is.

---

[35] Gettier (1963).
[36] Zagzebski (1994).

Furthermore, '[o]ne's evidence [for p] justifies [a] belief… if and only if one's knowledge justifies that belief'.[37] Justification is not conceptually prior to knowledge. Knowledge comes first. It justifies our beliefs, instead of being what is justified. It is thus a mistake to define knowledge in terms of justification.

It would take too much space to adjudicate this issue here. For my purposes, all I need is a necessary link between knowledge and justification. I will point out, though, that we can accept such a connection whether Williamson is right or not. If knowledge can be defined in terms of justification, then one cannot have knowledge without justification. If Williamson is right, then if something cannot justify any belief, it cannot be knowledge. At minimum, knowledge that p can justify the belief that p. One's knowledge that p implies that one's belief that p is justified.

The exact meaning of the term 'justify' can change from author to author. At one end of the spectrum, it can mean something like 'guarantee as true'. For example, X only justifies Y if X guarantees the truth of Y. If Y is not true, then X cannot be either. Towards the other end of the spectrum, 'justify' can mean something like 'increases the likelihood of'. For example, my belief that Helen is now in France may be justified by the fact that I saw her booking a stay in a Parisian hotel for today, as this fact makes it more probable that she is in France. However, it does not guarantee that she is in France (She might have changed her mind about going.). We might even take 'justify' to be weaker than this; it might mean 'make rationally permissible to believe that' or 'render intelligible a belief that'. I am not going to fully define the term 'justify' in this thesis, but my position is closer to the former end of the spectrum than the latter. If X justifies Y, then Y is much more likely to be true than not, even if Y's truth is not guaranteed.

It may still be the case that having a justified true belief is not the same as having knowledge, but as I have said, all we need is a necessary connection between justification and knowledge. One cannot have knowledge of X without having access to a justification of X, either because justification is a necessary constituent of knowledge, or (if we agree with Williamson's views) because knowledge of X always entails that X is justified.

5 – Why accept core-apriorist theories?

---

[37] Williamson (2002), p. 9.

One reason why core-apriorist theories have been so popular is because of the difficulty of showing that one can ascertain the truth of moral propositions simply through *a posteriori* methods. W.D. Ross claims that if you only provide *a posteriori* explanations such as sociological explanations for why we have moral belief X, this neither justifies nor refutes the belief. The truth or falsity of X is independent of the sociological explanation for us having the belief, and Ross believes that we use other methods to ascertain the belief's truth.[38] These methods, claims the core-apriorist, are *a priori* methods.

One might think that Ross rejects sociological explanations simply because of the difficulty (or impossibility) of using descriptive information to justify normative conclusions. If sociological explanations are purely descriptive, they have no normative implications. However, Hallvard Lillehammer points to another explanation of Ross's view, which is that sociology, like other sciences, analyses natural systems. Moral beliefs are not like natural systems, as moral beliefs have characteristics 'of being true or false, of resting on knowledge or of being the product of wishes, hopes and fears'.[39] Natural systems do not, yet it is these characteristics that give moral beliefs any normative authority they may have. Sociology and other sciences therefore do not provide the necessary tools to discover if moral beliefs are true or false. This objection is not the same as the objection that purely descriptive information cannot have normative implications, so it is a separate reason why the core-apriorist may reject the Central Claim.

Both objections, however, imply that sociological study simply does not have the tools to justify any normative moral conclusions. When sociological evidence is used in moral arguments, it does no normative work. It therefore does not justify the normative content of a moral argument's conclusion. The objections can be extended to apply to other types of *a posteriori* evidence. We can see an example of this with Selim Berker's objection to Joshua Greene's argument that consequentialist theories are preferable to deontological theories. Berker claims that the empirical psychological research that Greene cites to support his argument is irrelevant to the argument's conclusion. The core-apriorist will add to this that all the normative work in the argument must be done *a priori*, contrary to what Greene believes.

Greene argues that deontologists normally make moral judgements based on quick, automatic emotional responses, and if reasoning takes place it is to rationalise the judgements

---

[38] Ross (2002), chapter 1.
[39] Ross (2002), p. 13; Lillehammer (2016), pp. 105-106.

after they have been made. He claims that 'there is a remarkable correspondence between what rationalist deontological theories tell us to do and what our emotions tell us to do', and he bases his conclusions on a large body of neurophysiological experiments.[40] Since our deontological judgements are emotionally driven, rather than rationally driven, we have no reason to trust such judgements. Emotions are not a reliable guide to moral truth.[41] In general, then, we should be dubious about deontological theories. Consequentialist judgements are less likely to have this problem, and hence they are preferable.

This argument is problematic. Suppose that Professor X, a deontologist, holds that one should not torture children, even when the consequences of such torture would be good. Greene may claim, using *a posteriori* research, that Professor X's judgement was emotionally generated, and so she has no reason to hold this view. The smaller problem with this claim is that we cannot just assume that Professor X does not make her moral judgements rationally, especially if she is an ethicist herself. It is still an open question how far philosophical training can help us avoid mistakes when making moral judgements.[42] Deontological judgements may yet have a rational basis.

The larger problem is that it does not matter what Professor X believes. What matters is whether the argument that she provides for her belief is correct. Suppose that Joe beats me in a competition. Upset, I accuse him of cheating. I honestly believe that he has cheated, because I am unaware of how upset I am. I unwittingly use my reasoning to construct a plausible argument for how he could have cheated. It then turns out that he did cheat in the way I claimed. I did not know that he cheated, because my belief was not based on the argument I gave. However, the argument provides a method to know whether Joe cheated. It is irrelevant that I did not actually use this argument to form my belief.

At most, if an agent makes a moral judgement based on quick, automatic emotional responses, that could give us a reason to pause before accepting it. But that does not necessarily mean that we should reject the judgement. Obviously, the deontologist must provide convincing arguments for her judgements, but this shifts the debate from 'Why do you believe

---

[40] Greene (2007), p. 68. See also Greene (2014).
[41] I assume here that agents must use rational, not emotional methods to attain moral knowledge. Philosophers such as Jesse Prinz (2007) will object, but that debate is not relevant to the points I make here.
[42] Jonathan Haidt points to evidence that philosophers are more likely to use reason than others ((2001), p. 819), but Greene disputes this ((2014), pp. 719-720).

this?' to 'If this is true, what methods could be used to know it?' The deontologist's psychological state drops out of the picture.

Greene's argument tries to show that deontological judgements are decided by emotional responses, which are normatively irrelevant.[43] Selim Berker suggests that Greene's argument is something like the following:

1) Deontological judgements are influenced by emotional responses that are triggered by certain factors.

2) These factors are normatively irrelevant.

3) So the emotional responses that influence deontological judgements are triggered by factors that are normatively irrelevant.

4) So deontological judgements have no genuinely normative force.[44]

(2) is problematic for Greene, claims Berker, because the experimental results that Greene uses to make his argument do not support it.[45] Berker points out that the normative content of the argument – the part that is doing the philosophical heavy lifting – is not obtained from the *a posteriori* research that Greene cites. He cannot use his experimental results to ascertain which factors in moral judgement-making are normatively relevant, which means that he cannot use the results to conclude that deontological judgements rely on morally irrelevant factors. Consequently, we do not need to examine *a posteriori* how people actually make deontological judgements in order to reach the conclusion that such judgements are unjustified. It is unimportant precisely how people make these judgements. All Greene needs is the argument

5) Deontological judgements are based on certain factors.

6) These factors are normatively irrelevant.

7) So deontological judgements have no genuinely normative force.

We do not need any information about the emotional dispositions or neurophysiology of actual agents to reach the conclusion that consequentialism is preferable to deontological theories.

---

[43] Greene (2007), pp. 67-70.
[44] Adapted slightly from Berker (2009), p. 322.
[45] Berker (2009), p 322.

Berker does not say that we cannot use *a posteriori* research to find out which factors in moral judgement-making are morally irrelevant. He limits his criticisms to the neurophysiological research that Greene uses. However, his criticism is consistent with a stronger type of argument, one that has a core-apriorist conclusion:

8) Determinate core moral propositions have normative content.

9) Moral agents can know at least some true determinate core moral propositions.

10) If an agent uses the wrong method to ascertain the truth of the normative content of a true moral proposition, he cannot know the proposition. To know the proposition, he must use an appropriate method.

11) Any method that is based on *a posteriori* research will fail, for the same reason that Greene's attempt fails. The research will have no relevant normative implications, and if we try using the research to reach relevant normative conclusions, we will find that any such conclusions we draw are not based at all on the research. We reached the conclusions without it.

12) But moral agents can know some true determinate core moral propositions, so there must be a method to ascertain the truth of a moral proposition's normative content.

13) Since this method cannot utilise information obtained *a posteriori*, it must be *a priori*.

I am not going to develop this argument here, though it is worth saying that the core-aposteriorist will disagree with (11), I use the argument simply to show one way that one can support core-apriorist theories. Neither do I want to argue that Greene's argument is ultimately correct. My defence of the Central Claim is entirely independent of Greene's argument.

Some core-apriorists have used something like the argument (8)-(13) to defend their position. For example, Patrick Clipsham uses what he calls the 'What's at issue?' argument ('WA') to argue that there is something very different epistemologically between moral disagreements and non-moral naturalist disagreements. It is not entirely clear what he takes the WA to show, but it can be interpreted as a version of the argument (8)-(13).[46]

---

[46] Clipsham bases the WA on his interpretation of Allan Gibbard (2003), pp. 23-29. Gibbard might not agree with him, because Gibbard argues in that passage that the meaning of moral terms is empirically indeterminate, which is not what the WA claims. Neither does Gibbard use the WA against Shaun Nichols when Nichols criticises Gibbard's metaethics on empirical grounds (Nichols (2004), pp. 90-96; Gibbard (2006)).

Clipsham claims that we can resolve non-moral naturalist disagreements *a posteriori* in three ways.[47] If Jack and Jill disagree over whether something is a horse, they can do three things to resolve the disagreement:

14) agree on a complete definition of 'horse' and check *a posteriori* whether the thing matches the definition;

15) agree on what the *a posteriori* evidence is, so that they can match it to the criteria for something being a horse, having also agreed on how those criteria should be applied; or

16) agree on belief-formation norms regarding horses; that is, agree under what conditions we should believe that something is a horse.

If the disagreement cannot be resolved in one of these ways, it becomes 'baffling' what the disagreement is really about.[48] However, moral disagreement is different. Suppose that Jack and Jill disagree whether it is morally right to torture a terrorist for information that will save lives. It seems that they can

17) agree on a definition of 'right';

18) agree on all the relevant non-moral facts about the situation; and

19) agree on when agents should form beliefs about moral matters; but

20) disagree on whether torturing the terrorist is morally right.

Clipsham states that the disagreement may be entirely reasonable, and it seems intuitive that it is a real moral disagreement.[49] Deontologists and consequentialists, he claims, can disagree over moral dilemmas even when they agree on all the non-moral facts about the dilemmas, and they are not confused or irrational to do so. Such disagreements must be about moral propositions that cannot be justified by *a posteriori* evidence.

But why is this the case? The core-apriorist will give an answer such as the argument (8)-(13). Moral disagreements are normative disagreements, and so if we tried to resolve them by

---

[47] Clipsham uses the term 'empirical' rather than '*a posteriori*', but this makes no practical difference.
[48] Clipsham (2014), pp. 216-217.
[49] Clipsham (2014), pp. 221, 225-226, 228.

using *a posteriori* evidence to argue for or against a particular moral proposition, we would fail. We can only ascertain the truth of the normative content of the moral proposition by using an *a priori* method.

6 – The Core-Apriorist Thesis

In order to argue for the Central Claim, I will argue that the core-apriorist is wrong, which entails that the core-aposteriorist is right. To do so, I need to refute the positive thesis that the core-apriorist holds. This is the Core-Apriorist Thesis ('CAT'):

CAT: In the group of determinate core moral propositions that a possible moral agent can know, for any proposition in that group, either:

a)   the normative content of the proposition can only be ascertained by using exclusively *a priori* methods, or

b)   if the normative content of the proposition can be ascertained by using *a posteriori* or mixed methods, it can also be ascertained by using an *a priori* method, and the *a priori* method has explanatory precedence over the *a posteriori* and mixed methods.

Philosophers such as Parfit would hold part (a), and those such as Ross would hold part (b). Core-aposteriorists must reject it entirely, as it implies that the Central Claim is false. If part (a) is correct, then we cannot know the truth of any determinate core moral propositions with *a posteriori* or mixed methods. If part (b) is correct, then if the truth of a particular determinate core moral proposition can be known with *a posteriori* or mixed methods, it can always be known with *a priori* methods as well. The *a priori* methods have explanatory precedence over the other methods. The core-aposteriorist will therefore accept neither part (a) nor part (b).

It is possible to reject both the Central Claim and the CAT. This third position holds that if there is an *a posteriori* or mixed method to ascertain the truth of the normative content of a determinate core moral proposition, there is also an *a priori* method to do so, but the *a priori*

method has no explanatory precedence over the other method. I am going to ignore this position because when I criticise the CAT later in the thesis, I do so by showing that there are most likely some determinate core moral propositions that can only be known through the use of *a posteriori* information. This entails that the third position is automatically unsuccessful. Since I want to exclude this third position from the debate between the core-apriorist and the core-aposteriorist, I will make a little more precise what a core-apriorist holds. Core-apriorists accept the CAT, and hence reject the Central Claim. My strategy will be to show that we can reasonably reject the CAT, and this is good evidence that the Central Claim is correct. I will do this in two stages, by rejecting part (b) of the CAT in chapter 2, and arguing against part (a) in the rest of the thesis.

Part (b) introduces the idea of explanatory precedence. We can show what explanatory precedence is by using what I call the mathematics analogy.[50] An agent can work out that 2+2=4 *a priori* from particular mathematical axioms, or *a posteriori* by counting apples. He can get the correct answer with each method, and both methods are reliable. However, the *a priori* method takes explanatory precedence over the *a posteriori* method. The reason why 2+2=4 is because the calculation is justified by certain mathematical axioms that are true, rather than because certain apples were counted first or because apples were used instead of oranges. The *a priori* method explicitly uses the axioms to make the calculations, whereas the *a posteriori* method does not.[51] The *a priori* method thus has explanatory precedence, because if the axioms were false, the truth of the calculations may have been different. Changing apples for oranges could never affect the truth of the calculations.

If I count apples to find out that 2+2=4, I still have good reason to believe that 2+2=4. I can explain why I got the answer that I did, by grouping apples together in a particular way. However, mathematicians do not hold that 2+2=4 simply because every time we count physical objects, we find out that two objects plus two objects equal four objects. Mathematicians search for an explanation why 2+2=4 that has more explanatory power, and they find it by discovering certain mathematical axioms that (as far as we know) are true.

---

[50] I use the mathematics analogy as an illustration, and so I am not endorsing any philosophy of mathematics it may imply.
[51] The analogy assumes that mathematical knowledge can be *a priori*. We will see in section 1 of chapter 2 that this assumption can be challenged, but for the moment I will ignore this complication.

The mathematics analogy claims that moral knowledge is similar to mathematical knowledge in this way. Even if there are *a posteriori* methods to attain moral knowledge, these methods are only useful because there are *a priori* methods to attain the same knowledge, and so the *a priori* methods have explanatory precedence over the *a posteriori* methods. So if we use an *a posteriori* method to ascertain a determinate core moral proposition, part (b) of the CAT kicks in. The *a posteriori* method may let us attain determinate core moral knowledge, but only because there is an *a priori* method to attain the same knowledge that has explanatory precedence over the *a posteriori* method.

Core-apriorists are less likely to endorse part (b) of the CAT than part (a), but it should be included in the CAT for two reasons. First, one can find something like part (b) in the writings of some core-apriorists. For example, W.D. Ross appears to endorse it.[52] Second, part (b) does not actually concede that much to core-aposteriorists. As we will see shortly, when core-aposteriorists endorse *a posteriori* or mixed methods to attain moral knowledge, they often imply that the methods utilise fundamental reasons or explanations for the truth of various moral propositions. To use a very simple example, a core-aposteriorist may hold that the fundamental reason why moral proposition X is true is because natural property Y exists, and we can only ascertain via a particular *a posteriori* or mixed method that Y exists and implies the truth of X. Therefore if a method has explanatory precedence over other methods, it will be an *a posteriori* or mixed method, contrary to part (b) of the CAT.

It is helpful to see how the CAT applies to a particular example. Haidt et al have conducted experiments to show that people often make moral judgements based on their emotional reactions to moral scenarios. For example, they feel that incest is disgusting, so they believe that it is morally wrong under any circumstances, and they tend to persist in their belief even when all their stated reasons for their belief have been debunked.[53] Jesse Prinz uses these experiments as part of his argument that we attain moral knowledge by experiencing certain emotional dispositions.[54] Is incest wrong? If you are disposed to have particular emotional reactions towards it, then yes.

---

[52] Ross (2002), pp. 32-33. See also Ross (1929), p. 121, where he claims that 'that something can be inferred does not prove that it cannot be seen intuitively'.
[53] Haidt et al (2000).
[54] Prinz (2007), pp. 29-32.

If Prinz is right and the prohibition against incest is only knowable *a posteriori*, then it cannot be justified by an *a priori* determinate core moral proposition. Therefore, if Prinz is right, then part (a) of the CAT is wrong. Obviously, core-apriorists who endorse part (a) of the CAT will turn this around: since part (a) is correct, Prinz must be wrong. Haidt et al's experiments can only describe actual people's behaviour when they make moral judgements. The experiments do not reveal which moral propositions are true or how they can be known.

If part (b) applies, then Prinz's *a posteriori* method to attain moral knowledge may work, but there will also be an *a priori* method to attain the same knowledge, and this *a priori* method will have explanatory precedence over Prinz's *a posteriori* method. However, given the details of Prinz's metaethics, this is impossible. Prinz believes that the method works because moral properties simply are properties that dispose observers to experience certain emotions.[55] The fundamental reason why genocide is wrong is because it consists of mass murder, which provokes outrage in observers. This therefore implies that no other method can have explanatory precedence over Prinz's method, because there is no more fundamental reason. We thus see here a core-aposteriorist theory that would reject part (b) of the CAT because the theory entails that *a posteriori* methods have explanatory precedence.

Richard Boyd is another core-aposteriorist. His moral realism is similar to scientific realism, which holds that scientific theories seek to describe real phenomena and that 'ordinary scientific methods constitute a reliable procedure for obtaining and improving (approximate) knowledge of the real phenomena that scientific theories describe'.[56] We assess our scientific beliefs by observing the world around us, and we should do the same with our moral beliefs.

> '[G]oodness is a property quite similar to other properties studied by psychologists, historians and other social scientists, and observations will play the same role in moral enquiry that they play in other kinds of enquiry about people.'[57]

The properties that are discovered by psychologists, historians and social scientists are commonly discovered *a posteriori*. One cannot discover the results of the 1829 British general election without *a posteriori* investigation of contemporary evidence. But if moral properties are

---

[55] Prinz (2007), pp. 20-21, 92.
[56] Boyd (1988), p. 181.
[57] Boyd (1988), p. 206.

so similar to historical and psychological properties, it is very unclear how there could be any purely *a priori* method to attain knowledge of moral properties. If we cannot do so with historical properties, how could we do so with moral properties? Boyd does not even hint that an *a priori* method could be possible, and we can guess why. As moral properties are natural properties, we should be able to know them in the same way that we know other natural properties. If we cannot know natural properties *a priori*, then we cannot know moral properties *a priori*.[58] Since no *a priori* methods to attain moral knowledge are possible, this entails rejecting both parts (a) and (b) of the CAT.

The lesson to draw from Prinz and Boyd is that many philosophers who want to explain how moral agents attain core moral knowledge also have other metaethical aims, such as the metaphysical task of explaining what moral properties are, and this shapes their epistemological commitments. This is why core-aposteriorists do not want to endorse part (b) of the CAT. Their epistemological commitments entail that if any method has explanatory precedence over other methods, this method is an *a posteriori* or mixed method.

Having laid out the debate, we can now begin defending the Central Claim.

---

[58] I ignore here the objection that some natural properties are knowable *a priori*; for example, Kant thought that some knowledge of causation and time is *a priori* ((1998), A64/B89-A83/B109). It is clear that Boyd is thinking of natural properties that can only be known *a posteriori*.

Chapter 2 – The *a posteriori* strategy

In this chapter, I will first reject part (b) of the CAT. I will then examine two strategies to reject part (a). The first strategy, which involves examining the cognitive abilities of moral agents, initially looks promising, but it is a dead end. The second strategy, which seeks to refute the CAT, is more successful, and will take up the rest of the thesis.

1 – The mathematics analogy

Part (b) of the CAT can be defended with the mathematics analogy. If *a priori* mathematical knowledge exists, and if moral knowledge is sufficiently like mathematical knowledge, then that is good evidence that *a priori* moral knowledge exists.

The core-aposteriorist could try arguing against the possibility of *a priori* mathematical knowledge. She could claim that mathematical axioms cannot be known *a priori*, and so the propositions that they justify cannot be known *a priori*. Therefore there are no *a priori* methods to attain mathematical knowledge, and the analogy collapses. This is an obvious strategy to try. If mathematical knowledge is indeed similar to moral knowledge, then if there is no *a priori* mathematical knowledge, we have no reason to think that there is any *a priori* moral knowledge. However, this is actually a difficult route to take, as I will now show. We will see that the core-aposteriorist would do better to argue differently, and to claim that the mathematics analogy is actually redundant when it comes to defending core-apriorist theories.

How do we choose which mathematical axioms to accept? Justin Clarke-Doane points out that your philosophy of mathematics is very likely to influence the axioms you choose.[1] For example, George Boolos cheerfully declares himself untroubled by the existence of infinitely many natural numbers, but he balks at some of the theorems that are provable in ZFC, a type of set theory. ZFC's axioms prove the existence of orders of infinity so high that Boolos becomes sceptical. He admits that he does not have 'very good evidence' for rejecting such orders, and he does not categorically deny their existence. But the burden of proof, he claims, is on the set theorist to justify the axioms. ZFC's axioms are 'so removed from experience and the

---

[1] Clarke-Doane (2014), p. 241.

requirements of the rest of science (including mathematics)' that we need better reasons to accept the axioms that generate those theorems.[2]

Clarke-Doane writes that when we accept mathematical axioms, what we really do is begin with particular propositions that we deem plausible and seek general principles which systematise those propositions.[3] Boolos appears to be doing just that. Unable to accept unimaginably large numbers, he rejects the axioms that produce them. Certainly the axioms mathematically imply the existence of such numbers, but why should we accept those axioms in the first place?

Interestingly, Boolos appeals to science, not just mathematics, for evidence of the existence of numbers. This should trouble the core-apriorist, as much of science is *a posteriori*. If *a posteriori* evidence helps to show which mathematical axioms to accept, then the axioms are not accepted on *a priori* grounds. Mathematical knowledge is not *a priori*, so the core-apriorist cannot use the mathematics analogy to defend part (b) of the CAT.

However, we cannot just assume that mathematical knowledge is ultimately *a posteriori*. We have to show that it is, and there are philosophical theories that claim to do so. I will outline two well-known types of such theories here. Unfortunately, neither is particularly promising. If we want to show that mathematical knowledge is *a posteriori*, and so the mathematics analogy is harmless to the core-aposteriorist, there is a lot of work to do.

First, mathematical nominalism claims that there are no abstract entities, which numbers are typically taken to be. There is a large variety of nominalist theories, which hold that concrete entities in the universe play the roles that numbers are supposed to play.[4] Consequently, the nominalist can say that a mathematical proposition is true, but this depends on what the concrete universe is like. The structure of the concrete universe determines what is available to play the roles of numbers in a mathematical theory, and knowledge of the concrete universe may be *a posteriori*.

A second option is to become a fictionalist.[5] Fictionalism claims that numbers are fictional just like literary characters are, and so no mathematical propositions are true. Therefore we cannot really have any knowledge of true mathematical propositions, either *a priori* or *a*

---

[2] Boolos (1999), p. 121.
[3] Clarke-Doane (2014), p. 240.
[4] Burgess and Rosen (1987) provide an overview of the different types of nominalism.
[5] E.g. Field (1989), chapter 1.

posteriori, though we may use mathematical discourse as a way – not the *only* way – to describe physical phenomena, such as the Earth's orbit around the sun.

If either of these theories works, then the last thing that the core-apriorist wants is to show that moral knowledge works in the same way as mathematical knowledge. However, the theories come with various problems, so we cannot just pick one of them and conclude that the mathematics analogy fails. Fictionalism is in trouble if the only way that we can describe physical phenomena is mathematically. That would entail that fictional entities are essential for correct descriptions of the phenomena, which is highly implausible. If, however, fictionalism works, then the core-aposteriorist had better hope that the moral error theorist cannot now use the mathematics analogy. The moral error theorist believes that every possible moral proposition is false, and his version of the mathematics analogy would run like this: fictionalism entails that knowledge of true mathematical propositions is impossible. Mathematical epistemology works like moral epistemology, and therefore knowledge of true moral propositions is impossible. The core-aposteriorist does not want this result any more than the core-apriorist does.

If the core-aposteriorist chooses to endorse mathematical nominalism, she avoids the problems with fictionalism, because the analogy becomes harmless. If mathematical knowledge is like moral knowledge, then moral entities will be concrete entities in the universe as well. But it is an open question whether mathematical nominalism is plausible. For example, we may ask if there are enough concrete entities in the universe to play all the roles that mathematics requires. Do we really want to hold mathematical theorems hostage to the amount of concrete entities in the universe? What if there are not enough concrete entities to play the roles of numbers in a mathematical system? Mathematical nominalists are not even agreed on what 'concrete' means.

Other problems will also arise depending on the core-aposteriorist's specific moral theory. If she is a non-reductionist natural moral realist, for example, then she believes that moral properties are natural properties. But she also believes that we cannot reduce moral properties to non-moral natural properties, so it seems unlikely that she can say that particular concrete properties can play the role of moral properties. Doing so would imply some sort of reductionism.

Objecting to the mathematics analogy by showing that mathematical knowledge is *a posteriori* turns out to be a difficult task, and the core-aposteriorist may try instead to find another way forward. Fortunately, there is a much simpler way to neutralise it, by showing that it is of no use to the core-apriorist.

The mathematics analogy assumes that we attain moral knowledge in the same type of way that we attain mathematical knowledge. But why do we not get moral knowledge in the same type of way that we get knowledge of chemistry? In chemistry, we use *a posteriori* experiments to discover chemical laws, which is very different from discoveries in pure mathematics. We cannot work out how chemicals react with each other just by reasoning *a priori*. So why is morality like mathematics and not like chemistry? We can appeal here to the 'natural sciences analogy', which claims that we attain moral knowledge in the same way that we attain knowledge of the natural sciences.[6]

The core-apriorist may object that when we examine the world *a posteriori*, we discover what there is, but not what there necessarily is. I discover *a posteriori* that there are birds in my garden, but this is contingent. There may have been none. The core-apriorist could claim that when it comes to necessary facts, we can only discover what is necessarily the case by using *a priori* methods. This even applies to necessary facts about the natural world. It is a contingent *a posteriori* fact that Bert is a bachelor, but we discover *a priori* that since he is a bachelor, he is necessarily unmarried. So, the core-apriorist continues, scientific necessities must be *a priori*. If the natural sciences analogy holds, then moral necessities must also be *a priori*. The natural sciences analogy therefore does not help the core-aposteriorist.

The core-aposteriorist can reply that if moral necessities exist, then at least some of them are *a posteriori* necessities. Saul Kripke has argued that metaphysical necessities can be *a posteriori*. Suppose that it is true that water is necessarily H_2O. After all, it seems impossible that any substance could be water without also being H_2O. But we did not find this out by using *a priori* methods, and it is difficult to see how we could have done. Instead, we had to use *a posteriori* research, and so we discovered the necessary fact *a posteriori*.[7] Scientific necessities, if they are metaphysical necessities, therefore need not be *a priori*, so they cause

---

[6] Richard Boyd's moral epistemology, for example, is based on the natural sciences analogy.
[7] Kripke (1972), pp. 35-38, 100-105.

no problem for the core-aposteriorist's use of the natural sciences analogy. Why can moral necessities not act like metaphysical necessities?

It is important to realise that I am not arguing here that the natural sciences analogy in particular is correct. For one thing, the natural sciences analogy implies that there is an *a posteriori* method to ascertain the truth of the normative content of determinate core moral propositions, and so it ignores the possibility of mixed methods to ascertain the truth. It also ignores the possibility that there is an *a posteriori* method to do so that is very different from the *a posteriori* methods we use in the natural sciences. The reason I bring up the natural sciences analogy is to illustrate only one way that the core-aposteriorist may choose to reject the mathematics analogy. Both the core-aposteriorist who endorses *a posteriori* methods and the core-aposteriorist who endorses mixed methods are committed to rejecting it. Furthermore, one of the three types of method must be correct. If we cannot know the truth of a particular knowable moral proposition via *a priori* methods, *a posteriori* methods, or mixed methods, it is difficult to see how we can know it at all.

My point, then, is that if there are alternatives to the mathematics analogy that favour the core-aposteriorist, such as the natural sciences analogy, we need to know why the core-apriorist urges us to accept the mathematics analogy. I will show now that he does so because he makes a particular assumption, but this assumption renders the mathematics analogy redundant. If the assumption is false, then the mathematics analogy cannot work. If it is true, it already disproves the Central Claim, so the core-apriorist does not need the mathematics analogy.

First, if the mathematics analogy is wrong, then there must be a method to know at least some knowable moral propositions that is not an *a priori* method, otherwise the moral propositions are not knowable. The only types of method left are *a posteriori* methods and mixed methods, and therefore the analogy poses no threat to the core-aposteriorist.

Assume, then, that the mathematics analogy is correct. The core-apriorist wants to say here that there are always *a priori* methods to ascertain determinate core moral propositions that have explanatory precedence over *a posteriori* and mixed methods. But if this is so, then the mathematics analogy is redundant against the core-aposteriorist. The mathematics analogy relies on the key assumption that *a priori* methods always exist to ascertain determinate core

moral propositions. If this assumption is false, then the mathematics analogy fails. But if the assumption is true, then the assumption automatically disproves the Central Claim without the need for the mathematics analogy or explanatory precedence. It would show that there is no core determinate moral proposition the normative content of which can be ascertained as true only via *a posteriori* or mixed methods. Consequently, either the mathematics analogy is wrong or it is right but redundant, because the real work against the Central Claim is done by the key assumption.[8]

The core-aposteriorist can therefore ignore part (b) of the CAT. The real issue is whether there are any true determinate core moral propositions that can only be known via an *a posteriori* or mixed method, and this debate need not concern explanatory precedence.

2 – Moral cognitive abilities

We come now to the first strategy to defend the core-aposteriorist. There have been moves recently amongst ethicists to use *a posteriori* information about our actual cognitive abilities in support of core-aposteriorist theories. In this section, I will argue that these moves are frequently unsuccessful.

Shaun Nichols, a core-aposteriorist, thinks that in order to make what he calls 'core moral judgements', an agent needs to have a normative theory (which may be nothing more than just a collection of moral norms) and some emotional mechanism. Core moral judgements are moral judgements about the causing of harm.[9] Nichols claims that to make such judgements, we must have the ability to feel particular emotions; we have strong emotional reactions against causing harm, which is a judgement that it is wrong.[10] Furthermore, the reason why one moral norm overrides another may be because of our emotional reactions to the norms. Our emotional reaction not only tells us which norm overrides the other, but it also dictates whether the overriding happens. It is therefore difficult to see how we could know *a priori* when overriding happens, if the overriding depends on our emotional reactions.

---

[8] This is similar to Christopher Cowie's objections to companions in guilt arguments. Such arguments claim that moral facts are similar to other kinds of non-moral facts, and since these non-moral facts uncontroversially exist, there is no special problem that prohibits the existence of moral facts. Moral error theories are therefore unsuccessful. Cowie objects that these arguments either fail or only succeed if there are independent reasons to reject moral error theories, so the arguments are redundant ((2014); (2016)).
[9] Nichols (2004), pp. 6-7, 16-18. Nichols uses the term 'core' differently from me, and there is no necessary relationship between them.
[10] Nichols (2004), p. 135.

Nichols uses *a posteriori* psychological research concerning how people make core moral judgements to argue that if an agent cannot feel these emotions, she cannot make such judgements.[11] I want to argue here that work such as Nichols's gives the impression that we can use such *a posteriori* work directly to disprove the CAT, but this impression is mistaken.

R.J.R Blair has conducted influential experiments to show that psychopaths cannot tell the difference between moral transgressions (such as hitting people) and conventional or social transgressions (such as school pupils not paying attention to the teacher).[12] This is known as the moral/conventional distinction. Blair claims that psychopaths can recognise when a moral transgression is committed, but they treat all transgressions, moral or conventional, in the same way. Non-psychopaths differentiate between the types. Furthermore, psychopaths are much less likely to claim that a moral transgression is wrong because of the harm that the transgression causes.

If psychopaths cannot recognise the moral/conventional distinction, does this mean that they are incapable of recognising moral norms as moral, and that they cannot attain genuine moral knowledge? Since Blair's work, philosophers and psychologists have been arguing about what the distinction really shows.[13] Fortunately, I need not settle the debate here, as I am only using Blair's work as an example. Suppose, then, that psychopaths cannot recognise the moral/conventional distinction, and therefore they cannot recognise true moral propositions. To a certain extent, the psychopath's endorsement of particular moral norms is just him copying what non-psychopaths do. The next question is why psychopaths cannot attain moral knowledge. What cognitive deficit do they have that prevents them from identifying moral transgressions as moral?

Nichols argues that it cannot be a defect in the psychopath's rationality.[14] The problem is not that the psychopath lacks the ability to see the point of view of other people, or that he has a rational deficiency of some other type, or that he suffers from 'intellectual arrogance' and insists that he is right without being able to explain why. If Nichols is right, the psychopath can be just as rational as anyone else. His moral deficiency must therefore have a non-rational cause.

---

[11] Nichols (2002); (2004).
[12] Blair (1995).
[13] E.g. Kelly et al (2007); Prinz (2008); Aharoni, Sinnott-Armstrong and Kiehl (2012). It is also difficult to define exactly what the distinction is (e.g. Southwood (2011)).
[14] Nichols (2002).

If this is true, then rationality by itself is not enough for an agent to be moral. Suppose further that what prevents psychopaths from recognising the moral/conventional distinction is an inability to feel various emotions. Therefore, in order to attain knowledge of a moral proposition, human moral agents must be able to have certain emotional attitudes towards the propositions under certain conditions.

Take the following moral proposition:

Torture: Torturing sentient beings for pleasure is morally impermissible

Assume that *Torture* is a true determinate core moral proposition. If Nichols is correct about the psychopath, how can moral agents know that it is true? It appears that if most reasonably intelligent people with functioning moral cognitive abilities conclude that it is true, that is good evidence that it is true. 'Cognitive abilities' is a very broad term here. They include any abilities that agents can use to attain any sort of knowledge, whether mathematical, moral, scientific, perceptual, and so on. Moral cognitive abilities are the cognitive abilities an agent uses to attain moral knowledge.

Most non-psychopathic people have particular emotional dispositions when considering *Torture*, whilst the psychopath does not. Given what Nichols has told us, we discover this by using *a posteriori* psychological evidence. So now it seems that we can find out whether *Torture* is correct depending on what our emotional dispositions are when we think about it. If *Torture* never triggers any emotional reaction in any agent who considers it, then this is good evidence that it is not true. Our confidence in the truth of *Torture* depends on the cognitive abilities that we actually have, and we find out *a posteriori* about these abilities. The psychopath provides the opportunity to observe an actual agent who cannot attain moral knowledge, and we can examine how he is cognitively different from actual moral agents. *A posteriori* examination of the difference tells us what cognitive abilities are required for agents to obtain evidence of true determinate core moral propositions. Armed with this knowledge of our moral cognitive abilities, we know that if a particular proposition can be ascertained as true via these abilities, we can have confidence that it is true.

The idea that we can use *a posteriori* information about our cognitive abilities to attain moral knowledge can also be found in work by Owen Flanagan, Hagop Sarkissian and David Wong. They endorse 'ethical naturalism'.

'[M]oral philosophy should not employ a distinctive a priori method of yielding substantive, self-evident and foundational truths from pure conceptual analysis. The claims of ethical naturalism cannot be shielded from empirical testing. Indeed, the naturalist is committed to there being no sharp distinction between her investigation and those of relevant other disciplines (particularly between epistemology and psychology)… [E]thical science must be continuous with other sciences.'[15]

Replace 'naturalist' in this quote with 'core-aposteriorist', and this would be a good description of some of the core-aposteriorist's beliefs.

Taking their cue from W.V.O. Quine, Flanagan, Sarkissian and Wong separate what they call 'naturalized epistemology' into two parts: a descriptive-genealogical component and a normative component.[16] As far as moral epistemology goes, the first component describes the characteristics that humans have (such as sympathy, egoism, empathy, etc.), how emotions are involved in our actual moral judgements, how we develop moral systems, what actually happens when people make moral judgements, and so on. This involves the study of psychology, neuroscience, anthropology, and other sciences.[17] The second component involves 'the gathering together of norms of inference, belief, and knowing that lead to success in ordinary reasoning and in science'. We cannot discover these norms simply by analysing actual moral practice, but doing so is an essential element of the discovery. We discover the norms by looking at actual moral practice and then 'abstracting successful epistemic practices from unsuccessful ones… The database is, as it were, provided by observation of humanity'.[18]

A typical objection is that we cannot use the descriptive-genealogical component to tell us anything about the normative component. We cannot use descriptive information about our actual moral practices to tell us about what our moral practices *ought* to be. Flanagan,

---

[15] Flanagan, Sarkissian and Wong (2008), p. 5. Ethical naturalists do not believe that moral norms and values can be explained by invoking supernatural forces (p. 2).
[16] Quine (1969).
[17] Flanagan, Sarkissian and Wong (2008), p. 12.
[18] Flanagan, Sarkissian and Wong (2008), p. 12.

Sarkissian and Wong think that this objection is a little odd, considering typical practice in moral philosophy. They claim that '*everyone* thinks that philosophical psychology… has *implications* for ethics'.[19] Flanagan, Sarkissian and Wong are not claiming that we can prove that certain moral principles are correct merely by looking at descriptive information, but they do think that it is essential to do so.

They make two general claims. First, we can investigate actual moral practices *a posteriori*. This provides information about the cognitive abilities we use to engage in those practices. We must use particular abilities when discovering moral truths. The second claim is that despite the objection that we cannot simply go from the descriptive to the normative, we commonly use *a posteriori* information about our actual moral practices to help identify true moral principles. 'The normative component [of moral epistemology] involves the imaginative deployment of information from *any* source useful to self/social examination, forming new or improved norms or values, improving moral education practices, training moral sensibilities, and so on'.[20]

In other words, we can investigate *a posteriori* the cognitive abilities we use to reach moral truth and attain moral knowledge, and this information tells us something about how we *should* do so. The core-aposteriorist's hope is that some of these truths can be expressed with determinate core moral propositions. We have already seen how this idea could be fleshed out by using Blair's work on psychopaths. If Nichols is right, psychopaths cannot attain moral knowledge because they lack certain emotional abilities. If we investigate further how moral agents use these emotional abilities to attain moral knowledge, we will be able to discover what the content of this knowledge is.

Unfortunately, there is a major problem with this hope, which is that when we recognise moral truths, we do not commonly do so by investigating our cognitive abilities. When we make moral judgements, we think we are doing more than drawing on knowledge of what is happening cognitively. Take the argument

1)   There is a table in front of me.

2)   If I use perceptual ability PA correctly, I will know that there is a table in front of me.

---

[19] Flanagan, Sarkissian and Wong (2008), p. 13, emphasis in the original.
[20] Flanagan, Sarkissian and Wong, (2008), p. 17, emphasis in the original. The emphasis on '*any*' source implies that they are happy to use both *a priori* and *a posteriori* sources, in which case they would endorse mixed methods to ascertain the truth of the normative content of determinate core moral propositions.

3) I use perceptual ability PA correctly.

4) I know that I have perceptual ability PA.

Therefore

5) I know that there is a table in front of me.

Assume that all the premises are true. The question is what premise (4) is doing here. We can derive (5) from premises (1)-(3) alone; it is not necessary for me to know anything about the perceptual abilities that I use. All I need to do is use them correctly. I do not need to know what they are, or that I must use them in a particular way. Similarly, whilst we can investigate our moral cognitive abilities *a posteriori*, we do not need to know about these abilities when recognising moral facts. All we have to do is use the abilities, which does not imply that we need to know facts about them that can be used as premises in moral arguments.

So where did the core-aposteriorist go wrong? The problem is that she has reached too far. It is true that an agent cannot recognise any truths unless they have the cognitive ability to do so. Knowledge of the limits of our cognitive abilities can help outline the limits of what we can recognise as true or false, but these limits are extremely general. Knowing the limits of what we can morally know does not provide us with a list of particular true moral propositions. A psychopath may be unable to recognise any moral propositions as specifically moral because they lack certain emotions, but that does not mean that when we successfully argue for moral propositions, we must use evidence about our cognitive abilities as premises in the argument.

This is similar to mathematics. In order to do mathematics, we need certain cognitive abilities. If we do not have them, then we do not have access to mathematical truths. But we do not reach mathematical truths by finding out what these abilities are. The abilities allow us to attain knowledge of the premises needed to reach mathematical conclusions, but they are not part of the premises themselves. Furthermore, whilst we can investigate *a posteriori* the cognitive abilities we use to attain mathematical knowledge, this does not entail that we attain mathematical knowledge *a posteriori*. Similarly, the fact that we can use certain cognitive abilities to attain moral knowledge does not mean that knowledge of these abilities is used to attain moral knowledge. We only need to use those abilities. Suppose that, as Nichols thinks, we can use particular emotions to attain some moral knowledge. We do not need to know that

we are experiencing these emotions; merely feeling them provides us with moral knowledge. We need not even be able to put the feelings into words. We can also argue analogously with colour perception. If I investigate how humans see colour, I can work out the range of colours that we can see. But this knowledge is not part of the perceptual process that occurs when I look at something blue. I just know, thanks to my perceptual abilities, that I see something blue. I need have no knowledge about how my perceptual abilities work.

We can sum up this problem to make clear how it strikes against the core-aposteriorist. It may be possible to work out *a posteriori* what cognitive abilities we use to make mathematical calculations. However, we do not use this information to make mathematical calculations, and so the fact that we work out what our mathematical cognitive abilities are *a posteriori* does not imply that we work out mathematical calculations *a posteriori*. Similarly, if we can work out what our moral cognitive abilities are *a posteriori*, this does not imply that we recognise moral truths *a posteriori*. The hope that *a posteriori* analysis of our actual cognitive abilities will provide us with moral knowledge ultimately fails.

3 – The *a posteriori* strategy

The core-aposteriorist cannot argue against the CAT by appealing to *a posteriori* information about our cognitive abilities. Neither can she argue against the CAT simply by developing an *a posteriori* or mixed method to ascertain the truth of the normative content of determinate core moral propositions, because there may be an *a priori* method to do the same thing. Instead, the core-aposteriorist must show that core-apriorist theories are not viable. She can do so by using the *a posteriori* strategy, which takes the following form:

a) At least some true determinate core moral propositions are not *a priori* self-evident. These determinate core moral propositions are either *a posteriori* self-evident, or can be known by using other propositions as premises.

b) At least some of these propositions are not knowable solely by using *a priori* propositions as premises. Any attempt will fail either because some of the purportedly *a*

priori propositions contravene the principle *Sufficient Determinacy* or are really *a posteriori* propositions.

c) Since moral agents have determinate core moral knowledge, it is reasonable to assume that these determinate core moral propositions are knowable.

d) As they cannot be knowable via an *a priori* method, they must be knowable via an *a posteriori* or mixed method.

This strategy relies on the following principle, which I will argue for shortly:

Sufficient Determinacy: If a proposition is able on its own to justify the normative content of a determinate moral proposition, the justifying proposition is also determinate.

The *a posteriori* strategy denies that all knowable determinate core moral propositions are knowable via *a priori* methods. The CAT is therefore false. Since the propositions are knowable, though, they must be knowable via *a posteriori* or mixed methods. So the Central Claim is true.

The drawback of the strategy is that it does not provide a single argument that can be applied across the board. Such an argument against the CAT would work against all core-apriorist theories, whatever the particular details of those theories were. The *a posteriori* strategy does not have this benefit. What it does is take a particular core-apriorist theory and analyse it to show where the theory breaks down as a core-apriorist theory. It investigates the details of that particular theory, which means that any core-apriorist theory that has sufficiently different details will remain unaffected until it itself is subjected to examination. It is therefore an inductive strategy, rather than a deductive one.

This is a considerable drawback, as it would be more conclusive to have an argument that slices through all core-apriorist theories in one go. However, the strategy has considerable advantages. First, since it involves refuting core-apriorist theories by examining the content of those theories, it will be less vulnerable to the objection that it does not directly address the claims that core-apriorist theories make. Second, it claims nothing controversial about the cognitive abilities of actual or possible moral agents, so it avoids the problem discussed in

section 2 of this chapter. Third, whilst the strategy cannot provide certainty that all core-apriorist theories are incorrect, it can at least provide considerable evidence in the core-aposteriorist's favour. If it works against the most plausible types of core-apriorist theories, then the core-apriorist will have to rely on ever more unlikely theories to escape the strategy. The unlikely theories will have their own problems before they are touched by the *a posteriori* strategy. So although the strategy has a drawback, we should examine whether the price is worth paying. I believe that it is.

Before we can use this strategy, though, we need to show that *Sufficient Determinacy* is true. I will spend the remainder of this chapter showing that it is.

Determinate moral propositions have normative content, and for moral agents to know determinate moral propositions that are not self-evident, there must be one or more premises that justify the normative content. *Sufficient Determinacy* implies that if there is only one such premise, that premise must be determinate. If there is more than one, then the proposition that is the conjunction of the premises must be determinate. We will examine both cases separately, but before starting we should recall how I am using the term 'justify'. Following what I have said in section 4 of chapter 1, if proposition X justifies proposition Y, then Y is much more likely to be true than not. If proposition X is part of a justification of proposition Y, then Y is much more likely to be true than not, and this would not be the case if X were not part of the group.[21]

Take the following moral propositions:

Nice: One should be nice to others

Drive: One should not drink-drive

Assume that *Nice* is a core moral proposition and that we can attain knowledge of *Drive* by using *Nice* as its only normative premise. If *Nice* is *a priori*, this means that the argument for *Drive* is an *a priori* method to attain knowledge of *Drive*, since the normative content of the argument is provided by *Nice*.

Suppose that *Drive* is a determinate moral proposition. *Sufficient Determinacy* denies not that *Nice* is *a priori*, but that *Nice* can by itself be used as the sole normative premise for *Drive*.

---

[21] This last condition ensures that X is not a redundant member of the group.

The term 'nice' can be fleshed out in numerous ways, so if 'nice' has such imprecise application conditions, how can we use it in any justification of determinate moral propositions? Take the following argument:

1) One should be nice to others.
2) Drink-drivers are not nice to others.
   Therefore
3) One should not drink-drive.

We cannot verify the soundness of this argument without further information. How do we check the truth of (2) without working out first what the application conditions of 'nice' are? If the application conditions are too imprecise, then different agents will have different understandings of the terms, and they will all understand the argument differently. We thus need to tighten up the application conditions of 'nice' so that (1) becomes more determinate than it previously was. However, that means that *Nice*, as it stands, cannot be used in the argument (1)-(3). It will be replaced by *Nice**, a more determinate premise, which can be used to justify *Drive*. *Nice* fails to justify *Drive* because it contravenes *Sufficient Determinacy*. *Nice** does not.

Replacing *Nice* in the argument with another proposition that is just as non-determinate will not help. Whatever non-determinate proposition we replace it with, we will get the same problem. Different agents will interpret the terms in the proposition in different ways, because the terms' application conditions will be imprecise and open to multiple interpretations.

Another argument for *Drive*, with more determinate premises, is

4) One should not perform acts that are likely to kill and injure others.
5) People who drink-drive perform acts that are likely to kill and injure others.
   Therefore
6) One should not drink-drive.

If this argument is sound, then (4), being the single normative premise in the argument, complies with *Sufficient Determinacy*. One could try arguing against *Sufficient Determinacy* by

claiming that (4) must itself be justified by normative premises that contravene *Sufficient Determinacy*. For example, suppose that (4) is justified by the following argument:

7) One should be nice to others.

8) One is nice to others only if one does not perform acts that are likely to kill and injure others.

   Therefore

9) One should not perform acts that are likely to kill and injure others.

The problem with this argument is that it is too similar to the argument (1)-(3). Just as we cannot verify the truth of (2) without making (1) more determinate, we cannot verify the truth of (8) without making (7) more determinate. We still need to give the term 'nice' more precise application conditions in order to work out exactly when an agent is nice and under what conditions he should be nice (Respecting the rights of others, for example, may sometimes entail being not very nice at all.). If there is a lot of disagreement, then the term 'nice' is not doing very much. It seems to be obscuring the truth rather than illuminating it.

So we still have to make the application conditions of 'nice' more precise. When we do that, *Nice* becomes *Nice**, and the original problem reappears because *Nice** is a determinate proposition. *Nice* is not meaningless, but it fails to justify the normative content of determinate moral propositions because such propositions themselves contain terms with fairly precise application conditions. To justify them, therefore, we need more determinate premises than *Nice*. When we say 'Don't hit your brother because it isn't nice', we are making a number of assumptions about what being nice entails – not causing unnecessary pain, or respecting the brother's rights, or having certain virtues, and so on. We assume that the listener shares these assumptions, or imply that they should.

We can note here that we cannot provide counterexamples to *Sufficient Determinacy* by appealing to the distinction between abstract and concrete moral propositions, or the distinction between general laws and particular cases. We saw in section 3 of chapter 1 that such distinctions are not the same as the distinction between determinacy and non-determinacy. Consequently, providing examples of general laws that justify the normative content of particular

cases, or examples of abstract moral principles that justify the normative content of concrete moral principles, will not show that *Sufficient Determinacy* is wrong.

We now turn to the second case, which we can examine by looking at W.D. Ross's core-apriorist moral theory. Ross believes that there are *a priori* self-evident *prima facie* duties. The duties are *prima facie* because they are not actual duties, as they can be overridden in various circumstances. Ross only reluctantly calls them duties for this reason. They are, however, related to duties in a special way.[22] For example, you may have a duty to keep a particular promise because there is a *prima facie* duty to keep promises in general, but that does not mean that we have a duty to keep promises in all circumstances. If there are sufficiently good reasons for breaking a promise, such as avoiding the deaths of innocent people, you should break it.[23] *Prima facie* duties are also non-determinate, not because they can be overridden, but because it is an open question what the circumstances in which they can be overridden are.

Because there is always a possibility of *prima facie* duties being overridden, we cannot generate a list of such duties as a set of hard-and-fast moral rules, but they certainly feature in justifications of morally right acts. They are, to a certain extent, the default position. Generally speaking, we should keep promises, even when the net benefit is slightly lower than if we did not.[24] This means that when we try to ascertain the truth of the determinate moral proposition 'Given conditions C, I should keep my promise that I will X' (where C and X have sufficiently precise application conditions), part of the argument in favour of its truth will be that there is a *prima facie* duty to keep a promise. The argument will not be valid without it.

I want to highlight here the *prima facie* duties' status as non-determinate. Take 'One should generally keep promises'. This *prima facie* duty does not tell you specifically when to keep promises, just that you generally should. Furthermore, there is no single principle that says that one duty trumps another. That depends on a variety of factors that are not self-evident. So the *prima facie* duty is non-determinate. Other *prima facie* duties are even more non-determinate, such as the *prima facie* duty to 'improve the conditions of others in respect of intelligence'.[25] Does this mean that we have a duty to make others more intelligent? In what way? What types

---

[22] Ross (2002), p. 20.
[23] Ross (2002), p. 19.
[24] Ross (2002), pp. 34-35.
[25] Ross (2002), p. 21.

of intelligence are we talking about? It is difficult to work out just what the duty tells us to do, which entails that it will not tell us what to do in particular situations.

We can use Ross's theory to see how *Sufficient Determinacy* deals with multiple premises for the normative content of a determinate moral proposition. Take the following argument:

10) You have a *prima facie* duty to keep promises.

11) This *prima facie* duty can only be disregarded if the set of non-normative circumstances X applies.

12) You have promised to pay back today the money I lent you.

13) Regarding your promise in (12), nothing in the set of circumstances X applies.

14) The *prima facie* duty to keep promises holds regarding your promise in (12).

Therefore

15) You should keep your promise to me.

Assume that (10)-(13) are true, and that 'X' can be clearly defined. (10) and (11) are normative premises. This is clear in the case of (10), since it just expresses a *prima facie* duty. (11) outlines more determinately when one should abide by the *prima facie* duty, and when it is permissible to ignore it. (12) and (13) are not normative.

The conclusion is a determinate moral proposition. We know what the promise is, who has made it, and so on. Its normative content is justified by (10), which is non-determinate, and (11), which is determinate.[26] Following what I have said about single propositions that justify the normative content of determinate moral propositions, it cannot be the case that we can alter the argument to drop (11) and keep (10) as the sole premise justifying the normative content of (15). However, we can drop (11) and alter (10). Suppose we change (10) to

10*) You have a duty to keep promises unless the set of non-normative circumstances X applies.

---

[26] The core-aposteriorist can therefore argue against Ross on the grounds that even if *prima facie* duties are knowable *a priori*, they are non-determinate, and so cannot justify determinate moral propositions on their own. We need to combine them with other premises that are only knowable *a posteriori*. Since we must justify determinate moral propositions with a combination of *a priori* and *a posteriori* premises, this entails that we must use a mixed method, and so the CAT is false. I will not pursue this argument here, though.

(10*) is equivalent to the conjunction of (10) and (11). It is a single normative premise to justify the normative content of (15). It also complies with *Sufficient Determinacy*, as it itself is determinate. It tells us precisely when we have a duty to keep promises.

We can now see how *Sufficient Determinacy* implies that the conjunction of the premises that justifies the normative content of a determinate moral proposition is itself determinate. We can keep (10) and (11), or replace them with (10*). (10*) is determinate, since the application conditions of the terms in the proposition are sufficiently precise for us to know in most cases whether we should keep promises or not. As (10*) is equivalent to the conjunction of (10) and (11), the conjunction must also be determinate. Naturally, the core-aposteriorist will want to add that (10*) is not knowable *a priori*, but this is an additional claim. *Sufficient Determinacy* itself makes no claims about the *a priori* or the *a posteriori*.

There is thus good reason to believe that *Sufficient Determinacy* is true. Non-determinate propositions cannot by themselves justify the normative content of determinate moral propositions. Only determinate propositions can. This means that we can use *Sufficient Determinacy* in the *a posteriori* strategy. In the following chapters, I will use the strategy against four particular core-apriorist theories.

Chapter 3 – The Sidgwick, de Lazari-Radek and Singer theory

Whilst core-apriorists agree that there are *a priori* methods to know a determinate core moral proposition, they can disagree over what those methods are. We can broadly divide core-apriorists into two camps, intuitionists and coherentists, according to the particular methods that they choose.

This division is epistemological rather than moral, because one can be an intuitionist or a coherentist regarding non-moral types of knowledge as well. The difference between the two types of theory is shown by a problem with justification. If proposition A is justified by proposition B, we can ask what justifies proposition B. Suppose that proposition B is justified by proposition C. What justifies proposition C? If it is justified by proposition D, what justifies proposition D? An infinite regress opens up, as each new proposition introduced to justify the others will always need justification itself. Intuitionism and coherentism propose different solutions to this problem.[1]

Moral intuitionist theories claim that there are self-evident propositions, which justify themselves without the need to appeal to any other proposition, and the regress thus stops at that point. I use the term 'moral intuitionism' in this thesis in the same way that Hallvard Lillehammer uses the term 'ethical intuitionism'. It is 'the thesis that some ethical claims (such as self-evident principles of practical reason) can play the role of non-inferentially grounded justifiers or defeaters in ethical theory'.[2] This does not mean that the moral intuitionist thinks that the claims are justified by some type of moral intuition that moral agents have. He might do, but it is far from essential, as we will see in this chapter. We should also note that sometimes moral intuitionist theories are taken to say that we must be able to know self-evident moral claims thanks to some special epistemological faculty that we have. A common objection to such theories is that it is utterly mysterious how such a faculty is meant to work.[3] However, moral intuitionist theories need not make such a claim, so I will ignore this objection here.

Moral coherentist theories claim that moral propositions must form a coherent set. Each proposition in that set is justified by other propositions in that set. The regress becomes

---

[1] Intuitionism and coherentism are not the only possible solutions to this problem, but they are the main standard answers.
[2] Lillehammer (2011), p. 80.
[3] E.g. Mackie (1977), pp. 38-39.

harmless, because we are not always appealing to new propositions to justify the propositions that we have already identified.[4] A maximally coherent set of moral propositions is the largest coherent set of moral propositions such that if any other moral proposition were added, the set would become incoherent.

The core-apriorist can freely choose whether to be a moral intuitionist or a moral coherentist. However, introducing the distinction divides core-apriorist theories into two types. If the core-apriorist is a moral intuitionist, he will claim either that all determinate core moral propositions are *a priori* self-evident (thus rejecting step (a) of the *a posteriori* strategy) or that they are justified by *a priori* premises that are themselves self-evident (thus rejecting step (b) of the *a posteriori* strategy). If the core-apriorist is a moral coherentist, the core-apriorist will accept step (a), as he does not believe that determinate core moral propositions are self-evident, but will reject step (b). In this chapter and the next, I will concentrate on moral intuitionist core-apriorist theories.[5] In chapter 5, I will turn my attention to two moral coherentist core-apriorist theories.

By endorsing moral intuitionism, the core-apriorist hopes to show either that true determinate core moral propositions can be known *a priori* non-inferentially, or that true determinate core moral propositions are justified by *a priori* premises that can themselves be known non-inferentially. In this chapter, I examine what I will call the SLS theory (the Sidgwick, de Lazari-Radek and Singer theory). This core-apriorist theory has recently been developed by Katarzyna de Lazari-Radek and Peter Singer, based on their interpretation of Henry Sidgwick's moral philosophy. The second intuitionist theory, which I will examine in the next chapter, is Matthew Kramer's moral realism as a moral doctrine. The theories differ substantially in their details, but the core-aposteriorist can effectively use the *a posteriori* strategy to criticise them.

Henry Sidgwick bases his metaethics on a set of what he takes to be self-evident axioms, where 'self-evident' can be taken as another way of saying 'true and not justified by appeal to other propositions'. This does not mean that the axioms cannot be inferred from other propositions, and Sidgwick allows that this may be possible.[6] His point is that they do not need

---

[4] An obvious objection is that this justification is circular, and so moral coherentist theories are mistaken. I will not examine this objection here, but moral coherentists have responded to it (e.g. Brink (1989), pp. 122-125).
[5] Intuitionism was particularly popular in the early twentieth century, with prominent intuitionists including G.E. Moore (1993) and W.D. Ross (2002). More recently, it has seen a revival thanks to intuitionists such as Russ Shafer-Landau (2003) and Robert Audi (2004). Ross, Shafer-Landau and Audi are core-apriorists.
[6] Sidgwick (1879), pp. 106-107.

to be. Whilst he did not think that the axioms could justify all true moral propositions by themselves, he did believe that they are necessary for such justifications. Let us take Sidgwick's moral axioms to be at least some of the core moral propositions that the SLS theory endorses.

De Lazari-Radek and Singer believe that the axioms can be rationally ascertained *a priori*. I will specifically argue that at least some of Sidgwick's axioms in their original wording cannot be used to justify the normative content of determinate non-core moral propositions, since the axioms contravene *Sufficient Determinacy*. We can alter these axioms to correct this problem, but the amended axioms turn out to be knowable only *a posteriori*. Hence, if these axioms are core moral propositions, some of the core moral propositions are only knowable *a posteriori*. I will also argue that the SLS theory's response to Sidgwick's dualism of practical reason, and its theory of the good, can only be defended on *a posteriori* grounds.

Before starting, though, we should note a couple of points about reading Sidgwick's metaethics as a core-apriorist theory. There is a danger of attributing anachronistic views to Sidgwick if we are not careful.

1 – A historical introduction

The first point relates to terminology. Moral intuitionist theories, as I define them, need make no claims about moral intuitions. This is particularly important when examining Sidgwick's metaethics, because Sidgwick does not use the term 'intuitionism' as I use it. For Sidgwick, intuitionism claims that 'the common moral sense of ordinary mankind' can be used to attain moral knowledge.[7] As I use the term, intuitionism need make no claims about common moral sense, and I will entirely ignore Sidgwick's use of the term 'intuitionism' in this context.

De Lazari-Radek and Singer follow Sidgwick's use of the term. Elsewhere, Singer has been very wary of moral intuitions. He believes that we do not reach many of our moral intuitions by rationally deliberating, and for that reason we cannot use them as justifications of moral judgements.[8] As Singer reads Sidgwick, when Sidgwick appeals to moral intuitions in support of utilitarianism (Sidgwick's preferred normative moral theory), this is simply to confirm a result that

---

[7] Sidgwick (1981), p. 200. See also pp. 96-97.
[8] Singer (2005).

Sidgwick also argues for on other independent grounds.[9] Sidgwick's point, according to Singer, is that even if we are intuitionists in Sidgwick's sense of the term, we can accept his utilitarianism based on intuitionist grounds.

Sidgwick's metaethics is intuitionist (in my sense of the term) because the self-evident axioms that Sidgwick bases his moral theory on are non-inferentially grounded justifiers or defeaters of moral claims. This brings us to the second point, which is how far we can treat Sidgwick's metaethics as a core-apriorist theory. The core-apriorist theory that I want to examine may not entirely be Sidgwick's, particularly since I am looking at it through de Lazari-Radek and Singer's interpretation. This is important because Sidgwick does not appear to concentrate very much on whether his axioms are *a priori* or *a posteriori*.[10] He is more interested in showing that they are self-evident, which does not imply that the axioms are *a priori*. If I look at a patch of blue, it may be self-evident to me that the proposition 'I am looking at something blue' is true, but the proposition is not justified by appeal to any other propositions. However, the proposition is not *a priori*. A core-apriorist may assume that, given the nature of moral propositions, any self-evident moral propositions must be *a priori*, because there do not seem to be any *a posteriori* methods to attain knowledge of a self-evident moral proposition. Unlike the colour case, for example, it is dubious that we can tell that something is self-evidently morally wrong solely by perceiving it. However, I will not pursue this line of thought. It is important to note, however, that even if Sidgwick's axioms are not self-evident, this does not imply that they are not *a priori*. The core-aposteriorist therefore cannot defend the Central Claim merely by showing that the axioms are not self-evident.

Whilst Sidgwick might not agree that his metaethics is a core-apriorist theory, it is easy to develop it as one. For example, Sidgwick was well aware of the ethical controversy that Charles Darwin had caused with the theory of evolution.[11] If all our moral beliefs are the product of blind evolution, asked some critics, does that not show that they have no justification? Even if we reject this criticism, the idea that evolutionary research can influence moral philosophy is very much alive today. How far such research should affect our moral views is a hotly debated topic. But in his major work *The Methods of Ethics*, Sidgwick never mentions evolutionary theory in

---

[9] Singer (1974), p. 498.
[10] Hallvard Lillehammer points out that saying that Sidgwick's axioms are knowable *a priori* may be anachronistic ((2010), p. 373).
[11] Darwin (2004), part 1, chapter 4.

connection with his self-evident axioms. Elsewhere, he even writes that '[e]volution… has little or no bearing upon ethics'.[12] He argues that merely finding out how a principle (whether moral or non-moral) is derived tells us nothing about its truth or falsity, which is a point that we have seen W.D. Ross also make. Sidgwick argues that if we should distrust our moral judgements just because they are an evolutionary product, we should distrust all our judgements, as they are evolutionary products as well.

Evolutionary debunkers such as Sharon Street would probably point out that moral judgements are not like other judgements.[13] If a moral judgement aids the survival of the agent making the judgement, claims Street, the survival value of the judgement is unrelated to its truth. The survival value of non-normative judgements is often related to their truth, though. If I judge that there are no tigers near me when there are, I will be more likely to become a tiger's lunch than I would be if I had a true belief about the tigers nearby. However, I will not investigate Sidgwick's argument here. I use it simply to show that Sidgwick seems to believe that *a posteriori* research on the formation of moral judgements says nothing about an agent's ability to attain moral knowledge.

So whatever Sidgwick's exact attitude towards core-apriorist theories would have been, it does not take much to see his metaethics as a core-apriorist theory. De Lazari-Radek and Singer take it as such. It is not necessary for my purposes to dig too deeply into philosophical history here. Maybe the real Sidgwick is to a certain extent different from de Lazari-Radek and Singer's Sidgwick, but not so much that they are presenting an imaginary figure. However, since we cannot be certain about Sidgwick's exact attitude, I cannot say that the core-apriorist theory examined in this chapter is Sidgwick's theory. This is why I call it instead the SLS theory, to mark it as de Lazari-Radek and Singer's development of Sidgwick. The SLS theory claims that there are particular self-evident axioms that are knowable *a priori* and that can be used to attain moral knowledge. If the axioms are core moral propositions, then the SLS theory is a core-apriorist theory.

There is one more historical point to make, this time about Sidgwick's 'dualism of practical reason'. However, I will leave this discussion until after we have seen the self-evident axioms that Sidgwick identifies.

---

[12] Sidgwick (1876), p. 54.
[13] Street (2006). See also Lillehammer (2010), p. 368.

2 – What self-evident axioms are there?

Sidgwick thinks that there are only a few self-evident axioms that a plausible moral system will be based on. He lists four conditions that would 'establish a significant proposition, apparently self-evident, in the highest degree of certainty attainable'.

1) 'The terms of the proposition must be clear and precise.'

2) 'The self-evidence of the proposition must be ascertained by careful reflection.'

3) 'The propositions accepted as self-evident must be mutually consistent… [I]t is obvious that any collision between two intuitions is a proof that there is error in one or the other, or in both.'[14]

4) '[T]he denial by another [judge] of a proposition that I have affirmed has a tendency to impair my confidence in its validity.'[15]

We will see shortly that criterion (1) is extremely important, so we should clear up a possible misunderstanding. The words 'clear and precise' in the criterion are Sidgwick's words, and he does not use them as I do. My use of the word 'precise' applies to the application conditions of terms with a proposition. For example, a determinate moral proposition contains terms that have very precise application conditions. The less precise the application conditions are, the less determinate the proposition will be. However, Sidgwick is not talking about the application conditions of a term. Furthermore, he concedes that the axioms alone are not sufficient to justify many moral propositions. We need to add a normative moral theory (such as utilitarianism). I will say more about this shortly. In any case, I do not want to claim that Sidgwick uses the word 'precise' in exactly the same way that I do.

An example from de Lazari-Radek and Singer can help show how to interpret (1). They take the candidate axiom 'We ought to give every man his own' and argue that 'his own' is not sufficiently clear and precise.[16] It is not at all clear what 'his own' is supposed to mean. We can

---

[14] Sidgwick is slightly imprecise here, because this is not a condition for an axiom, but for a group of axioms. However, it makes no important difference.
[15] Sidgwick (1981), pp. 338-342.
[16] De Lazari-Radek and Singer (2014), p. 115; Sidgwick (1981), p. 215.

of course change 'his own' into something clearer, and so make the candidate axiom more precise, but we must be careful. First, we could end up with a vacuously analytic proposition, such as 'We ought to give each man what it is right that he should have'. Second, we could end up with a proposition that contravenes the other criteria.

The criteria do not say that the self-evident axioms must be able by themselves to justify any true determinate moral propositions. Sidgwick addresses this point, as he criticises purported axioms that do not 'enable us to give clear and unhesitating decisions'.[17] If a set of purported axioms leaves us completely in the dark about what moral judgements we should make, they are not worth much. The reason why determinacy (as I use the term) is not implied by criterion (1) is because a proposition can be a self-evident axiom even if it cannot directly justify determinate moral propositions. Axioms need not be determinate themselves. Sidgwick's aim is to have a set of axioms that enable us to give clear and unhesitating moral decisions when we join them to an appropriate normative moral theory. He believes that the most appropriate theory is utilitarianism. Therefore we can say that we have to alter the axioms by adding the 'other resources' of utilitarianism, so that the axioms are determinate. The axioms thus become determinate core moral propositions, and can be used to ascertain the truth of the normative content of determinate non-core moral propositions.[18]

Sidgwick states that

'[C]ertain absolute practical principles [that are manifestly true]… are of too abstract a nature, and too universal in their scope, to enable us to ascertain by immediate application of them what we ought to do in any particular case; particular duties have still to be determined by some other method.'[19]

By 'method', Sidgwick means the application of a normative moral theory, such as utilitarianism or egoistic hedonism.[20] De Lazari-Radek and Singer argue that Sidgwick can use the resources of utilitarianism together with his self-evident axioms to produce a determinate utilitarian moral

---

[17] Sidgwick (1981), p. 215.
[18] The term 'other resources' is de Lazari-Radek and Singer's ((2014), p. 148).
[19] Sidgwick (1981), p. 379.
[20] Sidgwick's use of the term 'method' is thus very different to my own use, and there is no necessary connection between the two.

system.[21] Sidgwick objects to purported axioms that cannot provide justification for what we should believe or how we should act even when we have settled on a particular normative moral theory.

I shall follow de Lazari-Radek and Singer in their identification of Sidgwick's self-evident axioms.[22] They are:

Treatment: '[I]t cannot be right for A to treat B in a manner in which it would be wrong for B to treat A, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment.'

Prudence: '[One ought to aim at] one's good on the whole… [where] Hereafter *as such* is to be regarded neither less nor more than Now.' (A good does not become lesser or greater simply because it will happen sooner rather than later, or later rather than sooner.)

Impartiality: '[T]he good of any one individual is of no more importance, from the point of view… of the Universe, than the good of any other; unless, that is, there are special grounds for believing that more good is likely to be realised in the one case than in the other.'

Aim: '[A]s a rational being I am bound to aim at good generally, - so far as it is attainable by my efforts, - not merely at a particular part of it.'[23]

We can derive the 'maxim of Benevolence' from *Impartiality* and *Aim*:

Benevolence: '[E]ach one is morally bound to regard the good of any other individual as much as his own, except in so far as he judges it to be less, when impartially viewed, or less certainly knowable or attainable by him.'[24]

---

[21] De Lazari-Radek and Singer (2014), pp. 147-148.
[22] It is not entirely clear how many separate axioms Sidgwick identifies, but we need not settle the matter here.
[23] Sidgwick (1981), pp. 379-382, emphasis in the original.
[24] Sidgwick (1981), p. 382.

The axioms are not analytic, and hence they are not vacuous. However, they do not by themselves justify determinate moral propositions. Should we ban slavery? *Treatment* only tells us that if A can (or cannot) be enslaved, then B can (or cannot) be either, just because they are different individuals. *Impartiality* gives us no way to measure the goods of individuals, so it cannot tell us if the good produced by slavery is more or less than the good produced otherwise. This also affects *Aim*. Importantly, the term 'good' in the axioms does not mean 'moral good'. It refers to any type of good, moral and non-moral.

However, we have already seen that we need to alter the axioms by adding the 'other resources' of a normative moral theory to them to solve this problem. We can use the following method to justify a true determinate moral proposition:

a) One or more of the self-evident axioms must be a premise in the argument for the proposition.

b) The 'other resources' of a normative moral theory (such as utilitarianism) may provide other premises. The normative content of the axioms and the other resources must be knowable *a priori*, otherwise the SLS theory is not a core-apriorist theory.

c) There may be other non-moral facts that are relevant to the justification of the determinate moral proposition. These facts can be used as premises, and may be ascertainable *a priori* or *a posteriori*. However, they do not have normative content. The normative content is provided by the axioms and the normative moral theory.

d) The premises collectively justify the determinate moral proposition. As the self-evident axioms and the 'other resources' of the normative moral theory are *a priori*, this entails that the determinate moral propositions that they justify have an *a priori* basis, and hence the SLS theory is a core-apriorist theory.

Direct act utilitarians hold that acts are morally right if they produce the greatest amount of utility, and that we are always required to perform such acts. They might argue that agents should not drink-drive as follows:[25]

---

[25] Sidgwick is not a direct act utilitarian. I use this argument only as an example.

e) Sidgwick's self-evident axioms are true.

f) *Benevolence* is true. (from (e))

g) Morally right acts are those that produce the greatest utility for everyone in general. (from (e), (f), and the 'other resources' of direct act utilitarianism)

h) You should perform acts that produce the greatest utility for everyone in general. (from (f) and (g))

i) Refraining from drink-driving produces greater utility for everyone in general than drink-driving does (*a posteriori* observation)

Therefore

j) You should not drink-drive. (from (h) and (i))

This gets us from self-evident axioms to determinate moral propositions. The axioms cannot justify determinate moral propositions on their own, but they are premises in the above argument, and so we can attain *a priori* determinate moral knowledge since the normative content of (j) is justified by the self-evident axioms.

There is a final important point. Why use utilitarianism as the normative moral theory? Sidgwick ultimately wants to show that utilitarianism is rationally justified by the self-evident axioms. If he can do this, then we get the result that the axioms justify the normative theory, and the normative theory in turn can be used in conjunction with the axioms to justify true determinate moral propositions. There is a problem with this claim, called the dualism of practical reason. I will return to this problem in section 6, as it is significant for the SLS theory. For now, though, I place it to one side.

3 – A closer look at *Aim*

In this section, I am going to argue that *Aim* in its current form contravenes *Sufficient Determinacy*, and any attempt to alter it so that it is more determinate carries a high risk that the altered version is only knowable *a posteriori*. This means that the core-apriorist cannot use it as an example of an *a priori* determinate core moral proposition. Other axioms, though perhaps not all of them, suffer from the same problem.

The axioms do not merely apply to moral good, but to good in general (Remember that *Benevolence* follows from the axioms; it is not an axiom itself.). De Lazari-Radek and Singer are explicit about this. When they discuss examples of *Prudence*, for instance, their examples do not concern moral good. The axioms apply to normative good in general, which can come in many forms – prudential, aesthetic, financial, and so on. For instance, in many cases it is good to have laws that comply with the axioms. This is partly because it could be morally good to have such laws, but this need not be the only reason. It is morally good to ban slavery, but it is also good for me on grounds of self-interest. It ensures that nobody can legally enslave me. I do not need to think it morally wrong to be enslaved; I might just want to avoid it because I think it would be unpleasant. Similarly, a legal system that was not impartial may become unworkable and lead to social breakdown. This would be a moral failure, but I might also fear it for personal, non-moral reasons. It might affect my friends, whom I care about because they are my friends, rather than because of some specifically moral reason.

When making laws, we have to ask what we want from them. What is the aim? How can we best achieve it? It cannot be assumed that we can discover this *a priori*. We may want a well-functioning society because we have seen *a posteriori*, or can predict *a posteriori* what happens if our society does not function well. This leads to a suspicion that Sidgwick's axioms may presuppose some facts about human nature. Maybe they are *a priori* only given the assumption that they apply to humans, so they may not be true if we tried applying them to aliens.[26] The core-apriorist could object that whatever type of good humans aim at, they would still be aiming at the good, and it is *a priori* that neither humans nor aliens can (for example) rationally choose a lesser good over a greater good. *Aim* and the other axioms apply to every type of good. But how easy is it to give a plausible interpretation of *Aim*?

Suppose I discover that the morally best thing to do would be to sell all I have, donate the proceeds to charity, go to a severely impoverished country, and spend my life performing charitable work that I would hate. This would make the world a slightly better place than if I continued my current lifestyle. According to *Aim*, I am bound to aim for the good, but it would be hard to say that it would be rational for me to change my life like this. I might morally applaud someone who did, but that does not mean that it is a rational action, or that I would be rational

---

[26] Lillehammer (2010), pp. 372-374.

to do the same myself. If I did, it would look rather like moral fetishism. Moral fetishists are agents whose every act is dictated by what they believe to be morally best, regardless of any other considerations. For example, they choose to do A rather than B because A is slightly morally better than B, even though A is much more difficult to perform (physically, psychologically, or in some other way), or even though B has much more non-moral value than A.

We can also use cases of non-moral fetishism to create problems for *Aim*. Consider the case of parents with a severely disabled child. It does not seem irrational for them to give up great sources of personal good to obtain extra goodness for him, just because he is their child, and even though his personal goodness has no chance of matching theirs. Neither does it seem intrinsically irrational for society in general to do the same for a group of disadvantaged people, simply because they are members of society who need help. *Aim* does not appear to allow for this possibility. Aiming at the good 'generally', as *Aim* demands, appears to mean aiming at everyone's good, not simply at the good of a disadvantaged minority.

The problem of moral fetishism is not simply a possible danger for utilitarianism, or even for consequentialist theories generally. If a Kantian decides that she should adhere to the categorical imperative at all times and in all situations, regardless of any other considerations, this too looks like a form of moral fetishism (Even if it is not, the Kantian must explain why it is not.). Lying is forbidden, claims the Kantian moral fetishist, whatever the cost, and that cost may destroy lives.[27] The idea that moral agents can and must abide by every true moral norm, no matter how great the difficulty or how trivial the moral benefit, is not obviously true, even if it is true.

Sidgwick seems aware of the danger of moral fetishism. He knows that if we are always consciously aiming at maximising utility, it may be counter-productive. We may gain more utility by not consciously aiming at it.[28] Sidgwick even writes that it may be better from a utilitarian perspective if many people are not utilitarian.[29]

I want to examine the idea that if the interpretation of *Aim* that we have been using is wrong, there may be other, more plausible interpretations. If I always aim at good generally, it would not only be psychologically exhausting, but counter-productive. So perhaps *Aim* is not meant to

---

[27] See Langton (1992).
[28] Sidgwick (1981), pp. 405-406, 413.
[29] Sidgwick (1981), pp. 489-490.

be read as saying that on every occasion we must aim at good. If there are multiple ways that we can read *Aim*, though, then the terms in *Aim* as it currently stands have rather imprecise application conditions. If we read *Aim* as meaning that we are always obliged to aim at good generally, then we are going to believe that *Benevolence* morally requires us to always regard the good of others as having the same weight as our own good. If we read *Aim* as meaning that our general aim should be the good, but that we do not have to aim at the good at every opportunity, then that will lead to a different reading of *Benevolence*. Maybe occasionally not aiming at the good is morally permissible even though generally it is not.

The objection I am developing is not directed against utilitarianism, but I will continue using utilitarianism as an example to show what the target really is. Direct act utilitarians may read *Aim* as

A-Aim: As a rational being, I am bound to aim at good generally with my every act, not merely at a particular part of it.[30]

The wording of *A-Aim* keeps as closely as possible to the wording of *Aim*, and I have only altered it to specifically apply to direct act utilitarianism.[31] *A-Aim*'s problem is that direct act utilitarianism appears to encourage a form of fetishism. Aiming for the greatest good with every act is highly difficult (if not impossible) and often counter-productive. This makes *A-Aim* extremely dubious, but utilitarians can turn instead to indirect forms of utilitarianism. Rule utilitarianism holds that rules, rather than acts, are morally right if overall they produce the greatest amount of utility, and that we should follow these rules.[32] For example, rule utilitarians may sometimes not perform a particular act that maximises utility, because the rules they follow tell them not to. The justification for this is that following the rules will in the long run produce a greater amount of utility than not following them.

Rule utilitarians could read *Aim* as

---

[30] Given Sidgwick's wording of *Aim*, I take it as an axiom about how agents should act, as opposed to what makes something good. I will discuss the SLS theory's theory of the good later.
[31] I do not imply here that direct act utilitarians (or any other utilitarians) must accept Sidgwick's axioms in some form. They may justify their utilitarianism in some other way. I am only examining how they read Sidgwick's axioms if they do accept them.
[32] E.g. Brandt (1979), chapter XV.

R-Aim: As a rational being, I am bound to always follow a set of rules that aim at good generally, not merely at a particular part of it.

It is harder to argue that holding *R-Aim* is highly difficult, impossible or counter-productive, but it may be done. If following a rule means that I should drown the disabled, then surely I need better justification for following the rule in this situation than that doing so in general will maximise utility. We need to amend the rule. However, if we change the rule to accommodate too many amendments, then there is a danger that rule utilitarianism will collapse into direct act utilitarianism, which simply requires agents to perform acts that maximise utility.

The rule utilitarian can argue back. For example, he can claim that just as direct act utilitarianism is practically impossible to live by, a rule utilitarianism that has too many amendments to its rules would suffer from the same problem.[33] Any reasonable rule utilitarianism must therefore be different, and there would be a limit on the number of amendments that could be made to the rules. However, utilitarians might instead opt for another form of indirect utilitarianism, motive utilitarianism. Motive utilitarianism states that

'one pattern of motivation is morally better than another to the extent that the former has more utility than the latter. The morally perfect person, on this view, would have the most useful desires, and have them in exactly the most useful strengths; he or she would have the most useful among the patterns of motivations that are causally possible for human beings.'[34]

The motive utilitarian could read *Aim* as

M-Aim: As a rational being, I am bound to always follow patterns of motivation that aim at good generally, not merely at a particular part of it.

This might solve the fetishism problem, because the moral agent would not always act to produce the morally best outcome on every occasion, and he can avoid the problems I have

---

[33] For a fuller response, see Hooker (2002), pp. 93-99.
[34] Adams (1974), p. 470.

mentioned with rule utilitarianism. All he needs is a set of motives that overall motivates him to produce greater utility than otherwise. This set of motives may be able to avoid motives that would lead to moral fetishism.

So the fetishism objection is only obviously a problem for the direct act utilitarian. We do not need to read *Aim* in a way that makes it vulnerable to the objection. But note the result: as *A-Aim*, *R-Aim* and *M-Aim* make clear, there are many different types of interpretation that we can choose. Plausible interpretations seem to be plausible, in part, because they take into account what human beings and their social world are actually like. We might be able to know with certainty which acts have the best results if we could see the future, or if we had enough time to calculate, or if we had sufficient information and cognitive skills, but we do not. We might have been able to perform the morally best act in every situation if our brains were wired differently, or if we never got tired, or if we were psychologically capable of giving up all our other goals, but life is not actually like that.

This leads to a problem for the core-apriorist. As it stands, *Aim* is in danger of contravening Sidgwick's criteria (1) and (4). It is not sufficiently clear how we are meant to apply *Aim*, and we can reasonably disagree with others about how we are supposed to apply it in arguments for determinate moral propositions. Because of this, it also contravenes *Sufficient Determinacy*. The terms in *Aim* do not have precise enough application conditions for it to be capable of justifying determinate moral propositions. It thus cannot help the core-apriorist, even if it is *a priori*. We can make the application conditions more precise (for example, by amending *Aim* so that it becomes *R-Aim*), but this just creates a new problem. It appears that we make the application conditions more precise in part by using information from *a posteriori* sources, which tell us what human beings are reasonably capable of doing and deciding. By amending *Aim* to become *R-Aim*, we use the other resources of rule utilitarianism, but these resources are not entirely *a priori*. *R-Aim* may be a determinate core moral proposition, as its terms may have sufficiently precise application conditions, and it can be part of the set of propositions that justify all other moral propositions. However, since it is not knowable *a priori*, a moral theory that uses it as a determinate core moral proposition cannot be a core-apriorist theory.

Since *Aim* in its original formulation risks contravening Sidgwick's criteria (1) and (4), it is unlikely to be a self-evident axiom, whether it is *a priori* or *a posteriori*. As far as (4) is

concerned, *Aim* is not uncontroversial in any case, as non-consequentialists are going to be suspicious of it. De Lazari-Radek and Singer ask

'Would it be morally right for you to seek your own happiness if doing so meant sacrificing the greater happiness of another human being, without counterbalancing gain to anyone else? Sidgwick asserts… [that] as a matter of conscience, the answer would be 'unhesitatingly in the negative'… [Sidgwick attempts to] demonstrate that the self-evidence of the axiom of benevolence is the only way in which utilitarianism can be put on a rational basis…'[35]

But would it be morally right for you to seek *another human being*'s happiness if doing so meant sacrificing your own greater happiness, without counterbalancing gain to anyone else? Perhaps not, but the parents of the disabled child may well disagree. Our own response would probably not be 'unhesitating' either. Utilitarians, though, have to answer this question with 'no', if they also answer de Lazari-Radek and Singer's original question with 'no'. If granting the axioms entails agreeing with utilitarians on both questions, non-consequentialists may have a hard time granting the axioms.

Sidgwick knew that the axioms do not get us all the way to utilitarianism by themselves. They do not tell us what the good is. However, they are very favourable to utilitarianism. *Impartiality* and *Aim*, together with *Benevolence*, imply that we should be consequentialists, so non-consequentialists will find the axioms highly dubious. Non-morally, the example of the parents with the disabled child shows that we might reasonably prefer a lesser amount of total non-moral good depending on its distribution amongst agents. As far as moral good is concerned, suppose that we calculate that my non-moral good if you were my slave would be greater than your non-moral good would be if you were a free person. Deontologists can reasonably object that even if slavery produced net non-moral good (however one defines that good), it would be morally impermissible, since nobody has the right to enslave anyone else. Indirect utilitarians may have an answer to this complaint, but the point is that it is not at all clear that *Aim* is self-evident. Deontologists may decide to give *Aim* up, but utilitarians must retain some version of it.

---

[35] De Lazari-Radek and Singer (2014), p. 110, quoting Sidgwick (1981), p. 382. See also de Lazari-Radek and Singer (2014), pp. 120-121.

It is one of the two premises for *Benevolence*, which is key for Sidgwick's defence of utilitarianism. Therefore it is even more important for Sidgwick that it can be fleshed out in a way that makes it plausible, and this places heavier reliance on the 'other resources' of utilitarianism. This brings us back to the original problem, that *Aim* can only be fleshed out in a way that makes it *a posteriori*.

De Lazari-Radek and Singer do not really provide a defence of *Aim*. Instead, they defend *Benevolence* by defending its other premise, *Impartiality*. They also argue that whilst *Benevolence* implies that the morally best act is always morally obligatory, the alternative is just as problematic. It is just as difficult, they think, to claim that the morally best act is not always morally obligatory.[36] This allows them to avoid the problem of moral fetishism on the basis that the alternative is no better, but this response is not very convincing. Their position implies more than just moral fetishism. It also means that there can be no morally supererogatory acts. We typically believe that even if it is not morally obligatory to, say, volunteer to collect donations for cancer research charities, it is morally better than just lying in bed all day. According to de Lazari-Radek and Singer, this belief is wrong. If it is morally better to collect donations than all other options, then it is morally obligatory.

De Lazari-Radek and Singer might be right or wrong about this, but the point is that *Aim* does not appear to be self-evident if it raises all these questions. If *Aim* risks contravening criterion (1) because it is not clear and precise, then it also risks contravening criterion (4) because of the amount of reasonable disagreement that agents can have about it.

We therefore have reason to doubt *Aim* on several grounds. First, in its original form, it may contravene criteria (1) and (4) for being a self-evident axiom. Second, even if it is a self-evident axiom, it contravenes *Sufficient Determinacy* and cannot be a determinate core moral proposition. Third, we can make *Aim* more determinate by altering it so that it becomes (for example) *R-Aim*, but this can only be done by using the 'other resources' of a normative moral theory, and we may only be able to know what these resources are through *a posteriori* investigation. So even if *Aim* becomes a determinate core moral proposition, it is not knowable *a priori*.

---

[36] De Lazari-Radek and Singer (2014), pp. 133-134.

The core-apriorist may object to my use of normative moral theories. I said in the last chapter that the core-aposteriorist cannot simply use *a posteriori* information about the cognitive abilities of moral agents to attain determinate core moral knowledge. However, I now seem to be using *a posteriori* information about moral agents to (for example) dismiss direct act utilitarianism. The core-apriorist may see a contradiction here. The limitations of a moral agent end up dictating how we interpret *Aim*, but they may not be the fundamental explanation of why *Aim* is true or not. Suppose that the correct interpretation of *Aim* is *R-Aim*. *R-Aim* is true, but the fundamental explanation of why it is true may not be that it is the only interpretation of *Aim* that human agents can reasonably use to guide their moral behaviour.

This is true, but the core-aposteriorist does not need to claim otherwise, and any problems with the use of *a posteriori* evidence lies with the SLS theory, not with the *a posteriori* strategy. It is the SLS theory that insists that we must interpret Sidgwick's axioms in the light of an appropriate normative moral theory in order to discover how we should act in various circumstances, and the theory it chooses is utilitarianism. But there are different forms of utilitarianism, and we have to choose between them. It turns out that this choice will depend upon *a posteriori* information about human agents. If it did not, then direct act utilitarianism may be the correct moral theory after all, and we would be condemned to live by moral standards that we could not possibly meet. The core-apriorist, if he is to use the SLS theory, has to find a completely *a priori* justification for the normative moral theory that he chooses.

Aim thus poses serious difficulties for the core-apriorist. Do any of the other axioms?

4 – *Impartiality* and *Treatment*

Could *Impartiality* contravene *Sufficient Determinacy*? *Impartiality* can be rephrased as 'When the goods of two individuals are equal, there is nothing to choose between them. When the good of one individual is likely to be greater than the good of another individual if it is realised, then the good of the first individual is more important that the good of the second.' Phrased like this, we can see a pressure point more clearly. How are we meant to read 'important'? If we go with de Lazari-Radek and Singer's interpretation of Sidgwick, it does not mean 'moral

importance'. The axioms are not specifically moral axioms. They are, rather, axioms that we base a moral theory on.

Take the following example: if Smith chooses to φ, White will get 20 units of good and Jones's goodness will be reduced by 5 units. The total increase of goodness is 15 units. If Smith chooses to not-φ, Jones will get 10 units of goodness and White will be unaffected. According to *Impartiality*, it is more important for Smith to φ than not-φ. But this only works if all the types of goodness under examination (moral, aesthetic, prudential, legal, etc.) can be compared. There must be a way of working out that a certain amount of aesthetic goodness is equivalent to a certain amount of prudential goodness, for example. Sometimes this can be done. It seems reasonable for me to give up some financial good as donations to the National Gallery, since the financial good I give up is much less than the aesthetic good at stake. But that does not imply that this sort of calculation can always be done (even in principle), or that it always helps us work out what it is morally better to do. Suppose that White is a loan shark who employs Smith to break Jones's leg, as a result of which White's other victims will be more likely to pay up. White's (financial) good may be greater if Jones's leg is broken than Jones's (prudential) good will be if it is not, so in such a case Smith should break Jones's leg. We would be reasonably alarmed by a moral theory that permitted this.

Perhaps we should read *Impartiality* as saying that when comparing different amounts of the same type of good, the larger amount is more important. But this would give an odd meaning to 'importance'. Suppose that a billionaire could get £1,000 or a starving family could get £900, but only one of these outcomes can occur. If we examine only the financial good involved, then it is more important for the billionaire to get £1,000. This sounds a very strange thing to say. Perhaps we should define 'financial good' to include the benefits that the money could bring, in which case more financial good is produced by giving the starving family £900. But this means that the financial good is based on other types of good (in this case, prudential good), and so we are actually comparing different types of good that are included in the definition of 'financial good'.

The core-apriorist can try making a concession. He could grant that when two goods are in conflict, we have to resort to *a posteriori* investigation to work out which good we should prefer. We can do this, for example, by working out the consequences of accepting either good and

discovering the utility that results. We then pick the good that produces the most utility. What is *a priori*, claims the core-apriorist, is what those goods are. If health is a good, we discover this *a priori*. We then work out *a posteriori* whether a particular level of health is better, worse or equal to a particular level of another good, and hence whether we should prefer it to that other good.

This has two problems. The first is within the context of Sidgwick's axioms. Even if the axioms are *a priori*, they give us no way of working out what is good. This is one reason why the axioms alone cannot justify determinate moral propositions. So if the core-apriorist is right, he must give us an independent way of working out *a priori* what these goods are. In other words, he must give us a theory of the good. We will come back to this point in section 7. The second problem is that if the axioms are ultimately not *a priori*, this also creates a difficulty. Identifying anything as good is not enough; we have to also know how to treat goods, and this is what the axioms tell us. The core-apriorist cannot just ignore this point. If the axioms are *a posteriori*, then we have to use *a posteriori* methods to ascertain the truth of determinate moral propositions relating to the comparisons of different goods. We make such comparisons regularly in everyday life, from deciding whether to go to the cinema or to the theatre, to deciding whether to give more government funding to healthcare or to education. If we can only work out the answers by using *a posteriori* methods, then the SLS theory turns out to have major *a posteriori* commitments. We may have an *a priori* theory of the good, but that is not enough to tell us how to behave morally. We need the help of the self-evident axioms, some of which now turn out to be only knowable *a posteriori*.

Rule utilitarians can solve the problems with *Impartiality* by interpreting it to mean that the good of one individual is not more important than that of another unless there are special grounds for believing that under particular rules, more good is likely to be realised in general than otherwise. Generally, loan-sharking will not produce a greater amount of good, so any reasonable set of rules will not allow it. However, we must once more use the 'other resources' of a particular utilitarian theory to interpret *Impartiality*. We are thus landed back with the problem that we encountered with *Aim*. As it stands, *Impartiality*, like *Aim*, appears to contravene *Sufficient Determinacy*, as the terms in it such as 'importance' have very imprecise application conditions. We can use the 'other resources' of a normative moral theory to justify determinate core moral knowledge, but this has a consequence: if we add more detail to *Aim*

and *Impartiality* by using a normative moral theory, our choice of theory is affected by *a posteriori* evidence (as we saw in our discussion of *Aim*). So *Impartiality*, like *Aim*, either contravenes *Sufficient Determinacy* or is known via *a posteriori* evidence.

If *Aim* and *Impartiality* are insufficiently determinate, it also seems unlikely that they can be self-evident. For they do not seem to be 'clear and precise' (contravening criterion (1)), and we can reasonably disagree over how they should be applied and which normative theory to use (contravening criterion (4)). Nevertheless, it is important to emphasise that there is no necessary connection between self-evident axioms and *Sufficient Determinacy*. It is not necessary that a proposition comply with *Sufficient Determinacy* for it to be a self-evident axiom.

Now that we have seen the type of argument that the core-aposteriorist can make against *Aim* and *Impartiality*, we can apply it to *Treatment*. *Treatment* risks contravening *Sufficient Determinacy* because it is unclear what the 'nature of circumstances' of individuals should be taken to be, and what a reasonable ground for the difference of treatment of individuals is. For *Treatment*, giving the terms in the axiom more precise application conditions will involve deciding which normative theory to use. Normative theories may agree that there is a difference between a right act and a wrong act, but disagree strongly on what that difference is, and whether that difference is always relevant in other cases. If the normative theory we use answers those questions with *a posteriori* evidence, then the axiom itself seems to be knowable on *a posteriori* grounds.

5 – *Prudence* and timing differences

Prudence claims that individuals should aim at what is good for them, and that the distance in time between an agent and the occurrence of the good should not by itself affect the agent's judgement of the good. In this section, I look at its claim about timing differences. In the next, I will look at a serious problem for Sidgwick that *Prudence* generates: the dualism of practical reason.

There are cases in which time appears to have an effect, but for many of such cases this impression is wrong. For example, I might prefer to receive £100 now rather than £110 in ten years because I reasonably expect inflation over the ten years to wipe out the difference, but my

decision is not actually influenced by the time between the two outcomes. The inflation influences my judgement, and it just happens to take place over ten years. My judgement would be just as reasonable if I could get £110 tomorrow, but the effect of the inflation remained the same. Another case would be if I chose to take £100 now because I know that I would find it difficult psychologically to wait ten years for £110. It may be rational for me to in effect pay £10 to avoid the wait, but it would be rational because of my psychological state, not the time between my decision and my obtaining the £110.

Whether timing differences are never relevant, though, is a more difficult question. According to *Prudence*, a life with a lot of pain at the start and a lot of benefit at the end is just as valuable as a life with a lot of benefit at the start and a lot of pain at the end, just as long as overall the amount of good in each life is equal. Intuitively, this seems wrong. We pity lives that start well and go downhill more than lives that start badly and improve. Saying that the two types of life are equally worthwhile seems to miss the point. Whilst he does not commit himself on the matter, John Broome suggests that we may attribute extra value to lives with good endings, and to the expectation that lives should grow better as they continue. If this does not happen, we get depressed, which is a negative cost.[37] If this is right, then the relevant difference is not actually the timing. It is the psychological effect.

Even if Broome's response works, there is still a more general objection to *Prudence*, which claims that temporal patterns matter in some way. If two agents have exactly the same amount of good in their lives, but that good is distributed throughout their lives differently, one of the lives may be rationally preferable to the other. The preferable life may even have a slightly lesser amount of good overall. *Prudence* denies that this is possible. A greater good is preferable to a lesser good no matter when they may occur in one's life. If White can get £100 now by suffering a very slight electric shock, or in a week just by waiting, *Prudence* would tell him to wait. The electric shock would add pain to his life, and waiting would not. Our intuitions may be different, though, even if we accept that White's life will contain slightly more pain if he is shocked. We also feel that it is better for pain to be in the past than in the future, but if timing differences are irrelevant, it should not matter whether I underwent major pain a year ago or whether I will undergo the same amount of pain tomorrow. Either event will cause the same

---

[37] Broome (2004), pp. 227-228.

amount of pain in my life.[38] The relief that I feel that a pain is in my past rather than in my future or my present is irrational.

De Lazari-Radek and Singer believe that intuitions against *Prudence*'s rejection of timing differences have no rational justification, and that some of them may be the result of evolutionary development (The intuitions were helpful to us in the past, but are no longer so.). There may be a rational reason to prefer pain in your childhood to the same amount of pain in your adulthood, but this is not because of the timing difference.[39] It might seem odd for core-apriorists to appeal to evolutionary evidence here, but it is a legitimate appeal. De Lazari-Radek and Singer argue that timing differences are irrelevant independently of evolutionary evidence. They only afterwards use such evidence to explain why we have intuitions about timing differences in the first place.

Whether the intuitions are rational or not, though, if *Prudence* has such counterintuitive consequences, it is unlikely to be self-evident. Given the amount of disagreement that it can generate, it may contravene Sidgwick's criterion (4). Unfortunately, this is not going to help the core-aposteriorist, because self-evidence is neither necessary nor sufficient for a proposition to be knowable *a priori*. It is irrelevant. Even if *Prudence* is not self-evident, it can still be true and *a priori*, which means that the core-aposteriorist cannot use it to argue against the CAT. She will have to rely instead on her examination of *Aim*, *Impartiality* and *Treatment* to do so – unless there are other criticisms of *Prudence* that she can use. We come now to the dualism of practical reason, a problem that Sidgwick famously confessed himself unable to solve.

6 – The dualism of practical reason

In this section, I first outline the problem. I then examine the SLS theory's solution to it, and show that the solution should deeply worry the core-apriorist.

Sidgwick wants to use *Prudence* as part of his moral theory, and part of the justification of his utilitarianism. However, he notes that *Prudence* could also be used in favour of rational egoism, which conflicts with the other axioms. In particular, rational egoism conflicts with *Benevolence*, because if I am obliged to favour my own interests, I cannot be obliged to treat

---

[38] Derek Parfit provides examples to test these intuitions, though he does not come down entirely on either side of the debate ((1984), pp. 177-178).
[39] De Lazari-Radek and Singer (2014), pp. 126-133.

the interests of others as equal to mine.[40] An egoist can use *Prudence* to rationally claim that he should only pursue his own self-interest. So long as he does not claim that everyone else must pursue the egoist's self-interest, he does not violate the principle that nobody's good is more important than anyone else's 'from the point of view of the Universe'. If we look at *Prudence* in isolation, leaving aside the other self-evident axioms for the moment, there is nothing incoherent in thinking that everyone should pursue their own self-interest, even if one agent's self-interest conflicts with another's.[41] So do we stick with *Benevolence*, or give it up to become rational egoists?

Sidgwick does not claim that the problem is insoluble, only that he cannot solve it. He suggests two possible solutions. The first is to weaken our epistemic standards. If we can take certain propositions about the natural (non-moral) world to be true simply because we have 'a strong disposition' to believe them, and because our beliefs would be incoherent without them, it will be difficult to reject a similar assumption in ethics.[42] Accepting *Benevolence* over *Prudence* will avoid incoherent moral beliefs. People usually believe strongly that rational egoism is pernicious, so it is better that we choose *Benevolence* over *Prudence*. We should be utilitarians rather than rational egoists.

Sidgwick's second solution is to go theistic. There may be a Divine Lawgiver who wants 'general happiness' amongst humans, and who will reward the just and punish the disobedient.[43] Sidgwick allows that the existence of God may be ascertained through reason, revelation, or both. Reason would demonstrate the existence of God through reasoned argument, such as the argument from design. The argument claims that the universe looks designed, designed things have a designer, and so this designer must be God. Revelation involves the experience of an event that demonstrates the existence of God, such as the witnessing of a miracle. Both methods have considerable drawbacks. Every argument for the existence of God that has been proposed is, at best, mired in controversy. Revelation only works if we can show that miracles are more likely to be genuine than not. Sceptics from David Hume onwards point out the difficulty (or even impossibility) of doing so[44].

---

[40] Sidgwick (1981), pp. 381-382.
[41] Sidgwick (1981), pp. 420, 497-498.
[42] Sidgwick (1981), p. 509.
[43] Sidgwick (1981), p. 504. Sidgwick does not mention it here, but he was deeply interested in paranormal research. If he had found evidence of a benevolent deity, it might have supported this particular solution.
[44] Hume (1999), section X, 'Of Miracles'.

Sidgwick does not develop either solution, and so leaves the dualism of practical reason unsolved. Even if they work, they may not be *a priori*. If his suggestion of lowering epistemic standards is correct, then we can ascertain if utilitarianism is true by seeing what our strong dispositions are on the subject. However, even if checking our dispositions is an *a priori* task (rather than an *a posteriori* task), it cannot be the whole story. Our dispositions may be hopelessly contrary to those of most other people, so we have to examine other people's dispositions, and this is an *a posteriori* investigation.

As for the theological solution, even if we think that miracles are not natural events (perhaps they are supernatural), it remains the case that miracles must be experienced by agents if they are to demonstrate to us the existence of God. If we see someone miraculously walking on water, we *see* them. The witnessing of a miracle is an empirical occurrence; there have been experiments conducted on the effectiveness of prayer to aid various illnesses, for example. Arguments for the existence of God do not fare much better, as they frequently rely on *a posteriori* premises. The argument from design collapses if it can be shown that many things that look designed do not have a designer.

It might be that there is no solution to the dualism of practical reason, in which case Sidgwick is in real trouble. Some philosophers have been even more pessimistic, arguing that the problem can be extended to show that no moral norms can be rationally required.[45] So the core-apriorist is left in an uncomfortable position. If he takes Sidgwick's solutions seriously, he has to show that they work without appealing to *a posteriori* evidence. Alternatively, he could reject Sidgwick's solutions, in which case he has to come up with another solution. De Lazari-Radek and Singer claim to have one, but we will now see that it is not *a priori* either, and hence it is of no help to the core-apriorist.

Their strategy is to use Sharon Street's evolutionary debunking argument ('EDA') to cast doubt on rational egoism, whilst simultaneously shielding *Benevolence* from the same criticism. This defence seems dangerous. If the EDA debunks rational egoism, we need a good reason to think that it does not debunk *Benevolence* as well, and hence that it does not debunk the SLS theory. Street's EDA claims that moral knowledge, if not impossible, is massively unlikely, and

---

[45] Hills (2003).

we can only reach it through sheer luck if we can reach it at all.[46] Since the SLS theory claims to give us a reliable way to reach moral knowledge, surely they cannot both be correct. De Lazari-Radek and Singer thus walk a tightrope. They have to show that the EDA only works in particular cases; it is neither a complete failure nor an unqualified success. We will see shortly that they cannot do this *a priori*. Even if *Benevolence* can escape the EDA, we will only be able to know that *Benevolence* is true *a posteriori*.

Street's EDA objects to value realists who believe that there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes.[47] It goes as follows:

1) The forces of natural selection have had a tremendous influence on the content of the value judgements that humans make. Evolutionary pressures shaped these value judgements so that the judgements helped humans to adapt to their environment.

2) The value realist must tell us what the relationship is between the selective pressures on our value judgements and the independent evaluative truths that he thinks that we can recognise. There are two options:

3) The first option is that there is no relationship. Evolution pushed human value judgements in a certain direction so that humans could successfully adapt to their environment, and it was just coincidental that the judgements were correct. This would be incredible.

4) The second option is that it was to our evolutionary advantage that our value judgements track independent evaluative truths. But this is at least highly unlikely, because our ability to survive in a certain environment is not improved by having true value judgements. We would have had the same success in surviving even if all our value judgements were false. So there is no evolutionary pressure on us to have correct value judgements. This forces us back to the first option, which is massively implausible.

5) So value realism is untenable.

---

[46] We should note that other types of evolutionary debunking arguments also exist. They can have stronger or weaker conclusions; some claim that there are no moral truths, for example (see Joyce (2016)). However, I will only consider Street's specific EDA.
[47] Street (2006), p. 110. Value realists include moral realists who believe that there are moral facts or truths that hold independently of our moral attitudes. Street notes that this does not include all moral realists, such as Peter Railton (Street (2006), pp. 136-137; Railton (1986)).

Street's EDA presents value realists with a dilemma. Either value judgements track evaluative truths, or they do not. If they do not, then they can only be correct by chance. So suppose that they do. Many of our non-value judgements can track truths because we evolved the ability to make such truth-tracking judgements. The truth of the judgements helps us to survive. However, our chances of survival remain the same whether our value judgements are true or false. So there is no evolutionary reason for our value judgements to track evaluative truths. It is thus mysterious how we could have the ability to do so. Value realists have of course argued that Street's EDA is wrong, but I will not take this line of argument.[48] I will instead ask the following question: assuming that the EDA works against rational egoism, why do de Lazari-Radek and Singer think that it does not work against *Benevolence*? My intention is to show that if their solution to the dualism of practical reason works, the solution can only be known *a posteriori*. The solution will therefore not help the core-apriorist.

De Lazari-Radek and Singer's specific claim is that *Benevolence* survives the EDA, but the self-interest principle ('SI') does not. *Benevolence* implies that if altruism increases the good overall, moral agents should be altruistic towards others, even when there is no chance of having any benefits repaid in any way.[49] For example, if you benefit a stranger because you hope the stranger will benefit you later, this is not altruistic. De Lazari-Radek and Singer make it clear that they do not use the term 'altruism' in the sense that biologists use it. In biology, an act is altruistic simply because the actor gains no immediate benefit from the act it performs. The ability to perform biologically altruistic acts evolved because such altruism ultimately benefits the survival of the actor, the actor's genes or other members in the actor's group.[50] The actor's motives (even if it has any) are irrelevant. De Lazari-Radek and Singer's 'altruism' excludes such acts.

SI claims that 'any agent has a reason to do what makes her life go better, the strength of the reason varying in proposition to the extent to which her well-being will be improved'.[51] This is not the same as Sidgwick's rational egoism, since the mere fact that I have a reason to do

[48] E.g. Enoch (2011), pp. 163-175; Clarke-Doane (2015).
[49] De Lazari-Radek and Singer (2014), pp. 185-187.
[50] Dawkins (2006) describes how biologically altruistic acts can promote the survival of the actor's genes. Sober and Wilson argue that biologically altruistic acts can promote the survival of the actor's group ((1999), p. 9).
[51] De Lazari-Radek and Singer (2014), p. 190. The idea comes from Roger Crisp ((2006a), p. 73), but I use de Lazari-Radek and Singer's formulation of the principle.

what improves my life does not mean that I cannot have stronger reasons to do things that make my life go worse. Reasons can conflict, and some reasons can be stronger than others. Still, rational egoism implies SI, so if SI is false, then rational egoism is false.

De Lazari-Radek and Singer claim that Street's EDA debunks SI, but not *Benevolence*. Since utilitarianism is justified by *Benevolence*, and rational egoism implies SI, we therefore have good reason to believe that utilitarianism is correct, and no reason to believe that rational egoism is. This solves the dualism of practical reason.

Street's EDA debunks SI, claim de Lazari-Radek and Singer, because SI clearly has evolutionary advantages. It is not difficult to see why organisms would perform acts that benefited themselves – avoiding pain, obtaining nutrition, finding shelter from hostile environments, and so on. Successful organisms survive, and survival usually increases the chances of reproduction. Evolutionary development has therefore 'programmed' us to have the intuition that SI is true, because it results in reproductive success. SI is a value judgement, and Street argues that if evolutionary development is the cause of our value judgements, this is not because such judgements track evaluative truths. The fact that evolutionary development programmed us to believe that SI is true has nothing to do with SI's actual truth value. So there is no reason to think that SI is true.

If Street's EDA debunks SI, why believe that it is ineffective against *Benevolence*? De Lazari-Radek and Singer think that we have the following reasons to believe *Benevolence*:

a) It is the result of careful reflection.
b) It is accepted by many different cultures and religions that have carefully reflected on it, from Christianity to Hinduism.
c) There is no plausible evolutionary reason why we believe *Benevolence*. It does not appear evolutionarily advantageous for us to do so.[52]

These criteria generalise to cover 'any process of establishing that an intuition has the highest possible degree of certainty':

---

[52] De Lazari-Radek and Singer (2014), p. 193. See also p. 195.

1) Careful reflection leading to a conviction of self-evidence;

2) Independent agreement of other careful thinkers; and

3) The absence of a plausible explanation of the intuition as a non-truth-tracking psychological process.[53]

De Lazari-Radek and Singer claim that *Benevolence* passes all these tests. They also appeal to other *a priori* beliefs we have that track truths, where the truth of the belief confers no evolutionary advantage. They say that '[l]ike our ability to do higher mathematics, [our belief in *Benevolence*] can be most plausibly explained as the outcome of our capacity to reason'.[54]

Street's EDA is important for the above criteria because of (3). There is a plausible evolutionary story about why we are inclined to endorse SI. Even if we endorsed SI after careful reflection, and even if most people would endorse SI after careful reflection, SI would still be less reliable than *Benevolence*, as *Benevolence* fulfils every criterion for an intuition to be reliable, and SI only fulfils the first two at most.

De Lazari-Radek and Singer's argument can be criticised by the core-aposteriorist on the grounds that even if their argument works, it is an *a posteriori* argument. We therefore use *a posteriori* methods to ascertain that *Benevolence* is more likely to be true than SI, and so the reason to prefer utilitarianism to rational egoism is also *a posteriori*.

Look again at (3). De Lazari-Radek and Singer claim that the ability to make moral judgements is similar to our ability to make higher mathematical judgements. Their point is that the ability to do higher mathematics is not a direct result of evolutionary development. Few, if any lives have been made more reproductively successful by knowing the truth of the Taniyama-Shimura Conjecture. There is no reason for evolution to directly give us the ability to understand the Conjecture or to prove it.

Since we do have the ability, though, there must be a reason why. There is a simple rough answer, though naturally it takes much biological work to provide the details. Whatever the survival value of knowing the Conjecture is, the ability to reason, even minimally, has a huge survival value. It helps organisms predict what predators will do, where best to find food, what to use for shelter, and so on. Basic mathematics may also be useful. However, once we have got

---

[53] De Lazari-Radek and Singer (2014), p. 195.
[54] De Lazari-Radek and Singer (2014), p. 193. I assume here that mathematical beliefs can be *a priori*.

the ability to reason simply and to make basic mathematical calculations, a whole new world opens up, and we are that much closer to using our abilities in ever more complex ways that have no direct survival value at all. The same type of process occurs with other human traits that do not involve reasoning. There is no direct evolutionary advantage in finding a three-star Michelin restaurant's food more pleasurable than the food at a Burger King. Both will ensure that you do not die of starvation, so why do we perceive such a large difference in culinary value? All that is evolutionarily needed is that nutritious substances taste good, and non-nutritious substances taste unappealing or bad. But evolutionary development is not so limited.

Evolution does not plan; evolutionary development is the result of a reaction to circumstances, and that reaction may have unexpected consequences later on, for better or worse. The EDA defender (who wants to argue that Street's EDA works for all value judgements, including *Benevolence*) may argue that *Benevolence* is similar. Our belief in *Benevolence* does not directly promote survival, but it is the consequence of a set of abilities that did promote survival, just as our ability to solve differential equations has been built from abilities that promote survival. The EDA defender claims further that our belief in *Benevolence* is the consequence of other non-truth-tracking beliefs that we have thanks to evolutionary development, so our belief in *Benevolence* is itself non-truth-tracking.

For de Lazari-Radek and Singer to refute this claim, and argue that criterion (3) supports *Benevolence*, they first have to confirm criterion (1). We can only divorce our belief in *Benevolence* from evolutionary influence once we have made a careful examination of evolutionary influence. And unfortunately for the core-apriorist, this examination must be made *a posteriori*. First, we must ascertain how evolutionary development works. How can we reasonably expect evolution to shape our beliefs? Does it shape different types of belief in different ways? Second, we must discover the *a posteriori* consequences of adopting *Benevolence*. Are the results beneficial from an evolutionary point of view, and if they are, in which environments? If de Lazari-Radek and Singer are correct, then our investigation will show that evolutionary development has not influenced beliefs in *Benevolence*, but it has influenced beliefs in SI, so it is more likely that we are justified in believing *Benevolence*. The trouble now is that the adoption of *Benevolence* becomes dependent on *a posteriori* investigation.

De Lazari-Radek and Singer can reply that although evolution gave us the ability to believe *Benevolence*, this does not imply that there was specific direct evolutionary pressure to believe it or not. Our reasoning pressures us to believe it, not our evolution. But even if de Lazari-Radek and Singer are correct about evolutionary pressure, this ignores the fact that there are other non-truth-tracking psychological processes that may also pressure us to believe *Benevolence*. We may believe that it is true because of the effect of holding the belief, rather than because of any rational reasons. For example, one might think that the world is filled with enough selfishness, spite and indifference to suffering. Being told that it is every man for himself would be depressing, especially if you are always the loser. A belief in *Benevolence* may be comforting. If this view sounds overly cynical, we can make it less bleak. The world is full of risks and dangers. For whatever reason, which may have nothing to do with reasoning, we have powerful inclinations to help the lives of others, just as we may have other powerful inclinations to do things that do not directly benefit us from an evolutionary point of view. It might not actually be that much of a leap to go from caring about our families (which is evolutionarily advantageous) to caring about our friends and then, for whatever reason, starting to care about strangers.

The wider problem here is that de Lazari-Radek and Singer only look at possible evolutionary explanations of *Benevolence*. (3), however, is not limited to evolutionary explanations. It claims that an intuition can only be reliable 'in the absence of a plausible explanation of the intuition as a non-truth-tracking psychological process'. This does not only include evolutionary processes. We can run *a posteriori* investigations on the effect of social influence on our acceptance of *Benevolence*, political influence, religious influence, and so on. A defence of *Benevolence* therefore may require a number of *a posteriori* tests to help us ascertain its reliability. Maybe we find *Benevolence* plausible not because of evolutionary influence, but because of social influence. This is similar to David Hume's claim that the reason why we developed the virtue of justice was that it helped to stabilise large societies. A growing society will eventually become so large that it is impossible for a single citizen to know everyone. Some citizens will be strangers to each other, and we need justice to ensure that they

will treat each other fairly, because they have no other ties of family or friendship to ensure that they will do so.[55]

This gives rise to a whole range of social/political/religious debunking arguments, which have a similar structure to Street's EDA. In effect, they claim that we hold our belief in *Benevolence* or SI because the beliefs are useful, not because they are true, and so we are not justified in believing that they are true. Furthermore, if discovering an evolutionary explanation *a posteriori* for our belief in *Benevolence* would damage de Lazari-Radek and Singer's argument, so would the *a posteriori* discovery of a social/political/religious explanation. In all these cases, *Benevolence* would fail to meet criterion (3).

(2) also relies somewhat on (1), as de Lazari-Radek and Singer claim that reliable intuitions require 'independent agreement of other careful thinkers'. This requirement is subject to how carefully we have reflected, how many others disagree, what their reasons for disagreement are, and so on. If we have good reason to be very confident about our reflection, then even large amounts of disagreement may not be enough to force us to re-evaluative our position.

Why not just use (1), then? Why do we need to rely on (2) and (3) as well? It is important here not to confuse two issues. The first is to ascertain the truth of an intuition. To do so, it may be that all we need is careful reflection. This is not what de Lazari-Radek and Singer are concerned with here, though. They are looking at whether we are justified in holding the intuition, and so what we consider to be careful reflection may not be enough (The rational egoist can insist that he has carefully reflected as well.). This means that even though (2) and (3) rely on (1), (1) also relies on (2) and (3). If many people who have carefully reflected disagree with us, maybe our own reflection was not that careful. If our intuition has a non-truth-tracking explanation, maybe we should hesitate to think it correct. Our reflection may be biased. (1) is still the most important criterion, however. If we take it away, (2) becomes a simple appeal to numbers, and (3) implies that if a belief does not have a non-truth-tracking psychological explanation, it is more likely to be correct. This is not quite right, as many beliefs may have non-truth-tracking explanations that are not psychological.

Simply dropping criterion (3), though, is not an option. If we do, we get the result that we should believe *Benevolence* rather than SI because we and other careful thinkers agree that

---

[55] Hume (2000), section 3.2.2.

Benevolence is more plausible. Suddenly the EDA is doing no work at all. If de Lazari-Radek and Singer could show that *a priori* reasoning alone is enough to do the job of showing that *Benevolence* is preferable to SI, they would not need to bring in the EDA. In fact, dropping (3) would entail leaving Sidgwick's problem unsolved. The entire point of his problem was that rational reflection alone does not appear to rule out rational egoism.

If we are very uncertain whether *Benevolence* meets criterion (3), then, this gives us reason to believe that we cannot isolate *Benevolence* from the EDA. If the EDA refutes our grounds for believing SI, we need to make sure that it does not refute our grounds for believing *Benevolence* as well. However, it turns out that this investigation of *Benevolence* is *a posteriori*. Furthermore, even if de Lazari-Radek and Singer insist that evolutionary development had no influence on our belief in *Benevolence*, it is difficult to dismiss all non-truth-tracking psychological explanations for the belief unless we make more investigations, which will also be *a posteriori*.

The prospects of solving the dualism of practical reason thus look bleak for the core-apriorist, as the SLS theory's solution relies on *a posteriori* information. Sidgwick's suggestion that we weaken our epistemic standards appears to depend on *a posteriori* observations of actual agents. His suggestion that we turn to God will be unattractive to anyone who thinks that moral truths and moral knowledge have no connections to a deity. Many purported ways to argue for the existence of a deity also depend themselves on *a posteriori* premises or observations.

If the SLS theory is right to use Street's EDA, we have the following method to justify the normative content of determinate core moral propositions. We start with Sidgwick's axioms. The axioms appear to justify either utilitarianism or rational egoism. Careful reflection alone leaves us uncertain which is correct. So we look around for widespread agreement on the axioms, and we examine whether SI or *Benevolence* is vulnerable to explanations of them that do not imply that they track any truth (which is an *a posteriori* investigation). *Benevolence* turns out to be better justified than SI according to the three criteria that de Lazari-Radek and Singer list, so we should choose utilitarianism over rational egoism, and we can include *Benevolence* as a determinate core moral proposition. We can also use the 'other resources' of utilitarianism combined with the axioms to justify other determinate moral propositions.

However, we only know that *Benevolence* is justified because of *a posteriori* investigation, so our acceptance of utilitarianism is based to a great extent on *a posteriori* investigation. The SLS theory is therefore not a core-apriorist theory.

There is one more part of the SLS theory that we have not yet looked at. Utilitarianism requires us to maximise the good, but we have not examined what the good is. Sidgwick identifies it as happiness, or pleasure, and so he endorses a hedonist theory of the good. De Lazari-Radek and Singer agree with him. Could the core-apriorist argue that we must identify the good *a priori*? In the next section, I will outline de Lazari-Radek and Singer's defence of hedonism, and show that their arguments are *a posteriori*. The core-apriorist will get no comfort from them.

7 – Theories of the good

It would take too much space to fully discuss theories of the good, so I will begin with only briefly outlining why de Lazari-Radek and Singer reject non-hedonist theories of the good. I will then move on to their defence of hedonism.

They identify three different types of theory of the good. The first is desire-based theories, according to which it is good for agents to have their desires satisfied. The problem with such theories, claim de Lazari-Radek and Singer, is that agents can have desires that later they will regret satisfying, and that if we try to restrict the types of desire that constitute the good, we cannot get a clear, definite set of desires that exclude all the problematic ones. For example, if an agent prefers A to B, B to C, and C to A, desire-based theories have no easy way to work out which desire for A, B or C should be satisfied.[56] Furthermore, if the satisfaction of an agent's actual desires is good, this might entail that the satisfaction of any crazy desires that the agent has is good. If we avoid this by saying that the satisfaction of the desires that the agent would have if he were 'rational and aware' is good, then what is good for the actual agent might be the satisfaction of desires that he never has. He may never actually be rational and aware.[57]

The second type of theory claims that there need be no connection between the good and anything that the agent would find attractive. The most prominent of these theories is

---

[56] De Lazari-Radek and Singer (2014), pp. 219-220.
[57] De Lazari-Radek and Singer (2014), pp. 225-226.

perfectionism, which holds that the good is the perfection of something. Typically, this is taken to be some human quality, such as wisdom, courage or virtue. So if the good is the perfection of virtue, it is good for an agent to be virtuous, even if he never finds virtue attractive or desirable. Sidgwick explicitly claims that perfectionism has fatal difficulties. For example, if the perfection of courage is good, then we need to know when courage becomes foolhardiness, as foolhardiness is not perfect. But we can only do this if we know when courage is no longer good. Since courage-perfectionism holds that courage is identical to the good, it denies that this could ever happen. Sidgwick believes that most qualities will lead to a similar problem, but he identifies a few that do not: wisdom, universal benevolence and justice. The trouble here is that we can only define these qualities by already presupposing the good. For example, if the definition of 'justice' implies the impartial distribution of what is good according to the right rules, then justice itself cannot be the good.[58] The right and the good are conceptually prior to justice. Furthermore, there is a risk that the reason why we should perfect something is become of the consequences that it will produce. This means that the theory is one of *instrumental* good, rather than intrinsic good.

Sidgwick, de Lazari-Radek and Singer choose instead to endorse a hedonist theory, which promotes pleasure as good and pain as bad. Sidgwick states that happiness is the ultimate good, where happiness is the attainment of pleasure and the avoidance of pain. The greatest possible happiness is 'the greatest attainable surplus of pleasure over pain'.[59] Some philosophers, such as the utilitarian John Stuart Mill, believe that there are different types of pleasure, and some are more valuable than others. Enjoying a performance of *Richard III* is a more valuable pleasure than enjoying a drug-induced stupor, even if the pleasure given by the drugs is more intense. As Mill puts it, '[i]t is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied'.[60] Sidgwick disagrees, as we can only make distinctions by appealing to something other than pleasure. So if Mill were right, good could not be identified with pleasure.[61]

Sidgwick gives the following definition of pleasure:

---

[58] Sidgwick (1981), pp. 392-393; de Lazari-Radek and Singer (2014), pp. 245-246.
[59] Sidgwick (1981), pp. 120-121.
[60] Mill (1987), p. 281.
[61] Sidgwick (1981), pp. 94-95.

'[a] feeling which the sentient individual at the time of feeling it implicitly or explicitly apprehends to be desirable; - desirable, that is, when considered merely as feeling, and not in respect of its objective conditions or consequences, or any facts that come directly within the cognisance and judgment of others besides the sentient individual.'[62]

De Lazari-Radek and Singer endorse this definition, and spend some time defending it. Curiously, much of their defence appeals to psychological research. For example, they cite psychological experiments to demonstrate that the satisfaction of a desire need not produce any pleasure.[63] They think that the experiments help to show that hedonism is separate from desire-based theories of the good, and so is not vulnerable to the problems that desire-based theories face.

One prominent objection to hedonism is Robert Nozick's 'experience machine' thought experiment.[64] Nozick imagines a scenario where scientists are able to plug your brain into a machine that would make you think and feel that you are having pleasurable experiences, such as successfully pursuing any goals you have. You would think that it was real life, as you would have forgotten that you were plugged into the machine. If you were made the offer to plug in, would you accept? Nozick thinks not, for three reasons. First, we want to actually do things rather than just think that we have done them. Second, someone hooked up to the machine is just 'an indeterminate blob'.[65] In a sense, she loses her identity, as she loses all the characteristics and traits that make her 'courageous, kind, intelligent, witty, [and] loving' when she is plugged in. Third, living in a 'man-made' reality is a poor substitute for the real thing. Nozick concludes that what is important to people is not merely their perceived experiences alone. This poses a problem for hedonism, because if the experience machine can give you more pleasure than you would get in your real life, hedonism would advise you to plug in. But most people, claims Nozick, would recoil from doing so.

De Lazari-Radek and Singer aim to show not that Nozick's intuition is wrong, but that we should not rely on it. Their aim is to show that the intuition cannot be used as an argument against hedonism. They point to two main flaws that the intuition apparently suffers from. The

---

[62] Sidgwick (1981), p. 131.
[63] De Lazari-Radek and Singer (2014), pp. 245-246; Olds and Milner (1954).
[64] Nozick (1974), pp. 42-45.
[65] Nozick (1974), p. 43.

first is that as Nozick phrases the thought experiment, it may involve a status quo bias.[66] Status quo biases are biases towards what we already have or states that we are already in. We are reluctant to change, even when we would benefit from doing so. So if we were already using the experience machine, and were asked if we would like to return to the real world, our reaction may not be what Nozick would expect. De Lazari-Radek and Singer cite an experiment conducted by Felipe de Brigard that tries to eliminate the bias. Test subjects were told that they were already plugged into the experience machine, and were asked if they would like to leave. Each subject was given one and only one of three vignettes that could influence their choice. In the neutral vignette, subjects were not told anything about real life outside the machine. In the negative vignette, they were told that in 'real life' they were an inmate at a maximum security prison. In the positive vignette, they were told that in 'real life' they were a multi-millionaire artist living in Monaco. 50% of the subjects given the positive vignette chose to remain plugged into the machine – apparently, being a multi-millionaire in real life was not tempting enough to leave.[67]

It should be said that considerably more work has to be done to confirm de Brigard's results. Only 46% of test subjects who got the neutral vignette chose to stay connected to the machine, so apparently people are more likely to leave the machine if they are told nothing about their real life then if they know that they are a multi-millionaire in real life (Incidentally, 84% of subjects with the negative vignette chose to stay connected.). This could be the effect of small sample sizes, as each vignette was only given to 24 subjects. The subjects were all non-philosophy undergraduates at an American university, which could also have skewed the results. It was not tested whether the test subjects' choices would have been different before they were plugged in, rather than after. We can also question the framing of the vignettes. A person plugged into the experience machine does not interact with anything outside the machine, but then prisoners do not usually interact much with anything outside prison. In such circumstances, why not take the option that makes you think that you can do what you like in freedom? As for the neutral vignette, where subjects were given no information about their life outside the machine, might they not have assumed that life outside the machine would be

---

[66] De Lazari-Radek and Singer (2014), pp. 256-258.
[67] De Brigard (2010).

exactly like the life they actually had? Does the actual life of the test subjects skew the results?[68]

De Brigard's experiment is interesting, but it is not clear how far it refutes Nozick's argument. De Brigard holds that it falsifies the following hypothesis:

'people care more about reality than they care about how their life is experienced from the inside… [so] regardless of how reality turns out to be, they would choose it over a simulated life.'[69]

Regardless of how reality turns out to be? If life were uniformly terrible for most people, then Nozick's original thought experiment would not be very convincing in the first place. Plugging in would be a way to escape misery. But as the quality of life increases, then even given the results of de Brigard's experiment, fewer people would want to plug in. More subjects chose to leave the machine in the neutral and positive vignettes than in the negative vignette. Maybe there is a limit above which an extra increase in the quality of life would not influence more people to leave the machine, but de Brigard is testing a far stronger claim than the one that Nozick is arguing for. Furthermore, as de Brigard mentions, his results appear to cause problems for hedonism as well.[70] If our intuitions are to pursue pleasure and avoid pain, or at least avoid the reduction of pleasure, then the percentage of test subjects who would unplug from the machine in the neutral vignette should be lower than those who would in the positive vignette. In fact, it is slightly higher.

An important consideration that de Lazari-Radek and Singer do not mention, and that might damage both Nozick's argument and de Brigard's experiment, is that your personal goals may have an effect on whether you agree to be connected to the experience machine. If you want to write a symphony, then arguably you could do so both in real life and if you were connected to the machine. But there are many goals that you cannot accomplish inside the machine. You cannot be loved by a real person. You cannot genuinely help others. You cannot make previously unknown astrophysical discoveries (Why think that the experience machine will give

---

[68] De Brigard does test whether subjects given the neutral vignette would unplug from the machine if they know that their 'real life' was very different in some undescribed way from their life in the machine, but this still allows the test subjects to think (or not to think) that their 'real life' outside the machine was their actual life.
[69] De Brigard (2010), p. 48.
[70] De Brigard (2010), p. 53.

accurate information about previously unknown astrophysics?). You cannot see the genuine Mona Lisa. When I donate to charity, I want my money to actually help people, and that could not happen in the experience machine. So if our reactions to Nozick's though experiment depend on what goals we actually have, then de Brigard's experiment can show us very little, as it does not take the goals of the test subjects into account. So many people have so many different types of goals that relying on intuitions about those goals to help identify the good is not helpful, even if we only use those intuitions as a first step towards a theory of the good. It is thus very unclear how far the hedonist can rely on de Brigard's experiment. He cannot use the results to argue that we care more about what we experience than about our connection to the real world.

A final point about de Brigard's experiment is that it is not even clear that he avoids status quo bias. In his set-up of his thought experiment, the test subjects are already plugged into the experience machine. Therefore the status quo bias would predict that they would prefer to remain plugged in. The amount of pleasure they get from the machine may therefore not be the deciding factor in their decision to stay.

The upshot of de Brigard's experiment is that it has too many problems to support hedonism. It might, however, be enough to show that Nozick is too quick to use the experience machine to show that hedonism is wrong. Since de Lazari-Radek and Singer are content to have this weaker conclusion, the problems I have raised may not worry them. But we do need more compelling reasons to think that their own position on hedonism is any better than Nozick's.

The second flaw that de Lazari-Radek and Singer identify with Nozick's argument is that we may be very nervous about agreeing to plug into the experience machine. Why are the scientists operating the machine so keen to make us the offer? What if the machine breaks? What if there is a power failure? Would you really trust the people operating the machine?[71] De Lazari-Radek and Singer do not mention any experiments that ask subjects about experience machine scenarios that deal with these matters, but presumably such experiments could be run. If it turned out that people would be more likely to plug in if they had sufficient reassurance, de Lazari-Radek and Singer could use the results in their favour.

---

[71] De Lazari-Radek and Singer (2014), pp. 258-259.

The trouble with the flaws that de Lazari-Radek and Singer identify is that they are identified *a posteriori*. If test subjects typically rejected the experience machine in de Brigard's experiment in all the vignettes, then it could not be used to claim that Nozick's intuition about the machine was wrong. It should be said that de Lazari-Radek and Singer do not merely point at people's actual intuitions to refute Nozick's point. They do not try to find out what is good just by checking what most people think.[72] But their methodology here is very close to the methodology they use to solve the dualism of practical reason. That is, we can check to see whether Nozick's intuition can be plausibly explained as a non-truth-tracking psychological process. It will not track the truth if it can be explained as a mere preference for the status quo, or if it has been massively influenced by fears about the experience machine's reliability, or the motives of the people running the machine. De Lazari-Radek and Singer even mention as a possibility Roger Crisp's suggestion that our intuitions about the experience machine are vulnerable to an evolutionary debunking argument.[73] This means that their defence of hedonism against Nozick suffers from the same doubts as their solution to the dualism of practical reason. It is not *a priori* because they are using *a posteriori* psychological research to defend hedonism. If they can find separate *a priori* means to defend hedonism as well, this is not a problem for the core-apriorist, but if they do not, it entails that their theory of the good is not *a priori*. Unfortunately, we will not see that they do not.

Could we not use evolutionary debunking arguments against hedonism? De Lazari-Radek and Singer think not. The EDA against hedonism would be something like this: our belief that pain is bad and pleasure is good has an evolutionary explanation. By avoiding pain and engaging in pleasurable activity, an organism typically has a greater chance of survival and reproductive success. So it makes evolutionary sense for people to believe that pleasure is good and pain is bad. But these beliefs are evolutionarily useful whether they are justified or not. A moral belief is not justified simply because it is useful. There is thus no reason to think that such beliefs are justified.

De Lazari-Radek and Singer object that we know that pain is bad because we can directly experience it.[74] We cannot be mistaken that violent headaches are bad, no matter what the cause of the headache is. The headache might have good consequences, but it cannot be

---

[72] De Lazari-Radek and Singer (2014), p. 261.
[73] De Lazari-Radek and Singer (2014), p. 260; Crisp (2006b), pp. 638-639.
[74] De Lazari-Radek and Singer (2014), pp. 266-269.

intrinsically good. De Lazari-Radek and Singer create a thought experiment to make their point. Thalia Wheatley and Jonathan Haidt ran an experiment in which they hypnotised test subjects to feel disgust when they read the word 'often'. The subjects then read a scenario about Dan, who runs student discussions about academic topics at a university. Dan 'often' chooses topics which participants would find attractive. The test subjects who had been hypnotised were more likely to think that Dan was doing something wrong.[75] Presumably, de Lazari-Radek and Singer claim, once the subjects found out the full details of the experiment, they would change their minds. They would know that the only reason they felt that Dan had done something wrong was that they had been hypnotised.

Suppose, though, that they had been hypnotised to have a headache when they read the word 'often'. When they read the scenario, they would feel pain, and they would think that they had a bad experience. If they were then told that the reason for the headache was because of the hypnotic suggestion, would they decide that they were wrong to judge the experience as bad? No, because it does not matter how the pain was caused. The fact that they were in pain would be enough.

It is not clear just what this example is intended to show. If de Lazari-Radek and Singer are trying to argue from analogy, then their point is that the test subjects would admit that their judgements about Dan were wrong, but not that their judgements about their headaches were wrong. However, this analogy breaks down rapidly. Wheatley and Haidt's experiment and the thought experiments are not testing for the same type of thing. The thought experiment concerns an experience; the test subjects are asked about their headaches. Wheatley and Haidt's test subjects are asked about a scenario, though, and not their experience of disgust. It would be extremely plausible for their test subjects to admit after the experiment that whilst they now knew that Dan had done nothing wrong, they still really did feel disgust. It does not matter how the disgust was caused, any more than it matters how the headache was caused – the disgust was real.

What if de Lazari-Radek and Singer are not using Wheatley and Haidt's experiment to make an analogy? If so, they can discard any mention of the experiment and simply point out that we judge pain as intrinsically bad no matter how the pain was caused, and no matter what the

---

[75] Wheatley and Haidt (2005).

consequences are. We cannot doubt that it is intrinsically bad, and that pleasure is intrinsically good. The SLS theory may even hold that we ascertain *a priori* that pain is bad and pleasure is good. Furthermore, even if pain has good consequences, it can only be at most an instrumental good depending on the situation. For example, a chest pain could motivate me to call an ambulance, which saves my life because the pain was the first sign of a heart attack. Once we accept this, we can bring in *Benevolence* to reach a moral theory. Suppose that someone offers you £1 to whack me on the head with a hammer, and the benefit to you will be less that then pain caused to me. Without *Benevolence*, you cannot deny that I will have a bad experience that I would prefer to avoid, but you can deny that you have a reason to care. You can adhere to *Prudence* and insist that you have a reason to pursue your own good, even if it negatively affects mine. If you accept *Benevolence*, though, you are morally obliged to turn down the offer. My well-being is just as valuable as yours, despite the fact that you cannot experience my pleasure or pain. Unfortunately, though, this brings us back to the dualism of practical reason, and the difficulties with the SLS theory's solution to the problem. On its own, hedonism cannot give a reason to care about other people's pleasure and pain. We need to solve the dualism of practical reason first.

Another point is that our belief that pain is bad need not be based on any rational or non-rational beliefs. The experience of pain and our reaction to it can be immediate. If you accidentally put your finger on an open flame, you will pull it away without thinking. It takes far more concentration and thought to keep it there. This does not mean that your reaction is irrational, but that it is non-rational. There are plenty of reasons why we do not like pain, but according to hedonism the reason why pain is bad is not because it has various consequences or evolutionary origins, but simply because we experience a certain sensation. We do not like feeling pain, and there is nothing more to be said. To say that this is rational or irrational misses the point, as physical sensations are not the sort of thing that can be rational or irrational. This means that the hedonist cannot *rationally* argue *a priori* that pleasure is good and pain is bad. We can see why by drawing an analogy with visual perception. Suppose I perceive a patch of blue because I am hallucinating, whereas in fact I am standing in front of a white wall. You can rationally argue that my perception is defective, that nothing blue is in front of me, that my hallucination has a particular cause and so on, and I can agree with all of it. But your arguments

will not make the hallucination go away. You cannot rationally argue that I do not perceive the colour blue. Equally, you cannot claim that I am irrationally perceiving the colour blue. Perceptions, being sensory experiences, are neither rational nor irrational. My *belief* that something blue is in front of me may be irrational, but my sensory experience is not. Similarly, if I feel pain, you may argue that the pain is unimportant or I should try ignoring it. And I can agree with you, but agreement will not make the pain go away.[76] Neither will it change my instinctive negative attitude towards it (I may agree that the pain is useful instrumentally, but that does not mean that I have a positive attitude towards the pain as pain. I have a positive attitude towards its usefulness, not the pain itself.). Pain, being a physical sensation, cannot be shown to be rational or irrational. De Lazari-Radek and Singer seem to implicitly agree with this point, because their arguments for hedonism are either that alternatives to hedonism do not work, or that objections to hedonism are mistaken. They do not provide a positive argument for hedonism. This may be because if we talked to agents who could never experience pleasure or pain, there would be no way to rationally demonstrate that hedonism is true.

Hedonism may still be correct, but it now appears that we identify what is good through our immediate experiences, which entails that we identify what is good *a posteriori*. The good of an agent who is incapable of feeling pleasure or pain would have to be something different, or perhaps the hedonist would claim that nothing would be good or bad for such an agent. This argument would be based on the fact that there are certain *a posteriori* experiences that the agent could not have. Ultimately, de Lazari-Radek and Singer's hedonist theory is based on *a posteriori* observations about the biological organisms we examine.

This feeds into a wider point, which is that de Lazari-Radek and Singer's conception of pleasure and pain does not have a lot of content. There is nothing that connects all pleasurable things other than that they are pleasurable. What, then, does it actually mean for something to be pleasurable? How do we identify what is pleasurable? The obvious answer is that we go out and look. We try different experiences, and we become good at estimating whether we would enjoy something based on similar past experiences. The proposition 'Pleasure is an intrinsic good' is not very determinate, because pleasure is more or less just what the agent finds pleasurable. It does not tell us what is pleasurable, and so the application conditions of

---

[76] I may agree that I should remove the cause of the pain or try to reduce it some other way, but this agreement will not eliminate the pain. *Acting* on the agreement may do so.

'pleasure' are very imprecise. We must therefore investigate *a posteriori* what agents find pleasurable, and fortunately there are a lot of pleasures that most people share – a pleasure in social contact, having varied interests, keeping busy with enjoyable tasks, various physical sensations, and so on. This means not that hedonism is an empty concept, but that it depends on *a posteriori* study of actual agents to flesh it out.

The core-apriorist could respond by trying to provide *a priori* tests for pleasure. We do not need to look at actual agents, he claims. Instead, we can work out from our armchairs that pleasure is intrinsically valuable. One way of doing so may be something like G.E. Moore's test. We take the concept of pleasure and ask whether a universe containing only pleasure would be good. Moore asks 'what value we should attach to [pleasure], if it existed in absolute isolation, stripped of all its usual accompaniments'.[77] Moore thought that the test shows that there are many things more valuable than pleasure; a pleasurable contemplation of beauty, for example, has an immeasurably greater value than the consciousness of pleasure alone.[78] Assume, though, that this test vindicated the hedonist. Since we worked this out *a priori*, does this not also vindicate the core-apriorist? No, claims the core-aposteriorist, because we did not actually work it out *a priori*. We cannot understand what pleasure is in the abstract, because there is nothing that connects all the pleasurable things other than that they are pleasurable. How, then, are we supposed to work out what a universe that only contained the value of pleasure would be like unless we brought our own experiences of pleasure to the test? If we had no idea what pleasure felt like, we would have no way of knowing whether it was good or not.

I have not reached the conclusion that the SLS theory's hedonism is wrong, though one may think that there are too many problems with it to be plausible. For example, one might agree with Mill that certain types of pleasure are more valuable than other types. My main aim has been to show that the SLS theory's hedonism has too little content to work as a theory of the good unless we add more content to the term 'pleasure' regarding what things are pleasurable and how we know that they are pleasurable. We can only do so, though, by examining *a posteriori* information about what agents find pleasurable. This theory of the good is thus not purely *a priori*.

---

[77] Moore (1993), p. 142. See also p. 236.
[78] Moore (1993), p. 145.

8 – Summing up

We can now sum up. The argument I have been making in this chapter seeks to develop the *a posteriori* strategy against the SLS theory. Return to the first two steps of the strategy:

a) At least some true determinate core moral propositions are not *a priori* self-evident. These determinate core moral propositions are either *a posteriori* self-evident, or can be known by using other propositions as premises.

b) At least some of these propositions are not knowable solely by using *a priori* propositions as premises. Any attempt will fail either because some of the purportedly *a priori* propositions contravene the principle *Sufficient Determinacy* or are really *a posteriori* propositions.

As far as (a) goes, three of the axioms are not determinate moral propositions in their original formulation. *Aim*, *Impartiality* and *Treatment* face the same problem. The terms in the axioms lack sufficiently precise application conditions, and thus cannot be determinate core moral propositions. They also risk contravening Sidgwick's criteria (1) and (4) for being self-evident axioms. Furthermore, the only way we can alter them so that they can be used to justify determinate moral propositions is by using the 'other resources' of a normative moral theory, such as rule utilitarianism. This is when step (b) becomes relevant. Whether those resources are *a priori* is an important question, as if they are not, the altered axioms cannot be either. So even if the altered axioms are determinate, we cannot assume that we can attain knowledge of them *a priori*. They may depend on *a posteriori* assumptions about actual agents, and hence they will not be *a priori*.

Prudence is a special case. According to Sidgwick, it is self-evident, but it conflicts with *Benevolence* and leads to the problem of the dualism of practical reason. We badly need a solution to this problem, otherwise we have no way to refute rational egoism. Unfortunately the SLS theory's solution is not going to help the core-apriorist. Its reliance on Sharon Street's evolutionary debunking argument to cast doubt on rational egoism implies that *Benevolence* must be defended *a posteriori*. If the solution works, then it works on *a posteriori* grounds.

Meanwhile, if the dualism of practical reason is not resolved, then the SLS theory has a major hole. Neither option helps the core-apriorist.

De Lazari-Radek and Singer attach a particular hedonist theory of the good to their utilitarianism and their metaethics. Unfortunately their theory appears to be *a posteriori*, which is further bad news for the core-apriorist. Hedonism states that pleasure is good and pain is bad, but the only way we can ascertain the truth of the SLS theory's hedonism is to examine our personal experiences, and this examination is *a posteriori*. The SLS theory's hedonism does not have an *a priori* basis.

I conclude that de Lazari-Radek and Singer fail to show that the SLS theory can be used to provide a set of *a priori* self-evident axioms that can be (or can be used to justify) *a priori* determinate core moral propositions. However, there may be another way that the core-apriorist can use moral intuitionist theories. Matthew Kramer proposes that we can accept or reject certain metaethical claims by examining the moral consequences that they produce. If accepting a purported claim leads to accepting morally repugnant consequences, we should reject the claim. Nazism is morally repugnant, so any set of purported claims that allow it must be wrong. Metaethical claims are not simply claims about morality, but are moral principles themselves. In the next chapter, I will examine Kramer's metaethical theory, and argue that even if it works, it likely does not work specifically as a core-apriorist theory.

Chapter 4 – Matthew Kramer's moral realism

1 – Basic moral principles

Matthew Kramer endorses a metaethical theory that he calls 'moral realism as a moral doctrine'.[1] He claims that true moral principles (such as the principle not to torture babies for pleasure) are based on certain basic principles that are self-evident. These basic principles are synthetic and knowable *a priori*. By 'knowable *a priori*', Kramer means that the principles are knowable non-empirically.[2] They include principles such as moral supervenience and moral objectivity. Kramer bases his definition of moral supervenience on Simon Blackburn's, having slightly modified it. Roughly speaking, moral property M supervenes on natural property N if M is not identical to N, and either:

a) it is impossible for M to change without N also changing; and

b) it is impossible for two things to have N to the same degree without them both possessing M to the same degree, or both failing to possess M.[3]

When Kramer claims that moral principles are objective, he means that they are mind-independent, determinately correct, uniformly applicable to all moral agents, true at all times and everywhere, impartial, and agreed upon by most people without biases or corrupting influences. Statements of moral principles are truth-apt.[4]

The claims of moral objectivity and moral supervenience are not only metaethical claims, as they are also moral principles. Whilst there are a few exceptions (which we will see later), Kramer thinks that metaethical claims are substantive moral principles, such as the principle not to torture babies for pleasure is. This is an uncommon move. Metaethical claims are typically seen not as moral principles, but claims about moral principles, facts and properties. The claim that moral facts are objective, say, is not seen as a moral principle itself. To Kramer, though, this is a false distinction. The claims discussed in metaethics are themselves moral principles,

---

[1] Kramer (2009); (2017).
[2] Kramer (2009), pp. 94-95.
[3] Kramer (2009), pp. 304-305; Blackburn (1973), pp. 114-115.
[4] Kramer (2009), p. 35.

and can only be defended morally. We cannot, for example, argue that moral supervenience is true via logical or non-normative metaphysical argument.[5] Instead, we must defend supervenience on moral grounds. The metaethical claim that moral properties supervene on natural properties is itself a moral principle, and not just a non-normative metaphysical claim about moral principles. *Basic* metaethical claims – in other words, self-evident metaethical claims – are therefore basic moral principles, and they justify various non-basic moral principles, such as the principle not to torture babies for pleasure.

Basic moral principles are *a priori*. The basic moral principles that Kramer examines, supervenience and objectivity, also seem to be expressed in fairly determinate ways. We can therefore say that core moral propositions express basic moral principles, and some of these propositions are determinate. So, because basic moral principles are *a priori*, Kramer's metaethics is a core-apriorist theory. In this chapter, I shall argue that Kramer's metaethics cannot successfully defend the CAT.

Why think that we can only defend basic metaethical claims as moral principles? Kramer endorses a version of what he calls Hume's Law, which claims that

'[t]here is no valid argument in which all the premises are non-moral (and logically consistent) and in which the conclusion is a substantive moral proposition.'[6]

Kramer thinks that there is no way that we can defend moral propositions by appealing solely to non-moral premises. Hume's Law thus entails two things. First, if a moral proposition can only be defended by appealing to basic metaethical claims, then the basic claims must themselves be moral. Second, since basic metaethical claims are basic moral principles, Hume's Law implies that they can only be defended morally themselves. We should be clear here about what such a defence would look like. The fundamental reason why moral supervenience is true is because moral supervenience is true; being a basic moral principle, it is self-evident. Basic moral principles 'are their own rock-solid foundations. True in all possible worlds, they and their

---

[5] Since various metaethical and moral claims can be metaphysical claims (such as the claim that moral supervenience is true), Kramer's moral realism implies that various moral principles can be expressed as metaphysical claims. But they must be irreducibly normative metaphysical claims, otherwise it may be possible to argue non-normatively that these claims are correct. This is impossible if they can only be defended morally.
[6] Kramer (2009), p. 6. The law's name is a reference to Hume (2000), section 3.1.1.27.

objectivity are unremittingly self-sustaining'.[7] However, there are other reasons to believe that basic moral principles are true as well, and Kramer thinks that such reasons are a 'superlative means for enhancing one's grasp of [the principles'] content and of the attractiveness of [their] content'.[8] Basic moral principles are not true because we can grasp them, or because moral agents find them attractive, but our grasp of the principles gives us reason to believe the principles.

Kramer's views may not be the most popular, but it is important to note that he is not alone. Ronald Dworkin is another prominent philosopher who argues that issues traditionally taken to be metaethical issues, such as moral epistemology and the possibility of mind-independent moral truths, are moral issues. One can only argue for mind-independence by using a moral argument, as opposed to, say, a non-normative metaphysical argument. Actually, Dworkin goes further than Kramer does, as he claims that metaethics entirely collapses into first-order normative moral philosophy.[9] At the first-order, our arguments are moral, and are designed to tell us which moral beliefs we should have and how we should morally behave. At the second-order, arguments are about the nature of moral properties, or how we can attain moral knowledge, or what we are actually doing when we engage in moral discourse and the like. Metaethical arguments do not directly tell us what our moral beliefs or moral behaviour should be, though they may have an effect on first-order moral arguments. Whilst Kramer explicitly refuses to collapse metaethics into first-order moral philosophy, both philosophers appeal to Hume's Law in their work. One might think that, since Hume's Law tells us that we cannot get moral conclusions from exclusively non-moral premises, we should be moral sceptics. There is no way of 'breaking into' morality from outside, via logical or non-normative metaphysical argument and the like. But Kramer and Dworkin believe that Hume's Law simply limits the methods with which we can attain moral knowledge.[10] It does not prevent us from attaining moral knowledge.

So Kramer is not making a lonely journey by himself. In fact, we may not need to completely accept his position in order to have some sympathy with it. For example, Simon Blackburn

---

[7] Kramer (2009), p. 2. Kramer means *morally* possible worlds, not logically possible worlds. I will return to this point later.
[8] Kramer (2009), p. 363.
[9] Dworkin (2011), pp. 8-11. Another recent example of a moral argument for a metaethical position is Melis Erdur's argument that moral realism fails on *moral* grounds (2016),
[10] Dworkin (2011), p. 17.

writes that philosophers such as David Wiggins and John McDowell compare explanations of the practice of ethical judgement by 'invok[ing] the very commitments of that practice'. They thus refuse to 'stand outside ethics', and Blackburn is not one to complain. He claims that

> '[e]thical notions require ethical sensibilities to comprehend them… [W]hy should it not require an ethical sensibility to comprehend an explanation of the views we hold? Only those who perceive friendship as good will understand why we do so…'[11]

However, Blackburn's analysis of supervenience, which he applies to moral properties, is a logical analysis.[12] He therefore parts company with Kramer regarding the moral status of the analysis, even though he indicates that metaethics can require specifically moral arguments. Blackburn's words here echo Hume again, this time concerning Hume's Sensible Knave. The Knave asks why he should not be immoral when he can get away with it and when it does not cause so much damage that it endangers society. Hume's response if that if the Knave really has no inclinations against immorality, there is no way to reason with him. We will never be able to convince him that he should act morally.[13] Kramer would say that this is what we should expect, given Hume's Law. If we do not accept moral notions in the first place, we cannot argue for moral conclusions.

We should briefly note that whilst Kramer does not call himself a moral intuitionist, we can very easily read his metaethics as an intuitionist theory, because basic moral principles are self-evident. This means that it fits Hallvard Lillehammer's definition of intuitionist theories that we saw at the start of chapter 3.

2 – A metaethical collapse?

An initial objection to Kramer's metaethics is that if metaethical principles are actually moral principles, then he cannot prevent metaethics collapsing into first-order normative ethics. If this difficulty is a fatal one, then the core-aposteriorist can use it to argue that we cannot use

---

[11] Blackburn (1988), p. 174. See also McDowell's comments on understanding the viewpoint of a moral community ((1998), pp. 198-218).
[12] Blackburn (1973), p. 115.
[13] Hume (1998), sections 9.22-9.25.

Kramer's metaethics to support the CAT. However, we shall see in this section that Kramer can resist it.

Kramer makes a distinction between abstract and concrete moral principles. Moral principle A is more abstract (and less concrete) than moral principle B if and only if the falsity of A would affect the truth value of a wider range of moral principles than the falsity of B would.[14] Kramer believes that metaethical theories carry with them 'highly abstract' moral implications, so they are highly abstract moral doctrines.[15] Consequently, a metaethical theory can be consistent with many conflicting moral principles that are more concrete. However, even if the theory is consistent with both of the concrete moral principles 'Telling white lies is morally permissible' and 'Telling white lies is not morally permissible', it can have many other moral implications.[16] The fact that the theory has moral implications, however abstract, makes it a moral theory.

Despite this, though, Kramer is at pains to deny that metaethics is just first-order ethics. A metaethical theory is not simply a first-order moral theory. First, some metaethical questions can be answered with non-moral arguments. Kramer believes that metaethical expressivism seeks to explain 'the sundry practices in which people form and communicate moral judgements'.[17] This is a metaethical issue, but it is not a moral one, because it asks nothing about the nature of moral principles and properties. Metaethical theories that do examine such issues, such as moral realism, are moral theories for this reason.[18] Second, moral realism and other metaethical theories still make 'metaphysical and substantive and epistemological enquiries', even though such enquiries are also 'substantive ethical enquiries'.[19] They are both types of enquiry at once, and presumably a key difference between, say, quasi-realism and utilitarianism is that utilitarianism says little about the metaphysics of moral properties or the possibility of moral knowledge. Quasi-realism says considerably more, which separates it from first-order moral theories.

---

[14] Kramer (2009), p. 10.
[15] Kramer (2017), p. 198.
[16] See also Kramer (2009), p. 4, which points out that philosophers who share the same first-order moral theory may have very different metaethical views.
[17] Kramer (2017), p. 204.
[18] Huw Price (2004) would disagree. He endorses 'subject naturalism', which claims that questions about the nature of objects and properties can only be answered by examining how actual humans use language. So in order to examine moral properties, it is necessary to (naturalistically) examine how we use moral language. Kramer's separation of theories such as expressivism from theories such as moral realism is thus mistaken. However, since this does not affect Kramer's arguments for his moral realism, I will set this issue to one side.
[19] Kramer (2017), p. 208-209.

In my discussion of Kramer I will follow him in restricting terms such as 'metaethics', 'metaethical' and the like to metaethical theories that in Kramer's opinion are moral theories. I will not be arguing that theories such as expressivism are core-aposteriorist theories. Kramer would not concede that such theories profit very much from *a posteriori* evidence in any case. He believes that 'the questions addressed by anthropologists and sociologists who study human beings' practice of moral discourse are not the questions addressed by expressivists', and that whilst expressivism does contain *a posteriori* elements, these elements are 'elementary platitudes'.[20] This relegates the *a posteriori* to a rather unimportant role, but this is not a battle that the core-aposteriorist has to fight. 'What do agents do when they express moral judgements?' is not an epistemological question. For epistemological questions, Kramer believes, we have to turn to theories such as moral realism.

Kramer also argues by analogy with logic and mathematics. Meta-mathematical and meta-logical theses, he claims, are second-order, but lie 'squarely within the mathematical or logical domain'.[21] He does not provide any examples, but an obvious choice is Gödel's incompleteness theorems. The theorems describe the limitations of mathematical systems under various conditions (and they thus describe something about the nature of those systems), but the theorems themselves are justified mathematically. If meta-mathematical enquiries are nevertheless second-order, then there is no reason why metaethical enquiries cannot be.

3 – Moral repugnance

It seems unlikely that the core-aposteriorist can argue against Kramer by claiming that he makes metaethics collapse into first-order moral enquiry. We should therefore move onto other criticisms. The next one I shall examine is interesting not because it works (I will show that it probably does not), but because its solution points to a more troubling objection to Kramer's metaethics. This new objection is based on the *a posteriori* strategy, and it is the one that the core-aposteriorist should push. If the new objection is right, then Kramer's metaethics cannot be a core-apriorist theory.

---

[20] Kramer (2017), p. 205, fn 17.
[21] Kramer (2017), p. 206.

If Kramer meets someone who does not see why objectivity must be a basic moral principle, how can Kramer defend his position? He cannot defend objectivity on logical, non-normative metaphysical or other non-moral grounds, because that would contravene Hume's Law. So there must be some way to defend objectivity on moral grounds. Kramer's answer is to appeal to moral repugnance.[22]

To argue for or against a purported basic moral principle, Kramer claims that we should examine the moral consequences of adopting or rejecting the principle. Consider a candidate basic moral principle, that moral relativism is true. Kramer points out that if this principle were right, then it could be morally permissible to torture babies for pleasure. This is so morally abhorrent that we should reject the principle.[23] Our 'concrete moral judgements' – that torturing babies for pleasure is wrong, that we should oppose racism, and so on – forbid us from accepting the principle. Kramer explicitly uses this method when defending moral supervenience. We have a moral reason to accept moral supervenience, he claims, because if we do not accept it, we risk our 'moral world… spinning out of control'. If moral supervenience were false, it would provoke 'unprincipled moral chaos'.[24] We would have no way to work out reliably what the moral implications of our actions were, and our choice of which actions to perform would be morally arbitrary. This chaos would falsify many other moral principles that we want to retain. Moral values such as 'fairness and equity and integrity' would be 'largely or wholly inoperative'.[25] Denying supervenience would be morally intolerable, says Kramer, so moral properties must supervene on natural properties.

Similarly, we should accept moral objectivity because of the consequences that would arise if morality were not objective. Whilst Kramer cannot appeal to logical absurdity when arguing against moral relativism (as moral relativism is logically possible), he draws analogies with logical principles. Accepting moral relativism, he claims, is like denying the logical law of non-contradiction.[26] Deny the law of non-contradiction, and you get an utter logical free-for-all. No matter what you prove, you can still affirm the opposite. Proving that p does nothing to disprove not-p. This leads to logical inconsistency, as it would be possible, if we rejected the law of non-contradiction, to accept it as well. We can thus defend the law by objecting to the consequences

---

[22] E.g. Kramer (2009), pp. 362-364.
[23] Kramer (2009), pp. 30-36.
[24] Kramer (2009), p. 360.
[25] Kramer (2009), p. 363.
[26] Kramer (2009), pp. 45-46.

of rejecting it. This defence of the law is not a justification of it, since the law is its own logical foundation. It relies on no other premises for its justification.

Moral relativism also results in a free-for-all, this time moral. Whatever moral principle is true for you, it may have no moral significance for anyone else. If moral relativism were true, thinks Kramer, morally bad consequences would result, because no moral argument would show the baby-torturer to be wrong unless the torturer voluntarily accepted the argument's premises. It would only show him to be morally wrong based on contingent, mind-dependent facts, such as which moral principles he himself accepted. This is not logically inconsistent, so we cannot say that moral relativism is logically mistaken, but it is morally mistaken.

The law of non-contradiction is necessary in classical logic, because classical logic would collapse without it. Would morality collapse if relativism were true? We could still make moral arguments that would be logically valid, so the problem is not a logical one. Nevertheless, Kramer holds that morality would collapse in such a case. Moral arguments would depend on contingent, mind-dependent premises, and Kramer rejects such a possibility because the consequences of this moral relativism are morally repugnant. They (morally, rather than logically) *cannot* be wrong. Morality would collapse without them. This, incidentally, is why Kramer says that basic moral principles are true in all morally possible worlds, rather than all logically possible worlds. A world is morally possible if and only if 'every normative state of affairs within it is consistent with the existence of all the basic principles of morality'.[27] As torturing babies for pleasure is not consistent with the existence of some basic moral principles, a world where it is morally permissible would be morally impossible.

To be clear, Kramer is not saying that moral relativism is repugnant because it will inevitably lead to people torturing babies for pleasure. It could be that every agent in the world was a relativist who firmly believed that torturing babies for pleasure was morally wrong. No agent would ever actually do it. However, as relativists, they would have to accept that if any society *did* believe the proposition 'Torturing babies for pleasure is not morally wrong', that belief would be true for that particular society. Furthermore, they would have to accept that if the universe contained no minds whatsoever, the statement 'It is not the case that torturing babies for pleasure would be morally wrong' would be true. Kramer rejects this, holding that some moral

---
[27] Kramer (2009), p. 158.

propositions are true or false regardless of whether the universe contains any minds or not. So Kramer's objection to relativism is not based on any consequences of relativism that may well arise in the actual world. Relativism is morally repugnant because of its moral implications, rather than because of any actual or possible danger it poses to actual agents. Even if we were certain that no relativist would torture babies for pleasure, relativism would still be morally repugnant.[28]

The most obvious objection to Kramer's argument needs a little development to really stick. The initial idea is that Kramer seeks to simply establish moral truth by 'horrified assertion', as Samuel Clark puts it.[29] Torturing babies for pleasure is so morally shocking to us that we reject any metaethical theory that implies that it could be morally permissible. However, our shock will not convince any society that does not find it shocking that it is morally impermissible. Such a society may use perfect logical and non-normative metaphysical reasoning, but that will not show the members of that society that torturing babies for pleasure is morally wrong, because Kramer has already said that we cannot reach moral knowledge purely through such reasoning.

Kramer certainly does not need to claim that our immediate or automatic responses to moral problems infallibly tell us which moral principles are correct. There are situations that appear or do not appear to be morally wrong at first sight, but we subsequently change our minds when given time to reflect on them, or when we discover new non-moral facts about the situations. Performing operations without anaesthetic sounds barbaric, but if we find out that no anaesthetic is available, and that the operation is the only way to save the willing patient's life, we are less likely to think it unethical. Any remaining discomfort is not moral discomfort. But it seems possible for a society to decide, even after a long period of reflection and non-moral fact-checking, that a particular metaethical principle is wrong because it has morally repugnant consequences, whilst another society in the same position would not find the consequences morally repugnant. There is no non-moral argument to show that one society is right and the other is wrong.

Clark therefore believes that the objection is fatal for Kramer's metaethics. We cannot discover a correct metaethical or moral system simply by checking what 'horrifies' us, even if we have perfect reasoning abilities and know all the relevant non-moral facts.

---

[28] Kramer (2009), pp. 32-33.
[29] Clark (2011), p. 426.

I am not going to argue that this objection is correct. In fact, I will shortly show that it fails. My purpose in developing it is not to show that Kramer's metaethics fails, but to concentrate instead on the solution. I will examine two possible solutions. The first is to appeal to what I will call the self-supporting condition, which concerns how basic moral principles relate to each other. The second defence is to draw a comparison between metaethical principles and first-order moral principles. The second defence is much more successful than the first, and Kramer can use it to withstand the objection. The trouble now is that the core-aposteriorist can use the *a posteriori* strategy to argue against Kramer's metaethics as a core-apriorist theory.

4 – The self-supporting condition

Basic moral principles do not stand individually. They are connected to each other. If two purported principles give morally good results individually but conflict irreconcilably over various moral questions, Kramer would think that we must revise them. This would be sensible, otherwise basic moral principles might conflict with each other. If two basic moral principles imply that we should act in contradictory ways, we would have no way to discover which principle to follow. We would also be unable to know which actions to perform.

One problem with calling sets of basic moral principles 'consistent' is that it seems to imply that it is possible to defend the sets by simply appealing to logical or non-normative metaphysical consistency, when Kramer denies that we can defend moral principles purely on logical or non-normative metaphysical grounds. Kramer allows that two moral propositions can logically contradict each other, but he also claims that no purported moral principle is wrong *just because* it is logically mistaken or inconsistent.[30] Incorrect moral principles are not necessarily logically irrational, but they are unreasonable, and logically irrational principles are a type of unreasonable moral principle. Unreasonable moral principles fall 'below a threshold of moral extenuability or intellectual credibility'.[31] So I will refer instead to the self-supporting condition, which states that all other things being equal, we should prefer a set of basic moral principles that all mutually support each other over a set that does not. The principles mutually support each other if and only if they do not justify contradictory non-basic moral principles, and they do

[30] Kramer (2009), pp. 289-294.
[31] Kramer (2009), p. 293.

not justify non-basic moral principles that collectively it would be unreasonable for a moral agent to hold. To be clear, it is the set of basic moral principles that is self-supporting, not the individual principles themselves. A set is self-supporting if and only if all the principles in the set mutually support each other.

The self-supporting condition should be used as a means to understand the basic moral principles that we accept. Like the appeal to morally repugnant outcomes, it is not the fundamental reason why a particular set of purported basic moral principles is correct or incorrect. It simply explains the attractiveness of the set.

Kramer can try defending the basic moral principles that he favours by appealing to moral consequences and the self-supporting condition. However, the second defence collapses into the first, because of the reason why we should endorse the self-supporting condition. As noted above, the self-supporting condition cannot be an appeal to logical or non-normative metaphysical consistency. There must therefore be another reason why the condition can be used as a defence of basic moral principles, and the reason can be found in the results of adopting the condition. If we did not adopt it, then we could endorse sets of basic moral principles that justify contradictory non-basic moral principles. This set might both justify ϕ-ing and not ϕ-ing in identical circumstances, in which case we would be unable to decide what to do. If we picked which principles to follow arbitrarily, this could have a morally repugnant result. So Kramer will reject sets of basic moral principles for not being self-supporting. Their lack of self-support is a reason to reject them.

However, this is an appeal to moral consequences. If we decide to act arbitrarily, we may act in a morally repugnant way. Meanwhile, if we cannot decide between two principles (assuming that it is logically possible to pick neither), then we may choose not to do anything, which itself may be morally repugnant. If a set of basic moral principles is not self-supporting, there are too many opportunities for morally repugnant consequences to arise, so we should accept the self-supporting condition to avoid these consequences. The appeal to the condition is thus not a separate defence from the appeal to morally repugnant consequences, but a variation of it. We can only defend basic moral principles by appealing to the moral consequences of accepting them directly or indirectly (i.e. via the self-supporting condition).

So appealing to the self-supporting condition will not defuse Clark's objection. We now turn to the second response to the objection, which develops a comparison between basic moral principles and first-order moral theories.

5 – First-order moral theories

Kramer thinks that a large number of different non-basic moral principles are consistent with basic moral principles. Basic moral principles, as we have seen, may be logically neutral amongst many conflicting non-basic moral principles, such as whether or not to tell white lies to save face in social situations.[32] It seems clear, though, that the basic moral principles you choose will affect what the correct first-order moral theory is. Kramer considers whether a form of utilitarianism could allow that it could ever be morally right to torture children for fun. It is not logically inconceivable, but it is inconceivable in the sense that it would be beyond 'the limits imposed on any decent person's imagination by [its] overwhelming unreasonableness'. Any utilitarianism that allows torturing children for fun cannot be 'a doctrine that can be endorsed under any optimal conditions for moral deliberation'.[33]

If a first-order moral theory lies outside the bounds of moral reasonableness, it cannot be correct. So if a first-order moral theory conflicts with basic moral principles, resulting in morally repugnant consequences, it cannot be correct either. This only holds, though, if we are more confident about the basic moral principles we choose than about the first-order moral theory. If a first-order moral theory clashes with the principle of objectivity, Kramer would abandon the theory. Objectivity is too morally important to give up. However, if we are more confident about a first-order moral theory than a purported basic moral principle that it conflicts with, then the purported principle may have to give way.

Though Kramer does not mention it, he might be able to use reflective equilibrium as a method with which we can reconcile basic moral principles and first-order moral theories. When we are morally evaluating a situation, and we find out that our moral judgements conflict, it is not always clear which judgements are the right ones, and so we do not know which moral principles are correct. John Rawls proposes that we should use reflective equilibrium, which

---

[32] Kramer (2009), pp. 9-10.
[33] Kramer (2009), p. 159, fn 2.

involves revising our judgements in acceptable ways until they conform to each other. We may also be able to revise what we understand the situation to be. Rawls claims that 'eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgements duly primed and adjusted'.[34]

Suppose that we have good reason to accept basic moral principle A and first-order moral theory B, but A and B are in conflict. By examining them more closely and looking at the outcome of the situations to which they apply, we may be able to amend them to become A* and B*, which do not conflict. Kramer appears to be favourable to this, though as I noted he does not explicitly use reflective equilibrium. If it helps us to avoid morally repugnant consequences, then it seems to be a tool that Kramer can use.

Kramer suggests that first-order moral theories provide an ethical purpose. This purpose might be to promote a certain result when moral agents interact, or to promote a particular aim that moral agents have, or some other outcome. First-order moral theories tell us what the purpose is. Kramer claims that in order to achieve this purpose, we need to accept various basic moral principles. The basic moral principles are attractive because they help us achieve this purpose.

'Any imputation of a general point or purpose to ethics… is a substantive ethical thesis. It attaches fundamental ethical value to an end, which transmits that value to the ethical norms and decisions by which the end is promoted. A focus on the intercourse amongst human beings – with an orientation towards advancing their positive interaction or limiting their negative interaction – is one attractive route for arriving at the cardinal value or purpose of ethics. Writers within the broad utilitarian tradition in ethics… generally favour such a route, but so do many others. Some writers within the Kantian tradition concentrate instead on the integrity of the individual rational agent as the primary concern of morality. Given that the deliberations and judgements of such an agent cannot partake of integrity unless the requirement of supervenience is heeded, a Kantian justification for that requirement is readily available.'[35]

---

[34] Rawls (1999), p. 18.
[35] Kramer (2009), p. 333.

The final sentence is important. If Kantianism implies that the primary concern of morality is the integrity of rational agents, and if this can only be the case if moral supervenience is true, then Kantianism cannot work without moral supervenience. Cue morally repugnant consequences, as no moral agent would be able to have integrity. Once more, the truth of Kantianism is not the fundamental reason why moral supervenience is true, because moral supervenience is a basic moral principle. The fact that Kantianism implies moral supervenience merely tells us why the Kantian finds it attractive. She cannot have Kantianism without it.

It is the same for consequentialism. Simon Blackburn also talks of morality having an aim, saying:

'[I]t is natural to associate projectivism with consequentialist moralities, in the following way. A projectivist [will attempt to give]… an explanation of the practice of moralizing. This turns to its function, and particularly to its social function. In [J.L.] Mackie's terms, morality is an invention that is successful because it enables things to go well among people with a natural inheritance of needs and desires that they must together fulfil.'[36]

Kramer states that the purpose that Blackburn finds is 'broadly consequentialist in character', and that 'the invocation of such a purpose is integral to Blackburn's account of supervenience'.[37] Once more, a first-order moral theory is used in support of a basic moral principle.

How does this help against Clark's objection? It points to a similarity between first-order moral theories and metaethical theories in Kramer's metaethics that he can exploit. It seems quite unremarkable to argue for and against first-order moral theories on moral grounds. For example, many problems associated with direct act utilitarianism are not that it is logically impossible for a moral agent to be a successful direct act utilitarian, but that it would lead to morally bad outcomes. One such problem, according to Bernard Williams, is that it entirely ignores the integrity of moral agents and their personal projects.[38] This is not an appeal to logical incoherence or non-normative metaphysical difficulty. It is a normative problem. Nobody

---

[36] Blackburn (1985), p. 164.
[37] Kramer (2009), pp. 347-348. Blackburn thinks that moral supervenience is logically or conceptually necessary (e.g. (1984), p. 184), whilst Kramer thinks that it is morally necessary, but otherwise Kramer endorses Blackburn's definition of supervenience.
[38] Williams (1973), pp. 108-118.

would claim that Williams is wrong because his objections are not logical or non-normative metaphysical objections. To demand that they be so would miss the point.

It would be quite possible for the utilitarian to agree that Williams is right, but claim that it is unimportant. Maybe the utilitarian does not care about our integrity. This is not an attractive position, and the utilitarian will not find many supporters. But without further work, we cannot claim that the utilitarian is logically irrational. What we can claim is that no sane, reasonable person would entirely dismiss the importance of our personal integrity. This is why Williams's argument should worry utilitarians. They should be trying to rebut it, not rejecting it just because their moral position is logically possible.

Suppose that the utilitarian says that under some conditions it would be morally permissible to torture babies for pleasure. Kramer steps in, saying that this simply is not morally plausible. No sane, reasonable person could truly believe it. Now suppose that a metaethicist declares that moral relativism is true, and that all of the horrors that Kramer fears are therefore true. Many morally repugnant acts, claims the metaethicist, are morally possible. Moral relativism is a metaethical position, but metaethical positions are substantive moral positions – as we have just seen, your metaethical position affects your first-order moral position and *vice versa*. Once more, Kramer puts his foot down. If moral relativism implies that morally repugnant acts are morally possible, then no sane, reasonable person would truly think that moral relativism is true. This is no more wishful thinking than Williams's claim that people should have their integrity respected by a first-order moral theory.

The point here is that first-order moral propositions can be criticised and defended on moral grounds. Therefore, if metaethical propositions are substantive moral propositions, we should be able to criticise and defend them on moral grounds as well. In fact, if Hume's Law is true, then this is exactly what we should expect. People who seriously believe that it is morally permissible to torture babies for pleasure, to drive over pedestrians for sport and to stab people for the thrill of it are morally unreasonable to the point of moral insanity. Their position may be logically coherent, but this is morally irrelevant. The fact that someone might not be able to recognise that a moral principle exists is not evidence that the principle is unknowable, any more than people suffering from dyscalculia prove that mathematical axioms are unknowable. Any metaethical theory that allows morally insane outcomes is a moral catastrophe.

We now have a response to Clark. Kramer states that metaethical claims can only be defended morally. Clark objects to this, claiming that such methods are far too weak. If two agents disagree over whether something is morally repugnant, there is no way to resolve the disagreement. The problem for Clark now is that if we have no problem using moral methods to defend substantive moral claims made by first-order moral theories, we should have no problems using moral methods to defend substantive moral claims that are also metaethical claims. In effect, Kramer can say the following: we can object morally to the substantive moral claim that torturing babies for pleasure may be morally permissible. This objection implies that any first-order moral theory that endorses the claim is wrong. Moral relativism, even though it is a metaethical theory, also endorses this claim. So the objection also implies that moral relativism is wrong. If there is something that stops the objection working against moral relativism, even though it works against first-order moral theories, it is up to Kramer's opponent to show what that something is. This response makes a rather strong claim, but Kramer cannot argue for a weaker one. He cannot simply argue that there are moral arguments as well as non-moral arguments against moral relativism. His acceptance of Hume's Law, which implies that substantive moral claims can only have moral justifications, will not allow him to do so.

This reveals a key point: Clark's objection can only make headway if Hume's Law is wrong. In effect, Clark asks 'How do we identify metaethical truths?' Kramer replies 'We use moral methods'. Clark then responds 'But we cannot use moral methods until we justify them, and we cannot do so by using other moral methods'. This implies that Clark is looking for a non-moral method, such as some sort of logical or non-normative metaphysical argument. Given how Kramer has set up his moral realism, though, Clark cannot do this without assuming from the outset that Hume's Law is untrue. Hume's Law claims that there is no non-moral method. There is no way to reach any moral conclusion without using moral premises, and one will only accept moral premises if one is already a moral agent.

If Hume's Law is correct, then Kramer can use similarities between first-order moral theories and metaethical theories to defend his view that we can identify moral principles by using moral repugnance. It appears uncontroversial that we can use our sense of what is morally repugnant to discover first-order moral truths; that is in effect what Williams's argument against utilitarianism relies upon, for example. Kramer can use this to argue that the discovery of

metaethical truths is similar. We are constrained by what is morally possible. Suppose that a society rejected moral objectivity and had a bizarre first-order moral theory that allowed them to torture babies for pleasure. Their first-order moral theory would be intolerable. It is hard to imagine a reasonable moral agent who would condone baby-torture. Suppose further that the society thought that torturing babies for pleasure was morally permissible because they had rejected moral objectivity. Why can we not go further now and claim that the society's metaethical theory must also be wrong because it is morally intolerable? We end up rejecting both the first-order moral theory and the metaethical theory for the same reason: they both make a substantive moral claim that is morally repugnant. If we can reject the claim on moral grounds when it is made by the first-order moral theory, why can we not do the same when the claim is made by the metaethical theory?

It is no objection that two agents could still disagree over whether torturing babies for pleasure is morally impermissible, because one agent is morally repulsed by it and the other is not. In order to find out who is right, we can appeal to what decent people who recognise the boundaries of what is morally possible think. Decent people do not torture babies for pleasure. Furthermore, decent people are often able to identify moral truths. Kramer does not claim that decent people can have perfect moral knowledge, but they will generally know what morality requires of them, and they will certainly be able to identify that torturing babies for pleasure is morally wrong.

Clark can reply that this just pushes the objection back a stage. How do we resolve a moral disagreement? By asking decent people what they would find morally repugnant. But then we need to know who the decent people are, otherwise we are no closer to finding out what is morally permissible. However, this objection again asks for a non-moral response. It assumes from the outset that Hume's Law is wrong. In order to find out who the decent people are, we need evidence that their moral judgements are accurate. If Hume's Law is correct, though, that evidence can only be moral.

Hume's Law implies that if an agent stands completely outside morality, there is no non-normative metaphysical or logical or other non-moral method to convince him of moral truths. We can only know true moral principles if we are already moral agents, the 'decent people' that Kramer talks of. Decent people may disagree over various moral claims, but that does not imply

that moral knowledge is impossible either. Maybe a particular claim fails to have a determinate truth value, or the disputants do not have sufficient moral or non-moral information to find it. Another possibility is that they disagree over a 'minor' moral principle, such as whether it is morally permissible to tell a white lie to save face in a social situation. Kramer allows that such principles need not be morally necessary – the basic moral principles are morally neutral concerning the truth of minor principles, so moral agents can agree on all the basic moral principles without agreeing on all moral principles that have determinate truth values. However, a lot of the time the disputants will manage to reach an agreement over a particular moral principle. They can recognise the boundaries of what is morally possible, so they will be able to recognise (for example) that torturing babies for pleasure is wrong.

This response can only work if Hume's Law holds. If it does not, then there could be non-moral methods to attain moral knowledge. Kramer would therefore be only telling us half the story if he continued insisting that we can recognise moral truths by discovering what is morally repugnant. He would owe us some non-moral explanation for why our sense of what is morally repugnant is a reliable guide to moral truth. But if Hume's Law holds, then such a non-moral explanation is impossible. At most, we can only object that Kramer has not identified the correct moral method to attain moral knowledge.

It is worth noting a weaker objection that Clark can make. Kramer appears to argue that being a moral agent is an all-or-nothing position; if you do not accept that something that is morally impermissible really is morally impermissible, then your moral position is morally impossible, and so you cannot be a moral agent. However, it may be possible for an agent to correctly ascertain everything that is morally possible or impossible except for one case. Usually this agent is entirely right when he says that something is morally possible, but it just so happens that when it comes to this particular case, he believes that it may be morally possible, and it is not. As an analogy, think of professional mathematicians who thought that Fermat's Last Theorem was an unsolvable problem. This was a reasonable belief to hold, as the Theorem stubbornly resisted a solution for centuries, but the belief was wrong. This does not mean that the mathematicians were not mathematical experts in general. It just meant that they were wrong about a particular mathematical question.

Kramer need not have any problem with the erring moral agent, though. If somebody consistently gets moral possibilities wrong, then it is reasonable to argue that they are not really a moral agent, because they turn out to be unable to say what is morally possible and what is not. However, the erring moral agent appears to genuinely be a moral agent; after all, he consistently gets most moral possibilities right, though he is wrong about this particular case. Kramer has various ways to explain what is happening here. It could be that the erring moral agent gets it wrong because the moral principles he accepts are not fully mutually supporting. For example, if he thinks that it is always morally impermissible to torture people for pleasure, he will have a hard time reconciling this with his belief that it may be morally permissible to torture babies for pleasure. Another possibility is that he just has a moral blind spot. Something prevents him from realising that this particular case is necessarily morally wrong. This makes him morally deficient, since he gets the matter wrong, but it does not mean that he is always wrong about what is morally possible, or even wrong most of the time. Kramer has no reason to insist that moral agents must be morally perfect. All he needs is to insist that they are generally reliable. Mathematicians need not get everything mathematically right in order to be good mathematicians, but they should usually be right.

Ultimately, if there is a problem with Kramer's epistemology, it is not that it implies that moral agents cannot attain moral or metaethical knowledge. The problem would be with Kramer's acceptance of Hume's Law, and Clark's objection does nothing to explicitly argue against it. It therefore has no particular bite against Kramer's epistemology. However, it is still a useful objection to investigate, as it vividly illustrates the two main pillars that Kramer's moral epistemology stands upon. Metaethical claims are substantive moral claims, and any argument for a substantive moral claim must be moral, rather than non-normatively metaphysical, logical, semantic, and so on.[39] The implication is that Hume's Law severely limits the ways in which agents can attain metaethical and moral knowledge, including knowledge of determinate core moral propositions. This renders Kramer's metaethics vulnerable to a new objection from the core-aposteriorist.

---

[39] This raises an interesting question: Hume's Law is a metaethical claim itself, so does that mean that it can only be defended morally? It seems that Kramer has to say yes, as Hume's Law makes the highly abstract substantive moral claim that no moral claim is justified if it cannot be defended morally. This emphasises the point that if you cannot understand moral concepts, you will have no way to attain knowledge of moral claims, so you would be unable to form moral arguments.

In order to develop this new objection, we first have to examine what exactly corresponds to core moral propositions in Kramer's metaethics. It turns out that they are not exactly identical to Kramer's basic moral principles.

6 – The core-aposteriorist's objection

Though he is a thoroughgoing objectivist, Kramer makes it clear that he does not believe that all moral questions must have determinately correct answers (i.e. answers with determinate truth values), because he thinks that (for example) moral vagueness exists and some moral values are incommensurable.[40] If every moral proposition was determinately true or false, there would be some determinate true moral principles that we would have to follow but that we could not possibly know, even in the best circumstances. This is impossible, because otherwise morality would be 'ludicrous'. Furthermore, some moral principles would be true for completely arbitrary reasons, which is also impossible.[41] However, determinate correctness plays a large role in Kramer's discussion of objectivity. Objectivists need not think that all moral questions have a determinately correct answer, but Kramer does think that *most* moral questions do. He aims for 'a vindication of the general determinacy of morality'.[42] Moral propositions that lack determinate truth values therefore form a fairly small subset of possible moral propositions.

Kramer states that basic moral principles alone may not by themselves tell you what to do in particular concrete situations. Just because you know that moral supervenience and moral objectivity are true, that does not mean that you know whether it is permissible to lie or not in a particular situation. So the set of basic principles is only a subset of core moral propositions, because the set of core moral propositions will be able to justify the normative content of all moral propositions that are determinately true. We need something more, and we find it with the normative first-order moral theory that we add to the basic moral principles. The basic moral principles plus the first-order moral theory will collectively justify the normative content of many moral propositions.

We can now see a similarity between Kramer's metaethics and the SLS theory. It turned out that whilst Sidgwick's moral axioms were core moral propositions, they could not be the *only*

---

[40] Kramer (2009), pp. 109-113, 121-122.
[41] Kramer (2009), pp. 111-112.
[42] Kramer (2009), p. 121.

core moral propositions in the theory. They could not by themselves justify the normative content of many moral propositions, as we needed to add the 'other resources' of a normative moral theory in order to do so. For example, in section 3 of chapter 2, we found that *Aim*, as it was originally stated, could not be a determinate core moral proposition. It lacked sufficient determinacy, and was consistent with different types of utilitarianism that could not all be true together. They only way that we could give *Aim* more detail and hence turn it into a determinate core moral proposition was to use the 'other resources' of a particular utilitarianism – say, rule utilitarianism – to alter *Aim* appropriately. The trouble was that these resources appeared to only be ascertainable *a posteriori*. They implied that *Aim* had to take into account what moral agents were like, and that they were capable of doing. Similarly, if we use first-order moral theories to help ascertain how we should morally act, we risk doing so on the basis on *a posteriori* observations. Kramer says that basic moral principles are highly abstract by themselves, so (for example), they do not tell us whether lying to save face in social situations is permissible or not. So if less abstract (that is, more concrete) principles are ascertained via first-order moral theories, and such theories can be influenced by *a posteriori* observations, then our knowledge of the principles is *a posteriori*.

The SLS theory also holds that the first-order theory, not the self-evident axioms, is what provides a theory of the good. Without a theory of the good, we are left utterly in the dark about how to behave, as we do not know what our actions should aim at. Kramer's moral realism is similar, in that it is the first-order moral theory we choose that provides us with the purpose or aim of morality.

There is one further point: we must select the *correct* first-order theory. If a first-order moral theory tells us that it is morally permissible to torture babies for pleasure, then it is morally repugnant, and we should reject it. It is also the case that, thanks to moral objectivity, a first-order moral theory cannot be true for me and false for you. Furthermore, if a moral question has a determinately correct answer, but two first-order theories give different answers to the question, then at least one of the theories is incorrect, otherwise the question does not have a determinately correct answer. So either there is only one correct first-order theory or there are two or more first-order moral theories that differ, but give the same answer to every such question. As there is no practical difference between the theories as far as the core-

aposteriorist's objection is concerned, I shall assume from now on that there is only one correct first-order moral theory.

The core-aposteriorist can now set up the following argument against Kramer's moral realism as a core-apriorist theory:

1) Most moral propositions have determinate truth values.

2) We must be able to know many moral propositions that are determinately true, otherwise we could not know how to morally behave, and that would be morally repugnant.

3) The normative content of all determinately true moral propositions is either justified by the basic moral principles, by the correct first-order moral theory, or by both together.

4) The normative content of all determinately true moral propositions are justified by the set of core moral propositions.

5) Therefore the set of core moral propositions is the set of basic moral principles and the principles of the correct first-order moral theory.

6) The normative content of many determinately true moral propositions cannot be justified by the basic moral principles alone (If they could, Kramer would not need to bring in first-order moral theories.).

7) Many of the moral propositions in (6) are determinate moral propositions, in the sense that the terms within the propositions have fairly precise application conditions.

8) In order to help justify the normative content of these determinate core moral propositions, the first-order moral theory cannot contravene *Sufficient Determinacy* (i.e. the first-order moral theory must provide determinate propositions that can help justify the normative content of determinate moral propositions).

9) It is only possible to select the correct first-order theory that abides by *Sufficient Determinacy* by using *a posteriori* information (This is shown by sections 3 and 4 of chapter 3, which showed how *a posteriori* information was necessary to make axioms such as *Aim* more determinate, and section 7, which showed that theories of the good may require *a posteriori* support.).

10) Therefore the correct first-order theory, in order to help justify the normative content of determinate moral propositions, must itself be (at least partly) expressed by determinate core moral propositions that can only be ascertainable through *a posteriori* or mixed methods.

11) So the set of core moral propositions in Kramer's moral realism must itself include determinate propositions that are only ascertainable through *a posteriori* or mixed methods.

12) Core-apriorist theories deny that this is possible.

13) So Kramer's moral realism cannot be a core-apriorist theory.

Kramer can make the objection that he only said that basic moral principles are *a priori*, not that first-order moral theories must be. However, this will not help this theory as a core-apriorist theory. Take the following proposition:

Drink: Driving when your driving ability is affected by your alcohol intake is always morally impermissible

This is a fairly determinate moral proposition; there may be borderline cases where it is unclear whether someone contravenes *Drink*, but one does not need to drink very much alcohol for it affect one's driving ability. Suppose that *Drink* is true. We cannot infer it simply by appealing to basic moral principles. Instead, we can construct a moral argument for *Drink* by appealing to basic moral principles together with a first-order moral theory. For example, a Kantian might claim that if *Drink* were false, it would contravene the categorical imperative.

If we work out the first-order moral theory we use with *a posteriori* or mixed methods, though, then we cannot say that we work out that *Drink* is true with an *a priori* method. The basic moral principles are not sufficient to imply the normative content of *Drink* on their own (partly because they cannot provide a theory of the good on their own). Kramer cannot use logical or non-normative metaphysical *a priori* arguments to defend the first-order moral theory either, because that would mean using such arguments to defend a moral position. This would contravene Hume's Law.

We have reached a position where the set of core moral propositions include basic moral principles and the principles of the correct first-order theory. The basic moral principles may all be *a priori*, but the correct first-order theory does not seem to be, because if they were they would risk contravening *Sufficient Determinacy*. In order to escape this conclusion, Kramer could try two versions of the same response. This response rejects premise (9) of the above argument, by claiming that it is possible to ascertain the correct first-order moral theory by using *a priori* information alone. Unfortunately, the response fails because Kramer does not tell us what this *a priori* information can be.

The first version of this response denies that the principles of a first-order moral theory could ever be discovered *a posteriori*, so even though a first-order moral theory is not a collection of basic moral principles, it must be justified *a priori* (or be *a priori* self-evident). The second version is to say that the correct first-order moral theory can be ascertained at least in part *a posteriori*, but it can also be ascertained entirely *a priori*, and that this *a priori* method has explanatory precedence over any *a posteriori* method. The second version may be preferable to the first, because it allows that we may be able to work out the correct first-order moral theory by looking at *a posteriori* information about actual moral agents. That is, it allows for multiple methods to attain moral knowledge.

The trouble with these arguments is that they need considerable development to look convincing. Unless we are actually told what this *a priori* method to attain determinate core moral knowledge is, dismissals of *a posteriori* evidence (following the first version) and appeals to explanatory precedence (following the second version) look rather weak. The key problem is that Kramer depends on the decent person's imagination and what the decent person would regard as morally repugnant. He may think that this provides *a priori* information about moral principles, but he does not say very much about how the decent person's imagination works and what *a priori* and *a posteriori* inputs it feeds off to generate moral judgements. Why could it not use the agent's personal experiences as inputs, for example? Since Kramer says very little about what information decent people use to make moral judgements, he leaves it open that they could use *a posteriori* information to ascertain the normative content of moral propositions.

This is not to say that basic moral principles or first-order moral theories are all wrong, or that moral propositions are relativist because their truth values depend on the decent person's

imagination. Saying that we do not actually know a particular moral principle does not mean that the principle is false. Neither is Kramer committed to saying that the fundamental reason why particular moral principles are true is because the decent person's imagination works in a certain way. Kramer could say that the decent person's imagination merely tracks the principles' truth value. It does not decide their truth value. However, Kramer does not give much attention to how the decent person's imagination operates. In the case of basic moral principles, this may not be such a problem, because the principles are supposed to be self-evident. First-order moral theories are not self-evident, though. If we try to claim that one of them is – say, that rule-utilitarianism is – then we can use Sidgwick's criteria for self-evident moral axioms to see how plausible the claim is, and it would be a struggle to defend it. Even if we concede that decent people exist (i.e. that there are people who can identify at least some truly morally repugnant outcomes) and that they can be identified, this says nothing about how they identify morally repugnant outcomes.

A minor issue to note is that we should not think that *a priori* truths are easier to identify than *a posteriori* truths. One argument in favour of Kramer could go as follows: moral truths are generally knowable, and if we cannot know whether a moral principle is true or not, that for Kramer is good evidence that it is not true. If there are too many unknowable moral principles, morality becomes absurd. It is far easier to know true moral propositions *a priori*, as we do not need to go out into the world and conduct difficult and costly experiments that are at the mercy of a thousand complications (Is our measuring equipment working properly? Have we allowed for all the confounding factors?). So there is a *prima facie* case that moral principles are more likely to be knowable *a priori* than *a posteriori*.

This is not a convincing argument, because there is no particular reason why it should be easier to know anything *a priori*. To take an extreme example, what could be easier to discover than the colour of one's skin simply by looking at one's hands? Yet this observation is perceptual, and hence *a posteriori*. On the other hand, it can be hard for a mathematical layman to list the axioms of the arithmetic we use daily, even though this is an *a priori* task concerning something very simple. So we should not be tempted to think that there is a relationship between the *a priori* and how easy it is to attain knowledge.

I have so far been talking about the *a priori* and *a posteriori* status of the first-order moral theory that decent people use, and not about the *a priori* status of basic moral principles. I have therefore left open the possibility that basic moral principles are indeed *a priori* self-evident. The core-aposteriorist could reject this possibility by arguing that at least some of them are not self-evident; instead, decent people identify these principles in the same way that they identify the correct first-order theory, and this involves using *a posteriori* inputs as well as *a priori* inputs to reach a conclusion. This objection is made more difficult by the fact that Kramer never lists the basic moral principles or really gives a method for identifying them; he concentrates mainly on examining the basic moral principles of objectivity and supervenience instead. It is therefore unclear how the core-aposteriorist's objection should go, though the core-aposteriorist could point out that the onus is on Kramer to show that all the basic moral principles are *a priori*. However, the objection also goes further than the core-aposteriorist needs to go, given the above argument (1)-(13) concerning first-order moral theories. Since there are determinate core moral propositions that do not express basic moral principles, we do not need to show that basic moral principles are *a posteriori*. All we need to do is show that some determinate core moral propositions are *a posteriori*. These particular propositions are amongst those used to justify first-order moral theories. This shows that Kramer's metaethics is not a core-apriorist theory, or at least that further argument is needed to show that it is. Therefore Kramer's metaethics cannot help to defend the CAT even if every basic moral principle is knowable *a priori*.

Kramer can argue that the core-aposteriorist's arguments presuppose that knowledge of the normative content of moral propositions is based on contingent facts. If we use *a posteriori* evidence and that evidence could have been different, then the moral principles that the evidence justifies may have been different. But this opens the door to relativism, which is rejected because it would lead to morally repugnant consequences. Therefore the claim that *a posteriori* evidence could be used to ascertain the normative content of a moral proposition is wrong.

We have already seen in section 1 of chapter 2 that Saul Kripke has argued that *a posteriori* metaphysical necessities exist, and that the core-aposteriorist can argue that moral necessities are sufficiently similar to metaphysical necessities so that they too can be *a posteriori*. If the

core-aposteriorist wants to use Kripke against the objection here that *a posteriori* moral propositions must all be contingent, the only thing to add is that we must be careful to keep non-normative metaphysical necessities and moral necessities separate. According to Kramer's metaethics, moral necessities cannot be non-normatively metaphysical, though they may be normatively metaphysical. If they were non-normatively metaphysical, it would be possible to generate a non-normative metaphysical argument for moral necessities, which would contravene Hume's Law. But the core-aposteriorist can accept this condition without difficulty. She can still hold that, say, we may be able to work out the limits of what is morally possible by examining moral agents in the world and how they react in various situations. The fact that torturing babies for pleasure is beyond the limits of any decent person's imagination is not something that can be ascertained *a priori*. To say that it is, we have to work out who the decent people are and on what grounds they reach the conclusion that torturing babies for pleasure is morally impermissible. It may turn out that the grounds are actually *a posteriori*. But just because decent people may use *a posteriori* information to work out normative conclusions, it does not entail that the conclusions that they reach are morally contingent. It does not imply that in a different morally possible world, decent people could have had imaginations that would lead them to think that it was morally permissible to torture babies for pleasure. Similarly, the fact that we discovered *a posteriori* that water is H_2O does not mean that it is metaphysically possible that water could have been something else. We might have called a different substance 'water', but that does not mean that the different substance would have been water.

One reason why the core-aposteriorist can argue in this fashion is because of the limits of what is possible. Kramer excludes all types of possibility except moral possibilities. Some types of possibility, such as logical possibility, appear to be knowable only *a priori*. It is usually thought that in order to work out if something is logically possible, you do not need to examine anything in the outside world. In classical logic, nothing can be A and not-A at the same time, and you do not need to look at any actual object to find this out. At the other end of the scale, there is physical possibility. If we define 'physically possible' as 'conforms to all the laws of physics as they apply in our own universe', we can often find out if something is physically possible or not based on *a posteriori* investigation. Could I fly if I jumped off a cliff? It is logically possible, but if we are right about the laws of physics (discovered through *a posteriori* research), then it is not

physically possible. Kripke tells us that metaphysical possibility lies between these two extremes; many metaphysical possibilities are discoverable *a priori*, but not all of them are. Why assume, then, that moral possibility has to lie at one extreme with logical possibility? If other types of possibility lie at various points along the scale between *a priori* knowledge and *a posteriori* knowledge, why can moral possibility not be one of them?

The core-apriorist may still insist that some moral necessities must be knowable *a priori*, but this is not enough. So long as some determinate core moral propositions express moral necessities that are only knowable *a posteriori*, the CAT fails. Kramer cannot completely isolate his metaethics from *a posteriori* evidence, partly because he has so severely limited the ways in which we can argue for moral principles. If the only way we can argue morally for moral principles is by using *a posteriori* evidence, then Kramer cannot switch to other forms of non-moral argument (logical, non-normatively metaphysical, etc.) to avoid the use of such evidence.

In conclusion, the core-aposteriorist's objection against Kramer is rather modest. The objection does not deny that basic moral principles exist, or that we can know them. All it does is point out that some of the determinate core moral propositions produced by Kramer's metaethics are unlikely to be knowable *a priori*. We can only know what these propositions are by examining what decent people would think about them, but Kramer says very little about how decent people actually think. He leaves open the possibility that they use *a posteriori* evidence to make moral judgements.

7 – Moral fixed points

The core-apriorist could claim that the moral principles we hold as most important can be justified entirely *a priori*, and so *a posteriori* evidence will only help to justify more minor moral principles. By 'most important', I mean the moral principles that most people actually believe have the strongest moral hold on us, or that are absolutely inviolable moral principles that nobody with even a trace of moral sensibility would transgress. Actual agents may still be morally good in spite of occasionally telling white lies, or feeling brief pleasure at a rival's embarrassment, but nobody morally good would torture babies for pleasure or become a serial killer. Anyone who thinks that it was morally permissible to be a serial killer would at best have a

moral system so bizarre that it may not really be a genuine moral system. The core-apriorist could claim that there is something 'special' about such principles such that we can only know them *a priori*.

If this is correct, it does not show that Kramer's metaethics vindicates the core-apriorist, because there still remain other true core moral proposition that are 'less important', and these propositions may only be ascertainable *a posteriori*. However, we should investigate this possibility, since it places limits on the set of moral propositions that can only be ascertained *a posteriori*, and it also limits the importance of the propositions in the set. The most important moral propositions are ascertainable *a priori*.

The core-apriorist needs to tell us what is 'special' about the most important moral principles such that they are knowable *a priori*. Terence Cuneo and Russ Shafer-Landau call these principles 'moral fixed points', and they propose that these principles are conceptual truths. The principles are conceptually necessarily true, and therefore they are knowable *a priori*. This differs from Kramer, as Kramer talks of moral and not conceptual necessity. However, I will ignore this difference because I will argue in this section that Cuneo and Shafer-Landau do not show that the most important moral principles are knowable *a priori*. Some of them may be, but there is no necessary connection between importance and the *a priori*.

A *minimally eccentric* moral system is 'a reasonably comprehensive and consistent body of moral propositions that apply to beings like us in a world such as ours'.[43] Any such system will include moral fixed points, such as:

- It is pro tanto wrong to engage in the recreational slaughter of a fellow person.
- It is pro tanto wrong to torture others just because they have inconvenienced you.
- There is some moral reason to offer aid to those in distress, if such aid is very easily given and comes at very little expense.

Cuneo and Shafer-Landau provide a longer list of moral fixed points, which they explicitly state is incomplete. The moral fixed points form the boundaries of the moral system, and if a moral

---

[43] Cuneo and Shafer-Landau (2014), p. 404.

theory goes outside these boundaries (say, by sometimes allowing the recreational slaughter of fellow people), then it must be wrong.[44]

Moral fixed points are conceptual truths. For example, take the first fixed point above. It belongs to the essence of the concept 'being wrong' that it applies to all and only those things that are wrong. If I killed a fellow person for recreation, my action would satisfy the concept 'recreational slaughter'. The concept 'being wrong' necessarily applies to recreational slaughter, so it is a conceptual truth that the recreational slaughter of a fellow person is wrong.[45]

Cuneo and Shafer-Landau do not give complete criteria for what propositions express conceptual truths, but they do provide a partial list. For proposition P to express a conceptual truth:

1) P must be necessarily true.

2) P must help fix the boundaries of the subject matter. For example, if the proposition 'God is a perfect being' fixes the boundaries of theological debate, then anyone rejecting that proposition would not be employing the theistic concept 'God'.

3) Denying that P would bewilder anyone competent with the constituent components of P. '[I]ts denial would be almost crazy.'

4) P is knowable *a priori*.[46]

Unfortunately, Cuneo and Shafer-Landau do not discuss these criteria in detail, so it is not always clear how we are supposed to interpret them. When Cuneo and Shafer-Landau say that P must be necessarily true, they cannot mean that P is conceptually necessary, because otherwise criterion (1) would say that conceptual truths must be conceptually necessary truths, which would be redundant. But then what necessity is relevant here? It may not be logical necessity, as they do not argue that moral fixed points, being conceptual truths, are logically necessary. It is also up for debate whether 'God is a perfect being' is logically necessary, if it is true.

Perhaps they mean metaphysical necessity. As we saw in the previous section, we can argue that certain metaphysically necessary truths are only knowable *a posteriori*, so only a

---

[44] Cuneo and Shafer-Landau (2014), p. 404.
[45] Cuneo and Shafer-Landau (2014), p. 410.
[46] Cuneo and Shafer-Landau (2014), pp. 407-408.

subset of metaphysically necessary truths can be conceptual truths, because of criterion (4). But this question will have to be left open for the moment.

Criteria (2) and (3) bring their own problems. Cuneo and Shafer-Landau's example of 'God is a perfect being' is not well-chosen. It is entirely reasonable to think that two people can have a coherent discussion about, say, God's relationship to man without assuming either that God is a perfect being or that God is not a perfect being. It may have no bearing on the discussion, and it seems difficult to hold that in this case, they are not really talking about the same thing if they would disagree on whether God is a perfect being.

The fact that the example is not well-chosen does not mean that there are no better examples. If someone denies that 3 is the whole number between 2 and 4, he not only rejects something that any mathematically competent person believes about the number 3, but the denial will infect practically everything he says or believes about the number. Even very basic arithmetic will produce wildly different results. The point, though, is that we have to be careful about what propositions actually fulfil criteria (2) and (3). It may be that 'God is a perfect being' satisfies criteria (2) and (3) when the people engaged in theological debate are theological experts.[47] If two theological experts debate God's attributes, they may implicitly assume that God is a perfect being, because within the group of experts this is taken as a necessary assumption for debate about the same thing to be possible. Maybe an expert would be reasonably expected to know this. One theologian would thus be bemused if another theologian denied that God was a perfect being. He may decide that he and the other theologian were not talking about the same thing. But if he met a layman who denied that God was a perfect being, would he assume that they were not talking about the same thing? Would he be just as bewildered by the denial? This is unlikely. More probably, he would think that the layman was just ill-informed.

So criteria (2) and (3) have different interpretations. The first is the 'layman' interpretation, which fits the mathematics example. You do not need to be a mathematical expert to know that 3 is the whole number between 2 and 4. If anyone denied this, whatever their level of expertise, we could reasonably claim that they were not using the concept '3' like everyone else does. The second interpretation is the 'expert' interpretation, which fits the theological example. If a

---

[47] I leave aside here the question of whether one actually needs to accept that God is a perfect being in order to be a theological expert. I am just assuming that one does for the sake of the example.

theological expert denied that God was a perfect being, he would be using different concepts from other theologians, and the other theologians would be bewildered by the denial.

The 'layman' interpretation is more plausible for moral fixed points. It does not take a lot of moral competency to know that the recreational slaughter of fellow people is morally wrong, so even if someone with very limited understanding claims otherwise, we would hold that he is not using the same moral concepts that we do.

Criterion (4) is the most important criterion as far as the core-aposteriorist is concerned. She claims that there is no necessary connection between *a priori* moral propositions and propositions that express 'important' moral truths. In other words, at least some propositions that express what we think are moral fixed points, and that help fix the boundaries of what is morally possible, are only knowable *a posteriori*, and so do not meet criterion (4). These moral fixed points are not moral conceptual truths. To make this argument, I will start with a different objection that Cuneo and Shafer-Landau explicitly address. Their response to it leaves them open to a new objection by the core-aposteriorist.

The objection that Cuneo and Shafer-Landau discuss runs as follows: moral fixed points concern beings like us in worlds such as ours. So moral fixed points must presuppose information about human psychology. But '[c]onceptual truths are… supposed to be true in virtue of the essence of their constituent concepts', and moral concepts do not imply anything about human psychology. The concept 'being wrong' carries no psychological implications. Therefore moral fixed points are not conceptual truths.[48]

Cuneo and Shafer-Landau's response to this objection concerns immediate and mediate conceptual essences. It is not entirely clear what they take an essence to be. They base their discussion on work by Kit Fine, who claims that

'[i]t is of the immediate nature, or essence of singleton Socrates [the set that contains only Socrates] to contain Socrates and of the immediate nature of Socrates to be a man, but it is only of the mediate nature of singleton Socrates to contain something that is a man. In general, the mediate nature of an object will be subject to chaining: the nature of any object (ineliminably) involved in its nature will also be in its nature. The immediate nature, by

---

[48] Cuneo and Shafer-Landau (2014), p. 433.

contrast, will include only what has a direct bearing on the object, excluding what derives from the nature of the other objects.'[49]

Cuneo and Shafer-Landau are well aware that Fine's use of terms such as 'immediate essence' is different from theirs. According to Fine, it is objects (such as sets and people) that have essences. If Y is an essential property of X, claims Fine, then X exists only if Y does. Cuneo and Shafer-Landau, however, talk of concepts.[50] Since they do not define what they mean by the essence of a concept, this raises the question of how we are meant to understand their comparison with Fine. Is it only an analogy? Is the essence of an object supposed to be related in some way to the essence of a concept? The problem of what the essence of a concept is meant to be remains unanswered.

The examples that Cuneo and Shafer-Landau use help, even though they do not define 'essence'. Cuneo and Shafer-Landau claim that 'it belongs to the immediate essence of the concept 'human being' that, if it applies to anything, then it applies to material objects'.[51] The essence of a concept, then, may be simply what the concept necessarily applies to. It is of the *immediate* essence of the concept 'human being' that it necessarily applies to material objects. If anything is a human being, it is a material object. It is also of the immediate essence of the concept 'material object' that it necessarily applies to things that are spatially extended. So it is of the *mediate* essence of the concept 'human being' that it necessarily applies to certain things that are spatially extended.

Consider the moral fixed point 'Recreational slaughter is wrong'. The immediate essence of the concept 'being wrong' does not contain any information about beings like us, but it does apply necessarily to the behaviour of rational agents (It is hard to see how it could apply to the behaviour of trees or spiders.). The immediate essence of the concept 'rational agent' necessarily applies to intentional agents, who meet certain criteria. Humans are necessarily intentional agents, so the immediate essence of the concept 'intentional agent' must apply to human beings. Consequently, it is of the mediate essence of the concept 'being wrong' that it necessarily applies to the behaviour of human beings. In effect, we go down a chain of concepts ('being wrong', 'rational agent' and 'intentional agent') that are linked by immediate essences.

---

[49] Fine (1985), p. 281. See also pp. 276-280.
[50] Cuneo and Shafer-Landau (2014), p. 434.
[51] Cuneo and Shafer-Landau (2014), p. 434.

The moral fixed point 'Recreational slaughter is wrong' is a mediate conceptual truth, since the mediate essence of the concept 'being wrong' applies to the behaviour of human beings, and such behaviour can include recreational slaughter. Go down the chain of concepts starting with the concept 'being wrong', and you will eventually reach concepts that concern human psychology, even though the concept 'being wrong' does not immediately do so.[52]

The core-aposteriorist can now start applying pressure to Cuneo and Shafer-Landau's account of moral fixed points. First, what precisely is it that the concept 'being wrong' applies to? Rational agents can behave in all sorts of ways, many of which have no particular moral implications. If the moral fixed points are conceptual truths, the essence of the concept 'being wrong' has to apply to a very particular type of behaviour (Cuneo and Shafer-Landau do not define what they mean by 'behaviour', which makes it more vague what precisely the concept 'being wrong' applies to. Behaviour may include different things according to different moral theories, such as the agent's intentions, motives and character.). What exactly is it about our behaviour that the essence of the concept 'being wrong' applies to?

At this point, the *a posteriori* strategy comes into play. The core-aposteriorist can try to claim that either moral conceptual truths contravene *Sufficient Determinacy*, or they are not conceptual truths as Cuneo and Shafer-Landau define conceptual truths. The result is that at least some moral fixed points are expressed by determinate moral propositions that are not knowable *a priori*. Therefore some moral fixed points are not conceptual truths.

Cuneo and Shafer-Landau tell us that the essence of the concept 'being wrong' applies to recreational slaughter. It is therefore a conceptual truth that recreational slaughter is wrong, since the truth satisfies criteria (1)-(4) for being a conceptual truth. Since it is a moral conceptual truth, it is a moral fixed point. We can thus use the following propositions as expressions of moral fixed points:

Slaughter: Recreational slaughter of fellow people is morally wrong

Wrong: Certain types of human behaviour are morally wrong

---

[52] Cuneo and Shafer-Landau (2014), pp. 433-435.

Wrong expresses a moral conceptual truth if *Slaughter* does, and therefore they both express moral fixed points according to Cuneo and Shafer-Landau. But *Wrong* is utterly hopeless as a determinate moral proposition. *Slaughter* is more successful, but we need to know much more about what exactly it is about recreational slaughter that is so morally terrible. We cannot find out simply by examining *Wrong*. Furthermore, the fact that it is obvious that *Slaughter* is true is irrelevant. The obviousness of a fact need not imply anything about the reasons why it is true, or how agents come to know it.

Assuming that *Slaughter* is true, we can go back to the criteria (1)-(4) for being a conceptual truth. Does *Slaughter* satisfy them? Let us be as generous as possible here. Let us assume that criterion (1) is satisfied if a truth is metaphysically necessary, and that *Slaughter* is metaphysically necessary. Let us also assume that anyone who denied that *Slaughter* was true would not be using the term 'morally wrong' as morally competent agents do, and that morally competent agents would be bewildered by the denial. Anyone who knows anything whatsoever about what is morally right or wrong would accept *Slaughter* as true. No great moral expertise is needed.

Does *Slaughter* satisfy criterion (4)? It seems that the core-aposteriorist can say no. If we take *Slaughter* as metaphysically necessary, then we can agree with Kripke that it may only be knowable *a posteriori*. Nor does there seem to be any necessary connection between satisfying criteria (2) and (3) on the one hand, and satisfying criterion (4) on the other. So why would we be forced to say that *Slaughter* is knowable *a priori*? We can flesh this point out a little by making a comparison between two particular moral systems. One is minimally eccentric, and the other is what I will call maximally eccentric. Cuneo and Shafer-Landau say that minimally eccentric moral systems 'apply to beings like us in a world such as ours'. They provide a list of criteria for 'beings like us' (which include criteria of having similar biology, psychology and rationality to us) and 'worlds such as ours' (which include 'the laws of nature and broad empirical statistical generalizations that are at least close to the ones that obtain here on earth').[53] We can conclude from these criteria that it would be possible for there to be beings not like us on worlds not like ours, who would not have a minimally eccentric moral system, relative to our own moral systems.

---

[53] Cuneo and Shafer-Landau (2014), p. 404.

Importantly, it must be possible that such beings could nevertheless have a moral system. It is an open question how different a moral system would have to be from ours for it to fail to be a genuine moral system. The thought runs as follows: moral systems, in order to be moral systems, have to fulfil certain criteria. Our moral systems on Earth meet this standard. If a moral system was wildly unlike our own moral system, it would not meet this standard, and so therefore it would not be a genuine moral system.

One way of developing this thought is what Gopal Sreenivasan calls the Hurley-Davidson argument.[54] The strong version of the argument claims that there is no possibility of a genuine moral system which provides moral values entirely incommensurable with ours. The two different moral systems have to share similar criteria, and if a moral system is incommensurable with ours, it does not meet those criteria. So if massively different beings from us cannot share moral values with us, then they cannot have a genuine moral system. The weaker version of the argument claims that if there is such a moral system, we could never come to understand it, because we would not be able to agree with it, and therefore we would not be able to translate it into our own moral language so that it matches our own values.[55]

Sreenivasan rejects both versions of the argument by claiming that in order to understand a moral system, we do not need to actually endorse it, and we do not need to be able to translate it into our own moral language either. Fortunately, I do not need to go into details because Cuneo and Shafer-Landau have not actually set up moral fixed points in a way that makes them vulnerable to the Hurley-Davidson argument. Though they leave open the possibility that there are moral systems with moral fixed points that are different from the ones that we human beings actually recognise, they do not go so far as saying that there are moral systems with values entirely incommensurable with ours. This allows us to talk of maximally eccentric moral systems, which I define as moral systems that are as different from our own moral systems as it is possible to be whilst still being genuine moral systems.

Cuneo and Shafer-Landau are inclined to believe that the moral fixed points are the same across every possible world. Is it even possible that a genuine moral system could exist in which recreational slaughter would be morally permissible? However, they are reluctant to insist

---

[54] The argument is named after Susan Hurley ((1989), p. 53) and Donald Davidson (1984).
[55] Sreenivasan (2011), pp. 3, 11, 24.

on this point, as they believe that it relies on modal intuitions that their critics may not share.[56] So they leave open two alternatives: if the moral fixed points are the same in every genuine moral system, then we can drop all talk of minimally and maximally eccentric moral systems. On the other hand, if even one moral fixed point can change between moral systems, then we must continue using the concept of eccentric moral systems.

This reveals a complication for Cuneo and Shafer-Landau. The specific moral fixed points that they use as examples are the fixed points for minimally eccentric moral systems, and it seems possible for them to be different in maximally eccentric moral systems. But moral fixed points are supposed to be moral conceptual truths. How can conceptual truths be different in different possible worlds, and on what basis? What has happened here is that we have not been sufficiently precise in setting out the moral fixed points. Rather than saying that

a) It is pro tanto wrong to engage in the recreational slaughter of a fellow person

is a moral fixed point, we should strictly speaking say that

b) For beings like us in worlds such as ours, it is pro tanto wrong to engage in the recreational slaughter of a fellow person

is a moral fixed point.[57] (b) can be true even if maximally eccentric moral systems do not hold that the recreational slaughter of a fellow person is wrong (Perhaps, as Cuneo and Shafer-Landau suggest, there are possible worlds where people who are killed spontaneously regenerate soon afterwards. Would killing be so morally horrible or impermissible in such a world? Our opinions are more likely to diverge on this question.). Whenever Cuneo and Shafer-Landau claim that (a) is a moral conceptual truth, it is just shorthand for saying that (b) is a moral conceptual truth, and this is true for all moral fixed points that they identify. We can ask whether a form of relativism does not start sneaking into their metaethics here, but the core-aposteriorist does not have to pursue this matter. The objection to Cuneo and Shafer-Landau does not depend on relativism.

---

[56] Cuneo and Shafer-Landau (2014), p. 405, fn 15.
[57] Cuneo and Shafer-Landau (2014), p. 405.

Suppose we meet people with a maximally eccentric moral system. For convenience, I will call such people 'aliens' and us 'humans', though this just means that they have a maximally eccentric moral system. Since they do, they would recognise different moral fixed points. The core-aposteriorist's objection is that we can only find out what the moral fixed points are for humans, and for aliens, by investigating the humans, the aliens and the worlds in which they live. This investigation requires *a posteriori* observations, so our knowledge of the moral fixed points that apply to us (or to the aliens) must be *a posteriori*.

What this means, in effect, is that somewhere along the chain of concepts *a posteriori* information can creep in. The immediate essence of the concept 'being wrong' applies to the behaviour of rational agents, fair enough. We can even concede that we discover *a priori* that this is so. But that only gets us as far as *Wrong*. It tells us nothing about what sort of behaviour is morally wrong. *Wrong* is not determinate, so it cannot justify the normative content of determinate moral propositions such as *Slaughter*. Furthermore, specific behaviour that is morally wrong may change its moral value if we switch from looking at humans to looking at aliens. We cannot just assume that the morally relevant differences between humans and aliens (or between the worlds that they inhabit) are discoverable *a priori*. The facts that *Slaughter* is obvious and that we humans would be bemused by agents who deny it are not sufficient evidence that knowledge of *Slaughter* is *a priori*. Consequently, it is possible for a proposition to satisfy criteria (1)-(3), but fail to meet criterion (4). This means that, if we stay with Cuneo and Shafer-Landau's definitions, moral fixed points that are expressed by determinate propositions may not be moral conceptual truths.

We may continue to insist that such moral fixed points really are moral fixed points, though they are not moral conceptual truths. Alternatively, we might decide that they are not moral fixed points after all. But does it really matter which option we pick? Assume that we pick the second, and call moral truths that fulfil criteria (1)-(3) but not (4) 'moral quasi-fixed points'. Moral quasi-fixed points appear to do everything that moral fixed points do, with the only difference being that we cannot discover them *a priori*. We must do so *a posteriori*. What reason, then, do we have to believe that there are moral fixed points, rather than just moral quasi-fixed points? None, claims the core-aposteriorist. It is up to the core-apriorist to show that there is one. For example, perhaps the core-apriorist can show that if a proposition cannot meet criterion (4),

then it cannot meet criterion (1). However, continues the core-aposteriorist, we currently have no particular reason to think that any advantage exists.

This objection to Cuneo and Shafer-Landau is another example of the *a posteriori* strategy at work. It may be ascertainable *a priori* that the concept 'being wrong' necessarily applies to human behaviour, but this cannot be expressed as a determinate core moral proposition, because we have no idea which types of human behaviour it applies to. It contravenes *Sufficient Determinacy*. But if we give more detail to the concept, we have to look at human behaviour and the world in which humans live, and we have no assurance that we can work out *a priori* from our investigations what is normatively relevant about particular human behaviour. We may end up deciding that recreational slaughter is morally wrong based on *a posteriori* observations concerning human beings. This may provide us with determinate core moral propositions, but the propositions will only be knowable via an *a posteriori* or mixed method. We cannot obtain *a priori* determinate core moral propositions.

Kramer claims that Hume's Law holds and that metaethical principles are moral principles, but he can reasonably deny that this implies that metaethical theories collapse into first-order moral theories. Whilst his claim that metaethical theories can only be shown as true though moral argument seems to lead to Clark's 'horrified assertion' objection, he has a viable defence against it. The trouble lies elsewhere, in the connection between basic moral principles and first-order normative moral theories. He does not do enough to show that moral agents can attain determinate core moral knowledge *a priori*, and so his metaethics fails to support the CAT.

After examining Kramer's metaethics and the SLS theory, we have found that core-apriorist theories that postulate *a priori* self-evident axioms or principles can suffer from a troubling problem. Either the candidate core moral propositions in the theories are not determinate, or they turn out to use *a posteriori* information to justify the normative content of determinate moral propositions. Can coherentist core-apriorist theories do any better? In the next chapter, we will examine two such theories: Michael Smith's moral realism and moral functionalism. Unfortunately for the core-apriorist, switching to moral coherentism does not help.

Chapter 5 – Moral coherentist core-apriorist theories

Both Michael Smith's moral realism and moral functionalism are moral realist theories, but they differ a fair amount in their details. Smith holds that there is a set of *a priori* moral platitudes that we must be disposed to endorse if we are to successfully master the correct use of moral terms. Whilst Smith does not define himself as a core-apriorist, the core-apriorist can read Smith's *a priori* platitudes as *a priori* determinate core moral propositions. The core-aposteriorist seeks to object that some of these platitudes are in fact only knowable *a posteriori*, and so Smith's moral realism cannot be a core-apriorist theory. Moral functionalism makes more use of *a posteriori* evidence than Smith does, as it claims that we can identify moral properties with particular descriptive properties by using *a posteriori* evidence. However, moral functionalism also purports to rely on *a priori* propositions, as it claims more generally that moral properties are identical to whatever it is that play particular moral roles, and this identification is made *a priori*. The core-aposteriorist's objection to moral functionalism is that these *a priori* propositions are actually *a posteriori*. The core-aposteriorist thus takes the following approach to both theories: she does not aim to disprove them, but seeks instead to show that they cannot be genuine examples of core-apriorist theories. In sections 1-3 of this chapter, I will look at Smith's moral realism in more detail. In sections 4-7, I will examine moral functionalism. In section 8, I will draw attention to an interesting feature that both theories share with the SLS theory and Kramer's metaethics, and I will discuss why they do so.

1 – Moral and colour platitudes

Smith's moral realism appeals to moral platitudes, which may be either *a priori* or *a posteriori*. Platitudes are claims that are platitudinous, in that they are claims that 'capture the inferential and judgemental dispositions *vis-à-vis* [particular terms] of those who have mastery of [those terms]… To have mastery of [those terms] is to be disposed to make inferences and judgements along these lines'.[1] Clearly this idea needs much more development, and in his book *The Moral Problem* Smith goes into detail about platitudes not only as they relate to moral

---

[1] Smith (1994), p. 30.

terms, but also as they relate to colour terms. He uses colours as a way to explain generally what platitudes are before moving on to moral platitudes in particular.

Colour platitudes are platitudes that people with normal colour vision accept. They include platitudes such as 'Most everything we see looks coloured', 'Red is more similar to orange than to blue' and 'Unperceived objects are still coloured'.[2] The obviousness of these platitudes is not the point.[3] We can only have mastery of colour terms such as 'red' if we are 'disposed to make inferences and judgements along [platitudinous] lines'. The platitudes might turn out to be 'incompatible or inconsistent with each other', so they do not necessarily need to be true, but we would need strong *a priori* reasons to give them up. The platitudes themselves are knowable *a priori*, and giving them up means giving up on using colour terms at all.

Mention of incompatibility and inconsistency brings up an important point. For Smith, to master a colour term, an agent must be disposed to accept a 'maximal consistent set of platitudes constitutive of mastery of the [colour term]'.[4] Since the set is maximal consistent, there is no platitude that can be added to the set that does not make it inconsistent. Furthermore, 'it would take something like… inconsistency' to make us give up the platitudes, as they are *a priori*. We can discover *a priori* whether they are collectively inconsistent or not. Examining the world *a posteriori* would not be necessary to discover the platitudes necessary for mastering colour terms.[5]

By analogy, we attain mastery of moral terms through our grasp of moral platitudes, which are also knowable *a priori*. Being disposed to endorse the platitudes in the set gives an agent mastery of moral terms. Smith is a moral coherentist, which means that the moral platitudes within a set are justified by other propositions within that set. It would take 'something like… inconsistency' to make us change the platitudes we include in the maximal consistent set of platitudes necessary for mastering moral terms. Smith divides the moral platitudes necessary for mastery of moral terms into five types:

1) Practical (e.g. 'Weakness of will, compulsion, depression and the like may explain why someone isn't moved in accordance with their moral beliefs')

---

[2] Smith (1994), pp 29-30.
[3] See also Smith (1994), p. 187.
[4] Smith (1994), p. 31.
[5] Smith (1994), p. 30.

2) Objectivity (e.g. 'When A says that ɸ-ing is right, and B says that ɸ-ing is not right, then at most one of A and B is correct')

3) Supervenience (e.g. 'Acts with the same ordinary everyday non-moral features must have the same moral features as well')

4) Substantial (e.g. 'Right acts are in some way expressive of equal concern and respect')

5) Procedural (e.g. 'We can work out which acts are right through a process of reflective equilibrium')[6]

If an agent is disposed to make moral judgements that are platitudinous, then she has mastery of moral terms. Smith claims that if an analysis of moral terms does not account for these platitudes, then it cannot be correct. It does not provide correct information about the concepts that the moral terms denote. This claim enables him to criticise reductive natural moral realism, which claims that moral properties can be defined solely in terms of non-moral natural properties. Forms of reductive naturalism that are subjectivist cannot account for platitudes of moral objectivity, and so must be dismissed. Forms of reductive naturalism that are objectivist cannot account for the platitude that we are often motivated by our moral beliefs, and so must also be dismissed. Therefore, Smith concludes, reductive naturalism is in trouble.[7]

It should be noted that whilst Smith allows for the existence of *a posteriori* moral platitudes, he holds that the moral platitudes that are necessary for mastering moral terms are knowable *a priori*. This allows us to interpret his theory as a core-apriorist theory, though Smith does not explicitly do so himself. To master moral terms, we need to know how to use them and in which circumstances we can correctly apply them. We master moral terms if we are disposed to accept certain *a priori* moral platitudes. The core-apriorist claims that these *a priori* moral platitudes are core moral propositions, and further states that these *a priori* moral platitudes (core moral propositions) are necessary for agents to master moral terms, and that agents need to master moral terms in order to ascertain the truth of the normative content of non-core moral propositions.

The core-aposteriorist's objection to Smith starts with the assertion that if this is so, some of the *a priori* moral platitudes must be determinate. First, some of the non-core moral propositions

---

[6] Smith (1994), pp. 39-40.
[7] Smith (1994), pp. 41-43.

are determinate. Second, in order to ascertain the truth of the normative content of these determinate non-core moral propositions, agents need to master moral terms. Third, the *a priori* moral platitudes, being core moral propositions, are necessary for the mastery of moral terms. Therefore, agents can use the *a priori* moral platitudes to ascertain the normative content of determinate non-core moral propositions. In other words, agents can use core moral propositions to ascertain the normative content of determinate non-core moral propositions. However, continues the core-aposteriorist, if all core moral propositions are non-determinate, this contravenes *Sufficient Determinacy*. Therefore some core moral propositions must be determinate, and so some *a priori* moral platitudes must be determinate if Smith's moral realism is a core-apriorist theory.

This result allows the core-aposteriorist to criticise the *a priori* platitudes that Smith identifies as necessary for the mastery of moral terms in three ways:

a) The moral platitudes are false.

b) The moral platitudes are non-determinate. Therefore, grasping them does not entail that an agent can attain knowledge of true determinate non-core moral propositions, otherwise it would contravene *Sufficient Determinacy*.

c) The moral platitudes are really only knowable *a posteriori*.

Option (a) would be a tricky line for the core-aposteriorist to take. In order to have moral knowledge, we have to be able to use moral terms, and we can only use moral terms correctly if we know how to apply them correctly. It is plausible that we have to accept some claims about these terms that Smith would identify as platitudes if we are able to use moral terms at all. The core-aposteriorist can object that Smith does not identify the correct moral platitudes for mastery of moral terms, but there are major obstacles to this. First, claiming that the platitudes about supervenience in particular are false risks refuting many core-aposteriorist theories as well. If moral properties do not supervene on natural properties, then there would be no way to investigate natural properties *a posteriori* in order to discover moral properties. The core-aposteriorist may well not want this result, given that she believes that we ascertain the truth of the normative content of some determinate core moral propositions using *a posteriori* evidence.

As for the other types of moral platitudes, the core-aposteriorist may not want to reject them either, depending on the content of her own metaethics. Second, the task of proving the platitudes wrong is made more difficult by the fact that the platitudes collectively have very little in common other than that they are *a priori* propositions about morality. Say that they all have characteristic X, where X is not the characteristic of being *a priori* or being a proposition about morality. If the core-aposteriorist could show that we should reject all platitudes with characteristic X because, say, platitudes with characteristic X are logically incoherent, this would be a simple and direct way to pursue option (a). But what would this characteristic be? The moral platitudes that Smith identifies are split into five types, and these types make claims about different aspects of morality (For example, substantive platitudes say nothing directly about supervenience.). Furthermore, the core-aposteriorist can hardly argue that the platitudes are false just because they are *a priori*. So the core-aposteriorist will have to laboriously go through each of the five types of moral platitudes to show that all the relevant platitudes are false.

As option (a) looks unpromising, the core-aposteriorist is left with options (b) and (c). In other words, the core-aposteriorist can try using the *a posteriori* strategy in order to show that Smith's moral realism cannot be used to defend the core-apriorist. Option (c) appears to be the best option, because if we pick option (b), we have to show that *all* the moral platitudes that Smith believes are necessary for the mastery of moral terms are non-determinate, and so they cannot justify the normative content of determinate non-core moral propositions. It is in fact very difficult to argue for this, as some of the platitudes are clearly determinate. Take the platitude that Smith gives as an example of a platitude about moral objectivity: 'When A says that φ-ing is right, and B says that φ-ing is not right, then at most one of A and B is correct'. Smith does not say whether moral propositions with indeterminate truth values exist, but even if they do, they would be rare compared to the number of moral propositions with determinate truth values. The platitude thus appears to be a determinate moral proposition.[8]

If we choose option (c), though, we only have to show that *at least one* of the moral platitudes is only knowable *a posteriori*. If we can do this, this immediately shows that the

---

[8] It should be noted that calling platitudes of objectivity *moral* platitudes can be controversial. We have seen that Matthew Kramer thinks that principles of objectivity are moral principles, but he does so by blurring the lines between metaethics and first-order ethics a great deal. If we want to keep the areas firmly separate, then we may believe that only Smith's substantial moral platitudes are truly moral platitudes, rather than just metaethical platitudes. However, the core-aposteriorist's objection to Smith in this chapter will follow a different path, so I will not pursue this thought.

Central Claim is correct given the moral platitudes that Smith endorses, and so his moral realism is not a core-apriorist theory. Option (c), then, is the best choice for the core-aposteriorist to take. My intention in section 3 of this chapter is to show that many of the purportedly *a priori* moral platitudes that Smith uses are actually only knowable *a posteriori*. Since these platitudes are still necessary for the mastery of moral terms, it entails that if the platitudes necessary for the mastery of moral terms are core moral propositions, some core moral propositions are only knowable *a posteriori*. Furthermore, since some of these *a posteriori* core moral propositions are determinate, Smith's moral realism cannot be a core-apriorist theory.

Before making this argument, though, we can first ask whether the colour analogy that Smith uses is a good one. The analogy breaks down at the point where Smith claims that in order to attain mastery of colour terms, an agent requires perceptual experience of colour. This by itself is not enough to cause Smith any problems, but it points to an important commitment that Smith has concerning moral motivation. If we reject the commitment, we reject his moral realism. The core-aposteriorist may try doing so, but unfortunately her ability to criticise Smith on this ground depends on her own particular metaethics. However, if the core-aposteriorist uses the *a posteriori* strategy, and pursues option (c) instead, she will not have this problem.

2 – The colour analogy

Smith says of colour platitudes:

> '[The platitudes governing the use of the term 'red'] have a *prima facie a priori* status, and gain *a priori* status *simpliciter* by surviving as part of the maximal consistent set of platitudes constitutive of mastery of the term 'red'.'[9]

With this in mind, we can compare colour platitudes to moral platitudes. Certain colour/moral platitudes are necessary for mastery of colour/moral terms, and these platitudes must be *a priori*. There are *a posteriori* colour/moral platitudes, but such platitudes are not relevant. Take,

---

[9] Smith (1994), p. 31.

for example, the *a posteriori* colour platitude 'Red is the colour of blood'. This platitude is obvious and widely believed, but it is 'still [a] contingent and *a posteriori* truth… about the colours, a [truth] whose rejection, in relevant circumstances, would be neither here nor there with respect to whether or not we possess and are capable of using colour concepts'.[10]

To clear up a possible misunderstanding, Smith does not require agents to explicitly know all the platitudes necessary for the mastery of colour/moral terms in order to master them. All he requires is that agents are disposed to make inferences and judgements in line with the platitudes.[11] Intuitively, this seems right; young children can master colour terms even though they would be hard-pressed to list all the platitudes that they make colour judgements with.

Despite the fact that we need not know any particular *a posteriori* colour platitudes in order to master colour terms, such as 'The clear sky is blue' or 'Ripe tomatoes are red', agents cannot master colour terms simply by working out the necessary platitudes *a priori*. Smith states that

'we learn colour terms in part by being presented with paradigms of the various colours, paradigms which, for us, fit within a natural visual similarity space. In acquiring mastery of colour terms, we then acquire a disposition to judge visually presented cases of particular colours to be the particular colours that they are.'[12]

We cannot attain mastery of colour terms simply by being disposed to accept a variety of *a priori* colour platitudes. We need to actually go out and look at colours so that we know how they, in Smith's words, 'hook up' with the platitudes.

The analogy between colour platitudes and moral platitudes breaks down here. Looking at colours is an *a posteriori* investigation, and it is essential for mastering colour terms correctly. If Bert says that he accepts every *a priori* platitude necessary for mastering colour terms, but gets all his perceptual colour judgements wrong, then it seems that Smith would deny that Bert has really mastered colour terms. If the analogy continues to hold between colour terms and moral

---

[10] Smith (1994), p. 51.
[11] Smith (1994), p. 50.
[12] Smith (1994), p. 51.

terms, then, it appears that we must use *a posteriori* investigation as well as *a priori* moral platitudes in order to master moral terms.[13]

Obviously, the core-apriorist will not say that we need to have *a posteriori* information about morality in order to master moral terms. So at this point he might drop the colour analogy, but Smith does not do this. He appears to hold that whilst the analogy breaks down at the point where perceptual experience is required in order to master colour terms, the analogy still usefully shows that simply being disposed to accept *a priori* moral platitudes is not enough. Just as we need something extra, *a posteriori* perceptual experience, to master colour terms, we need something extra to master moral terms. This 'something extra' is moral motivation.

Smith calls himself an internalist. He holds the following two claims that make up internalism:

Practicality Requirement: If an agent judges that it is right for her to φ in circumstances C, then either she is motived to φ in C or she is practically irrational.

Rationalism: If it is right for agents to φ in circumstances C, then there is a reason for these agents to φ in C.[14]

Practicality Requirement implies that if a practically rational agent has no motivation to φ in circumstances C, she *cannot* judge that it is right to do so. For example, she cannot truthfully claim 'I know that it is right for me to donate to charity, but I have no motivation to do so'. A practically rational agent with no moral motivations may accept all true *a priori* moral platitudes (and reject all false *a priori* moral platitudes), but she would not have mastered moral terms. She fails to meet *Practicality Requirement*, and so she cannot make genuine moral judgements.

David Brink rejects *Practicality Requirement* because he believes that agents who can make genuine moral judgements but who have absolutely no motivation to act morally are possible.[15] He calls such agents 'amoralists'. Smith replies by using the colour analogy. Even if an agent who has been blind from birth can use colour terms more reliably than many other agents, 'the

---

[13] Smith's thoughts on colour are reminiscent of Frank Jackson's (1986) thought experiment about Mary, who knows every physical fact but has never seen colours. Despite her knowledge, Jackson argues, she could not know what colours look like. This is not the same as Smith's points about colour platitudes, though, because Jackson is not trying to show that mastery of colour terms requires *a posteriori* information. His aim instead is to reject physicalism, which holds that if one knows all physical facts, then one knows all facts. Jackson claims that since Mary knows all physical facts but does not know what colours look like, physicalism must be false.
[14] Smith (1994), pp. 61-62.
[15] Brink (1989), pp. 27, 46-50.

ability to have the appropriate visual experiences under suitable conditions is partially constitutive of possession of colour concepts and mastery of colour terms… [The blind agent] does not *really* make colour judgements at all'.[16] Similarly, an agent may make generally reliable moral claims and use moral terms, but unless he is motivated to abide by his moral judgements, he is not really making *genuine* moral judgements.

Smith does not claim that the colour analogy proves that Brink is wrong, and he provides an independent argument for *Practicality Requirement* elsewhere. Rather, the analogy purports to show that Brink cannot just assume that the amoralist is possible. At this point, accepting *Practicality Requirement* is just as *prima facie* reasonable as rejecting it.

We need not determine here whether *Practicality Requirement* is true or not. I bring it up to show a use that Smith has for the colour analogy, and to show that we can reject Smith's metaethics by rejecting *Practicality Requirement*. If we reject it, then the colour analogy breaks down. We need something extra, *a posteriori* perceptual experience, to master colour terms, but if we reject *Practicality Requirement*, then we do not know what the 'something extra' is that we need in order to master moral terms. This creates problems for Smith's metaethics because he cannot give up *Practicality Requirement*. If he does so, then various purportedly *a priori* moral platitudes about motivation that he wants to retain will turn out to be false. This means that the core-aposteriorist could criticise Smith's moral realism by rejecting *Practicality Requirement*, and many philosophers have done so, whether they are core-aposteriorists or not.[17] However, the core-aposteriorist might want to retain *Practicality Requirement*, depending on her own core-aposteriorist theory. Since I do not want to dictate what a core-aposteriorist must hold any more than I have to, I will drop this criticism and concentrate on how we can use the *a posteriori* strategy against Smith. As we will see, the *a posteriori* strategy only leads to the conclusion that many of Smith's purportedly *a priori* moral platitudes are actually likely to be *a posteriori*. This is enough to show that his metaethics cannot be a core-apriorist theory, as it follows option (c) from the options that the core-aposteriorist can use to argue against Smith. It does not show, however, that any of the platitudes that he uses are wrong, or that *Practicality Requirement* is false (or true, for that matter). The core-aposteriorist is therefore free to endorse or reject it as she chooses.

---

[16] Smith (1994), p. 69, emphasis in the original.
[17] E.g. Copp (1997), pp. 48-53, Shafer-Landau (2003), pp. 145-147; Nichols (2004), pp. 72-75. Nichols, as we have seen, is a core-aposteriorist. Shafer-Landau is a core-apriorist.

3 – *A priori* moral platitudes

We have seen that Smith divides the platitudes necessary for mastery of moral terms into five types, and that the types need not necessarily share any connection that would make it easier for the core-aposteriorist to argue that they are also false. Neither do we have reason to believe that they all contravene *Sufficient Determinacy*. This leaves us with option (c), the strategy of showing that at least some of the platitudes are *a posteriori*. We can do this with at least three of the five types of platitude. The remaining two types, the platitudes of objectivity and supervenience, are more likely to be *a priori*, but this is not enough to show that Smith's moral realism successfully works as a core-apriorist theory. The core-apriorist therefore cannot use Smith's moral realism to defend the CAT against the core-aposteriorist.

The three types of moral platitude that I will examine are practical moral platitudes, substantial moral platitudes, and procedural moral platitudes. Specifically, I will examine the examples that Smith himself gives of such platitudes. The example he gives of a practical moral platitude is

Practical: Weakness of will, compulsion, depression and the like may explain why someone isn't moved in accordance with their moral beliefs.

One might be able to argue that we can define 'weakness of will' in such a way that it is knowable *a priori* that an agent with weak will is less moved to act in accordance with their moral beliefs (though how informative this definition will be is another matter – what is weak will, anyway?) Agents with weak will are less likely to perform difficult acts, and being moral is not always easy. But defining 'depression' in a similar way is trickier, not least because many depressed agents are very good at covering up their condition. They may act no differently from how they would if they were not depressed.[18] It is true that many (though not all) depressed agents lack the motivation to act, but we find this out by observing them. Defining depression so that it necessarily involves lack of motivation is not the answer, as depression is recognised as

---

[18] One might object that they still do not *feel* like acting, even though they do act, so their motivation is negatively affected. But it is difficult to flesh this out. If they freely choose to act morally and successfully do so, in what sense do they lack motivation?

producing a variety of symptoms, and an agent need not display all of them to be diagnosed as suffering from depression. Even if it is part of the definition of 'depression' that depression affects an agent's will power, this appears to be an *a posteriori* discovery, made by observing depressed agents. And in fact the platitude goes further than just saying that depression may affect an agent's motivation. It says that depression may affect an agent's *moral* motivation. There may be many depressed agents who lack the motivation to perform some types of action without the depression affecting their moral motivation.

A risk also arises for the core-apriorist with the word 'may', as it means that we are left uncertain just when depression, weakness of will and the like are going to affect an agent's motivation and how they would do so. As depression need not affect motivation, we need to know under which conditions it does, and this varies from agent to agent. So not only does it seem that *Practical* is actually knowable in part *a posteriori*, but it turns out not to be determinate. Therefore because of *Sufficient Determinacy* it cannot by itself justify determinate moral propositions. 'May' is not the only term in *Practical* with this problem, either. If many agents disagree on the conditions under which, say, depression adversely affects motivation, how it does so, and what we should take motivation to be, then the term 'depression' has imprecise application conditions. This leads back to the problem that we can only reasonably give the term more precise application conditions by examining depression and depressed agents *a posteriori*. It is unlikely that we can know *Practical* via an *a priori* method.

Substantial moral platitudes are also problematic for the core-apriorist. The example Smith gives of such platitudes is

Substantial: Right acts are in some way expressive of equal concern and respect.

The core-aposteriorist wants to argue that if *Substantial* is correct, then we can only find out *a posteriori* that it is correct. For example, perhaps we have to investigate actual societies to discover that *Substantial* is true.[19] If Smith wants to say that *Substantial* is necessary for mastery of moral terms, on the other hand, then he has to argue instead that it is *a priori*. In this case, the core-aposteriorist can use option (c) in an indirect fashion. First, she can argue that

---

[19] We shall discover later in this chapter that the moral functionalist can hold something like this claim.

Substantial contravenes *Sufficient Determinacy*. After that, she can argue that although we can amend *Substantial* to correct this problem, we can only do so by using *a posteriori* information.

Substantial falls foul of *Sufficient Determinacy* because of the phrase 'in some way expressive'. What is this phrase supposed to mean? Equal concern and respect are not the be-all and end-all of morality, otherwise Smith could just say that right acts are acts that best display equal concern and respect (This may be problematic on other grounds as well, as saying this comes close to endorsing a reductive moral theory, and Smith rejects reductive theories.). How far, then, do right acts have to accommodate equal concern and respect? Does it vary from case to case? Could there ever be right acts that ignore them entirely? With so many questions, it becomes increasingly unclear just which determinate moral propositions can be justified with *Substantial* as a premise.

Smith can of course say that the correct moral theory must be consistent with *Substantial*. If *Substantial* is true, then obviously moral theories that reject it cannot be accepted. The point, though, is that consistency is not enough. If an agent masters moral terms, then she knows how to use them correctly when making determinate moral claims. If all she knows is *Substantial*, then there will be many occasions on which she does not know whether she is using the term 'right' correctly, or whether she is using it inappropriately. For example, take the moral proposition 'It is right to treat your children in exactly the same way as you would treat any other children'. All children should be treated with equal concern and respect, according to this proposition, but very few people would actually agree with this proposition. Surely I should give my children preferential treatment, even if nobody else is obliged to give them preferential treatment.

For Smith to avoid this problem, he needs to provide other substantive moral platitudes that are more determinate; maybe these are platitudes that replace the phrase 'in some way expressive' in *Substantial* with terms that have more precise application conditions. However, the onus is on Smith to show this. Furthermore, this move allows the core-aposteriorist to argue that we can only make the necessary changes by appealing to *a posteriori* information, and this is where option (c) comes back into the picture. If we want to exclude the possibility that it is morally right to treat your child exactly like you treat any other child, we need a reason to exclude it. What is to say that this reason can be identified *a priori*? As we saw with the SLS

theory, if we bring in a first-order normative theory to provide a reason, it may give an *a posteriori* reason.

So far, the core-aposteriorist can use option (c) to argue against Smith's use of certain practical and substantial moral platitudes. There are three other types of platitudes: procedural platitudes, platitudes of objectivity and platitudes of supervenience. Procedural platitudes may also be vulnerable to the core-aposteriorist, depending on what actually justifies an agent's endorsement of a particular moral proposition or particular moral claim. For example, take

> *Procedural*: We can work out which acts are right through a process of reflective equilibrium.

If substantial platitudes can be *a posteriori*, then we can work out which acts are right *a posteriori*. First, we work out what sort of characteristics right acts have, and then we work out which particular acts (such as donating to charity, helping the impoverished, and so on) have these properties. If reflective equilibrium helps us to work out what is morally right, we discover that it is successful *a posteriori*, since it relies on *a posteriori* evidence about rightness and the characteristics of various acts.

The other two types of platitude that Smith thinks are required for mastery of moral terms, the platitudes of objectivity and supervenience, are much harder to question as *a priori* or determinate. It is difficult to see just what *a posteriori* evidence there could be to demonstrate the truth or falsity of such platitudes, and it is doubtful whether such platitudes must be non-determinate. The examples that Smith gives are:

> *Objectivity*: When A says that φ-ing is right, and B says that φ-ing is not right, then at most one of A and B is correct.
>
> *Supervenience*: Acts with the same ordinary everyday non-moral features must have the same moral features as well.

Neither *Objectivity* nor *Supervenience* seems non-determinate. It is also unclear how the core-aposteriorist could show that objectivity and supervenience can be defended *a posteriori*. She

could deny that the platitudes are true, but this would require an independent argument, and as I have mentioned above, rejecting platitudes of supervenience is particularly problematic for her.

However, this is not a very large problem, because there is no requirement to show that all of the moral platitudes that Smith believes are necessary for mastery of moral terms are *a posteriori* or non-determinate. The Central Claim is true if there is a single determinate core moral proposition with normative content that can only be ascertained as true with an *a posteriori* or mixed method. Platitudes about supervenience and objectivity cannot by themselves justify the normative content of many true determinate moral propositions. For example, take the determinate moral principle 'One should not drink-drive'. Platitudes about supervenience and objectivity cannot by themselves justify the normative content of the proposition. They can provide reasons why the proposition is true for every case of drink-driving if it is true, but not that it *is* true. One reason for this is that we cannot derive any sort of theory of the good or of the right merely from platitudes of objectivity and supervenience, but we need such theories in order to identify what is morally good or right. Theories of the good and of the right are more the business of substantial moral platitudes, so we must examine these platitudes as well in order to justify the normative content of the proposition. We may also need to consider practical and procedural platitudes. But we have seen that practical, substantial and procedural platitudes may not be ascertainable as true *a priori*, in which case we cannot justify the normative content of the proposition with an *a priori* method. Therefore the CAT is likely to be untrue, despite the fact that the platitudes about supervenience and objectivity may be both determinate and knowable *a priori*.

The core-aposteriorist is helped here by how weak the Central Claim is. The Central Claim does not claim that the truth of the normative content of all determinate core moral propositions can only be ascertained with an *a posteriori* or mixed method, but only that the normative content of at least one determinate core moral proposition is. The core-apriorist wants to say that the moral platitudes necessary for mastery of moral terms are knowable *a priori*, and hence are *a priori* core moral propositions. All determinate core moral propositions are platitudes that are knowable *a priori*. However, there is good reason to believe that some of the platitudes necessary for mastery of moral terms are not knowable *a priori*, but *a posteriori*. The core-apriorist therefore cannot use Smith's moral realism to defend the CAT. Smith does not

successfully defend his claims that the platitudes necessary for mastery of moral terms are knowable *a priori*.

Before moving on, we can raise the point here that the *a posteriori* strategy's objections to Smith's moral realism as a core-apriorist theory have something in common with its objections to the SLS theory and Kramer's metaethics. The *a posteriori* strategy refutes none of these theories outright. It simply states that, despite appearances to the contrary, the theories can only work if they allow for determinate core moral propositions of which the normative content is only knowable via an *a posteriori* or mixed method. This does not imply that *a priori* determinate core moral propositions do not exist. Both Kramer's metaethics and Smith's moral realism, for example, can continue endorsing *a priori* determinate basic moral principles and *a priori* determinate moral platitudes respectively. However, none of the theories is successful specifically as a core-apriorist theory.

In sections 4-7 of this chapter, I will turn to the final metaethical theory that I will examine, moral functionalism. The core-apriorist can read moral functionalism as a core-apriorist theory, and such a reading entails that the CAT is true. The core-aposteriorist's counterargument is not that moral functionalism is wrong, but that this core-apriorist reading is untenable.

4 – Moral commonplaces and A-extension propositions

Moral functionalism was created by Frank Jackson and Philp Pettit, and I will examine it as Jackson builds on it following their initial joint paper that developed it.[20] It claims that we are able to attain moral knowledge by ascertaining the commonplaces about morality that are typically accepted by people in general. Jackson and Pettit do not explicitly claim that moral functionalism is a core-apriorist or a core-aposteriorist theory, but it is easy to read it as core-apriorist because of certain *a priori* commitments it endorses. After outlining moral functionalism and identifying what these *a priori* commitments are, I will examine two objections. The first objection is developed by Nick Zangwill, who claims that moral functionalism leads to an unacceptable relativism. We will find that Zangwill's argument needs considerable tightening to really cause problems for moral functionalism, and it is not as strong as the second objection.

---

[20] Jackson and Pettit (1995); Jackson (1998).

The second objection aims against moral functionalism's *a priori* commitments by claiming the commitments are not really *a priori*. This implies that moral functionalism cannot show that determinate core moral knowledge is *a priori*.

Moral functionalism claims that moral terms are 'used in a way that presupposes a large network of connections with other terms, both evaluative and descriptive'.[21] We accept commonplaces about the terms that we use, and this allows us to identify particular descriptive properties that ordinary moral thinking tells us are moral properties. We are thus able to grasp moral concepts. For example, when we use the term 'fairness', it is a commonplace for us that if an action is fair, that is usually a reason to do it. Any agent who can use moral terms competently will agree that various commonplaces such as this are true.[22]

'[T]he meaning of relevant moral terms will be fixed by roles which certain commonplaces give them, and so moral thinking is bound to involve the attempt to use commonplaces as a base, and holding on to as much as that base as possible, or at least to the parts considered most secure, to fix opinions on particular questions.'[23]

Grasping the commonplaces correctly means that we grasp the concepts relating to the commonplaces correctly, which means that we can correctly use the terms that denote moral properties. Once we know about the roles that moral properties play, and the moral commonplaces that we accept, we can find out what descriptive properties we can identify with the moral properties.

The reason why moral functionalism is a moral coherentist theory lies in the fact that moral terms are connected with certain evaluative and descriptive terms that are based on what our moral commonplaces actually are. There is no requirement that any of our commonplaces must be self-evident, so moral functionalism is not a moral intuitionist theory. Instead, morality is based on a network of commonplaces that we accept and that are all consistent with each other.

What *a priori* commitments does moral functionalism have? Jackson claims that

---

[21] Jackson and Pettit (1995), p. 22.
[22] Jackson and Pettit (1995), pp. 22-23.
[23] Jackson and Pettit (1995), p. 26.

'[w]hat is a priori according to moral functionalism is not that rightness is such-and-such a descriptive property, but rather that *A* is right if and only if *A* has whatever property it is that plays the rightness role in common folk morality, and it is an a posteriori matter what that property is.'[24]

'Common folk morality' is the moral theory that endorses the moral commonplaces accepted by most people in a particular society.[25] Moral functionalism presupposes that we look at the commonplaces typically endorsed by people, and that we use that as the basis of a suitable moral theory.

There are two ways to read Jackson's remarks, and the one we choose affects whether we see moral functionalism as a core-apriorist or a core-aposteriorist theory. We might think that since Jackson says that we identify which descriptive properties are moral properties *a posteriori* by looking at common folk morality, moral functionalism must be a core-aposteriorist theory. However, if we dig deeper into Jackson's metaphysics, we find more evidence for a second reading.[26] To know what moral properties are, we must know *a priori* that moral properties play a particular role. Only after that can we use *a posteriori* evidence to ascertain which descriptive properties are moral properties, since they play the same role. The core-apriorist can claim that the knowledge that Jackson claims is *a priori* in the above quote is *a priori* determinate core moral knowledge. The core-aposteriorist will object that the *a priori* knowledge is not actually *a priori*.

Moral functionalism represents Jackson's metaphysics as it applies to ethics. He believes that one important task of metaphysics is what he calls 'the location problem'. If the proposition 'Kicking babies for fun is morally wrong' is true, where can we find this property of moral wrongness? The property will have various features, and if our account of moral wrongness entails that the property has features A, B and C, then those features 'have a place in the account'.[27]

To demonstrate, Jackson provides an example concerning the supervenience of the psychological on the physical. He first defines 'physicalism' as the metaphysical doctrine that

---

[24] Jackson (1998), pp. 150-151.
[25] Jackson and Pettit (1995), p. 24; Jackson (1998), pp. 117-118.
[26] Much of the following discussion relates to Jackson's metaphysical views, and it should be noted that these may not be Pettit's views.
[27] Jackson (1998), p. 5.

we can give a complete account of our world in terms of physical particulars, properties and relations.[28] Given physicalism, the following is true:

(b)   Any world which is a *minimal* physical duplicate of the world is a duplicate *simpliciter* of the world.[29]

Minimal physical duplicates are duplicates that have exactly the same physical properties and no further non-physical properties. Two worlds with exactly the same physical properties would not be duplicates if one of the worlds also contained extra non-physical properties (such as ectoplasm).

Suppose that ϕ is a complete, true physicalist account of our world. It will also be true, given (b), of any world that is a minimal physical duplicate of our world. Suppose that γ is a true sentence about the psychological nature of our world, such that it would only be false if the world's psychological nature were different. Given (b), every world at which ϕ is true would be a world at which γ is true, so therefore ϕ entails γ.[30] Our physicalist account of the world entails the supervenience of the psychological on the physical, and hence we should accept this supervenience.

Importantly, Jackson sharply divides the concepts that actual agents generally use from what he calls the 'essence' of what these concepts relate to.[31] For example, what is essential about water is that it is H_2O, but we do not need to know that in order to grasp the concept 'water'. People grasped the concept perfectly well long before they knew its chemical composition. This distinction enables us to call moral functionalism a core-apriorist theory, as I will now show.[32]

Jackson accepts that the investigation of which commonplaces we endorse can be *a posteriori*. The *a priori* element appears when we talk about the application of terms in the actual world or in counterfactual situations. Let T be a particular term. T may apply to various entities, events, relations, etc., and this defines its extension. It may have different extensions in different possible worlds. The A-extension of T in *w* is the actual extension of T in world *w*

---

[28] Jackson (1998), p. 6. As I use this only as an illustrative example, I will not discuss what 'physical' means.
[29] Jackson (1998), p. 12, emphasis in the original.
[30] Jackson (1998), p. 25.
[31] For the avoidance of doubt, Jackson's use of the terms 'essence' and 'concept' is unrelated to Cuneo and Shafer-Landau's use of the terms.
[32] One can try refuting Jackson's metaethics by refuting his general metaphysics. This will not be my strategy against moral functionalism, so I will not examine this move. However, I will touch on it later in this chapter.

where *w* may be our world or another possible world. In our world, the A-extension of 'water' is all and only the occurrences of water.

We may also ask, given the assumption that we are talking about our actual world, what T would apply to in various counterfactual situations. The answer would give us the C-extension of T, which would tell us what T would apply to in possible worlds. For example, the term 'water' applies to all the watery occurrences in the world, so in our actual world the A-extension of 'water' would be all and only the occurrences of H_2O. In another word, where substances of chemical composition XYZ are called 'water', the A-extension of 'water' in that world would be all and only the occurrences of XYZ. But since in our actual world water is H_2O, the C-extension of 'water' in the actual world is the occurrences of H_2O in any possible world.[33] The C-extension would not, for example, include any occurrences of XYZ, even though people on possible worlds with XYZ may call XYZ 'water'.

For 'water', the A-extension and the C-extension in the actual world are the same. 'Water' applies to all and only the occurrences of H_2O under both extensions. But, says Jackson, there is an epistemological difference. Before we found out that water was H_2O, we did not know its essence. So we could not determine its C-extension.

'[I]n order to pick out water in a counterfactual world, we need to know something about *relationships* between the counterfactual world and the actual world that we could only know after discovering that in the actual world H_2O plays the watery role.'[34]

Before we found out that water was H_2O – an *a posteriori* discovery – we could not say whether the stuff that plays the watery role in a counterfactual world was part of the C-extension of 'water'. For consider a counterfactual world that is identical with the actual world except that the stuff that plays the watery role in the actual world is H_2O, and the stuff that plays the watery role in the counterfactual world is XYZ. Until we find out that in the actual world water is H_2O, how do we work out that the stuff playing the watery role in the counterfactual world does not fall under the C-extension of water?

---

[33] Jackson (1998), p. 49.
[34] Jackson (1998), p. 50, emphasis in the original.

Knowledge of A-extensions (in our actual world) need not be *a posteriori* knowledge of the actual world. 'Water' is whatever plays the watery role in the actual world. We do not need to know that water is H_2O in order to know that. And this applies to any counterfactual world as well, because when we work out A-extensions of 'water' in counterfactual worlds, we do not compare watery substances in those worlds to the watery substance in any other world.

Since we can know A-extensions in our actual world without knowing what the actual world is like, Jackson concludes that such knowledge is *a priori*.

'The sense in which conceptual analysis involves the a priori is that it concerns A-extensions of worlds… and accordingly concerns something that does, or does not, obtain independently of how things actually are. When we do conceptual analysis of K-hood, we address the question of what it takes to be a K in the sense of when it is right, and when it is wrong, to describe some situation in terms of 'K', and so we make explicit what our subject is when we discuss Ks. The part of the enterprise that addresses the question of what things are K at a world, under the supposition that that world is the actual world, is the a priori part of conceptual analysis, because the answer depends not at all on which world is in fact the actual world.'[35]

In the actual world, it is *a priori* that the A-extension of 'water' is just the watery stuff of our acquaintance. Water is whatever happens to fulfil the watery role. What precisely that stuff is (H_2O) is an *a posteriori* question, but we do not need to know that in order to know that water is the watery stuff of our acquaintance.

Jackson cannot think that it is always true that K is whatever plays the K-role, as that is clearly wrong. London is not the city that happens to play the Londony role, because 'London' is just the name of that particular city. There is no Londony role, because London would be London no matter what its characteristics or properties were. Similarly, demonstratives ('this', 'that' 'those' and so on) are not associated with any role. The proposition 'This was given to me at lunchtime' does not imply that anything plays a 'this-y' role, or that such a role exists.

---

[35] Jackson (1998), p. 51.

So the proposition ''K' refers to whatever plays the K-role' is false if K is some name or some demonstrative. Nevertheless, Jackson wants to claim that for many entities and properties, such as water, 'K' does refer to whatever plays the K-role. The reason why 'water' refers to occurrences of H_2O is because H_2O plays a certain role. Various terms can have *a priori* A-extensions, even if not every term can, and furthermore some of these terms are moral terms. In future, when I talk of terms referring to things that play certain roles, I will mean terms that have A-extensions and C-extensions in Jackson's metaphysics.

What is the property of fairness? The A-extension of the fairness property (at our actual world) will be the thing that we are acquainted with because it fulfils a certain 'fairnessy' role, and we know this *a priori*. What we must discover *a posteriori* is exactly what descriptive property fulfils the role. Jackson does not say who 'we' are, presumably because he does not think that we should be too precise. He is also wary of defining roles in a way that deviates too much from the way that the relevant terms are commonly used. If I define 'number' in a certain way, regardless of how others typically use and define the term, then it may be easy for me to define 'number' in such a way that entails that numbers actually exist, and hence that mathematical realism is correct. But as Jackson rightly points out, this is false progress. Nobody who did not share my definition would agree that I have proven mathematical realism correct, and it might turn out that very few people share my definition. It is only when we examine numbers 'according to *our ordinary conception* [of them] or something suitably close to it that we can make any real headway'.[36]

The very idea of examining moral commonplaces in order to ascertain moral properties is *a priori*. It is *a priori* that moral properties are simply the things that play the 'moral property' roles, which are particular roles in the eyes of moral agents. We can roughly say that moral properties are those things that fulfil certain functions according to our folk morality (As we will see shortly, this is somewhat a simplification, but it is a good starting point.). Metaethical theories that analyse moral concepts in ways that are completely alien to our ordinary conception of them are most likely wrong, and the way to find out what our ordinary conception of them is is to go out and look at what moral commonplaces to endorse. What is *a priori* is our knowledge of the A-extensions of moral terms. What is *a posteriori* is our knowledge of the essence of moral

---

[36] Jackson (1998), p. 31, emphasis in the original.

properties, and we investigate the outside world to discover what moral properties are actually made up of.

It appears that both A-extension propositions (propositions that express A-extensions) and C-extension propositions can be commonplaces. 'The property of moral rightness is whatever plays the moral-rightness role' is a commonplace, and it expresses an A-extension. However, Jackson and Pettit also believe that for many people it is a commonplace that saving lives is more important than being fair.[37] This is an *a posteriori* commonplace, since humans might have come to believe that it is a commonplace that being fair was more important than saving lives instead, and we would find that out through observing them. So commonplaces may either express A-extensions or C-extensions, and thus be either ascertainable *a priori* or *a posteriori*.

Jackson's views on colour properties also point to something important that we can carry over to his views on moral properties. He identifies colours with physical properties, and so makes them objective and observer-independent, in that they may still exist independently of any possible observers. What is not objective and observer-independent is precisely what particular descriptive properties are colours. That identification can only be made by examining which colour commonplaces are accepted by actual agents.[38]

Despite the significant *a posteriori* elements in moral functionalism, it ultimately rests on an *a priori* foundation. There is more to say about this, but for the moment let us develop moral functionalism further to see more about how it treats the attainment of moral knowledge.

5 – Folk morality

A folk morality consists of the moral commonplaces that people generally accept. Obviously, most people are not philosophers, and they can hold moral commonplaces that badly conflict with each other. The answer to this problem is not to junk folk morality, but to recognise that this kind of folk morality is a *common* folk morality, a folk morality that perhaps has not been developed or worked out to any real degree. Instead of rejecting it, we must develop it into a *mature* folk morality. Moral functionalism claims that we can drop various commonplaces if we need to when they conflict with other more plausible commonplaces. We try to keep as many

---

[37] Jackson and Pettit (1995), p. 23.
[38] Jackson (1998), p. 100.

commonplaces as possible, but we do not have to keep them all. As people continue debating moral issues and refining their moral views, we work out which commonplaces to keep and which to drop. This gets us closer to a mature folk morality, a 'folk morality that has been exposed to debate and rational reflection'.

'The idea is that mature folk morality is the best we will do by way of making good sense of the raft of sometimes conflicting intuitions about particular cases and general principles that make up current morality.'[39]

Suppose that it is a commonplace for Alice that allowing a morally wrong action to happen is just as morally wrong as performing it. Shooting someone is just as morally wrong as allowing your friend to shoot someone when you could stop it without any danger or inconvenience to yourself. Meanwhile, it is a commonplace for Bethany that allowing a morally wrong action to happen is not as morally wrong as performing it. You may be morally culpable to some degree for the shooting, but not as much as the actual shooter would be (After all, it is not your finger on the trigger.). If Alice and Bethany tried to work together through the reasons that they held their commonplaces, they may be able to agree on reasons to prefer one of the commonplaces to the other. Say that they agree that Alice's commonplace is preferable. They would both reject Bethany's commonplace and so they would be closer to having a mature folk morality than they were previously. They may also use the agreed commonplace as evidence for further moral claims that they could come to think of as new commonplaces. This raises the possibility that if we refine different common folk moralities, they will converge on a mutual folk morality. I will say more about this later.

6 – Zangwill's objection

We have now seen enough of moral functionalism to starting considering objections to it. In this section, I am going to examine a criticism from Nick Zangwill. Zangwill wants to show that moral functionalism cannot avoid sliding into a disastrous form of relativism, and so it should be

---

[39] Jackson (1998), p. 133.

rejected. Unfortunately, the objection as he develops it depends on an assumption that the moral functionalist can refuse to accept. We can, however, adapt the argument to avoid making the assumption by using the distinction between determinate and non-determinate propositions. This causes more trouble for the moral functionalist than Zangwill's original objection, but it is still not enough to conclusively refute moral functionalism.

Jackson and Pettit point out that actual speakers in actual societies use moral terms in a way that conforms with various commonplaces, even if the speakers cannot articulate those commonplaces. This inability is not automatically a mark against moral functionalism. Many English speakers use perfect grammar, but we cannot assume that they can list all the grammatical rules that they follow. However, as Zangwill points out, this assumes that actual speakers will broadly agree on what the commonplaces entail (A group of competent English speakers will broadly agree on when an English sentence is grammatically incorrect, even if they cannot list the rules of English grammar.).[40] Jackson justifies moral functionalism's use of commonplaces on the grounds that in order to have genuine moral disagreement, we need 'a background of shared moral opinion to fix a common, or near enough common, set of meanings for our moral terms'.[41] But if the disputants share no or very few moral commonplaces, then Zangwill thinks that according to moral functionalism there should be no genuine moral disagreement at all. The disputants would use the same terms, but they would be talking about different properties, because they would be using different concepts. Since it seems clear that the disputants can have genuine moral disagreements despite sharing so few moral commonplaces, Zangwill concludes, moral functionalism must be mistaken. The disputants have different folk moralities, and moral functionalism must concede that there is no reason for the disputants to give up their separate commonplaces. So the disputants can adhere to their different moral theories, and moral functionalism leads to relativism.

Jackson hints at a solution by saying that there can be genuine moral disagreement between disputants who have a 'near enough common' understanding of moral terms. They talk about the same properties, but disagree as to the exact nature of those properties. The trouble with this response, Zangwill thinks, is that it is difficult to flesh it out in a way that does not cause more problems. He focuses on when the use of indexicals make a difference in moral

---

[40] Zangwill (2000), p. 277.
[41] Jackson (1998), p. 132. 'Genuine moral disagreement' is genuine disagreement about a moral issue, where the disputants are not just talking past each other.

judgements. We typically accept that I can give preferential treatment to my child and not your child specifically because my child is my child and your child is not. But we do not accept that I should give preferential treatment to people of my race just because they are of my race. So when is it morally significant that I have a personal connection to something? Our society constantly faces this problem in one form or another. Can I keep refugees out of my country because they take away resources from my fellow nationals? May I eat different species simply because they are not human?

Jackson thinks that it is a commonplace that acts of intentional killing are wrong.[42] Zangwill objects that this depends on how one views the moral relevance of indexical differences. If intentional killing includes the killing of chickens for food, then meat-eaters might object to the commonplace. If we are not allowed to intentionally kill assassins in self-defence, this could also raise objections. And, Zangwill notes, if the commonplace means that we should not intentionally kill Jews, then ardent Nazis would object.[43] Whether the commonplaces are ones that we find reasonable or not, it seems that even when moral commonplaces look perfectly legitimate there is always the risk that someone will say 'Yes, *except…*'

We cannot just decree that Nazis are conceptually confused about the moral terms that they use. During his Nuremberg trial, Hermann Göring shocked the prosecutor by stating that loyalty to Germany was morally the most important thing, and to be absolutely loyal required intentional killing.[44] It appears at first sight that we can reasonably disagree with him about this. He knows what loyalty and duty are just as well as we do, and we and he disagree over our moral judgements about them.

According to Zangwill, this cases problems with moral functionalism. It seems that we can intelligibly debate the ethics of genocide with Göring. Göring's moral views might be horrendously outlandish, but that does not mean that he cannot understand the moral terms that we use. However, it appears that since our moral views are so different, we and Göring accept very few of the same moral commonplaces about the moral terms used in the debate. Perhaps we mutually accept none of them. Moral functionalism must therefore hold that we cannot have a genuine moral disagreement. We and Göring cannot mean the same thing by the moral terms

---

[42] Jackson (1998), p. 131.
[43] Zangwill (2000), pp. 280-283.
[44] I assume that Göring was being honest about his beliefs, and could actually grasp moral concepts, even if his moral judgements were horrendously wrong. If he was not, we can assume that he was for the sake of the argument.

that we use, and so we will just be talking past each other. Zangwill further claims that even if Göring has a mature folk morality, it is 'not remotely plausible' that he will ever come to agree with us.[45] Since this is the case, and since there is no way of resolving the disagreement, there is no way to show that Göring is wrong. He and we accept different moral commonplaces, and that is all that can be said. Therefore moral functionalism leads to a disastrous form of relativism, which allows Nazism to be morally permissible depending on the moral commonplaces that agents accept.

This objection needs considerable tightening. The problem with it as it is currently stated is that we and Göring may actually agree on many of the same commonplaces. More specifically, we may agree on many of the same determinate moral propositions that express commonplaces. We may agree, for example, that it is a bad thing to kick dogs for fun, and to allow one's child to drown in a lake. Göring may generally be just as horrified by serial killers, selfish swindlers and destructive psychopaths as we are. So it just seems false that we cannot share moral commonplaces, or that we cannot have a genuine moral disagreement. This should lead us to tighten up Zangwill's objection, which becomes the amended objection that we cannot have a genuine moral disagreement with Göring about the intentional killing of Jews because we do not share any determinate moral commonplaces that are relevant to whether it is morally permissible or not to kill Jews.

The reason why I said that we may agree on many determinate moral propositions, and left non-determinate moral propositions to one side, is that it is easier to agree on what determinate propositions entail. We may agree with Göring that everyone should be nice, but if we interpret 'nice' in wildly different ways, then we are not agreeing on very much, and so we do not share many of the same commonplaces regarding being nice. This problem is much smaller for determinate propositions. In effect, if we agree on a very determinate moral proposition, there is less room for someone to say 'Yes, *except…*' and so there is a greater chance that we do share the same moral commonplaces.

The amended objection is still mistaken, though, because we probably can still have a genuine moral disagreement with Göring. Surely we and Göring can agree that intentional killing is generally wrong. The disagreement arises when we try to make the commonplace more

---

[45] Zangwill (2000), p. 284.

determinate. When we replace 'generally' with more specific criteria for ascertaining when intentional killing is wrong, we and Göring will disagree on what the criteria are. Nevertheless, the point remains that we agree on something. Most of the time, Göring would claim, intentional killing is wrong. This means that we do share a commonplace with Göring. The disagreement arises over how to flesh it out, and it seems implausible to claim that the disagreement cannot be genuinely moral because we share so few determinate commonplaces with Göring.

So what happens if two disputants really do share none of the same moral commonplaces? In this case, moral functionalism really would imply that the disputants could not have a genuine moral disagreement, as they would be talking about utterly different things. But it is not clear why this should be a problem for moral functionalism. If the disputants did have a genuine moral disagreement, what would it be about? If they mean utterly different things by the terms that they use, how could they have a disagreement? Would they not just be talking past each other?

Maybe we can alter Zangwill's objection in another way to argue more successfully against the moral functionalist. We can try using the distinction between determinate and non-determinate moral propositions that express commonplaces. Among the commonplaces that we and Göring share might be 'Intentional killing is generally wrong', or 'We should work for global justice'. But these are non-determinate. The moment we start unpacking them (say, by defining what global justice is, and what it implies), we are going to start disagreeing. For Göring, global justice requires the supremacy of the Aryan race. For us, global justice requires ending the threat of Nazism. This means that we can have a moral disagreement with Göring on the issue. We share the same non-determinate commonplaces about global justice, and we may also share some determinate commonplaces that are consistent with endorsing global justice. The trouble is that we do not agree on all the commonplaces that are consistent with endorsing global justice, particularly those possible commonplaces that relate to racial issues. Due to the commonplaces that we do share with Göring, we can have a moral disagreement with him about those issues. The disagreement is what results when we take the non-determinate commonplace 'We should work for global justice' and try to work out what determinate moral commonplaces follow from it.

At the moment, we do not share a folk morality with Göring. We might agree that we should work for global justice, but we certainly do not agree about the treatment of Jews. Our folk

moralities are going to diverge. Although Göring shares with us certain non-determinate commonplaces, our respective folk moralities endorse different determinate commonplaces. As soon as we try making the non-determinate commonplaces determinate, our moral views split from Göring's. So how can the disputants reach determinate moral commonplaces that they can all accept? The moral functionalist says that the disputants can refine their commonplaces so that they come to mutually agree on a mature folk morality. The more we talk about our moral views, the more we gradually come to settle upon mutually-agreed determinate moral commonplaces. Perhaps if Göring were a rational person, we would be able to convince him that his way of fleshing out the non-determinate commonplaces he holds is wrong. We may all finally be able to agree that 'Intentionally killing Jews is wrong', a more determinate commonplace, is correct, rather than Göring's determinate commonplace 'Intentionally killing Jews is not wrong'. Or we might reach a point where Göring's refusal to accept our commonplaces is just a sign of an irrational refusal to engage with us. Any reasonable observer would conclude that Göring's commonplaces are the wrong ones to hold.

This response assumes that different folk moralities are capable of converging in such a way. No doubt many can, but as Jackson admits, there is no guarantee that there is a single correct answer to a moral question. There could be two or more mature folk moralities that cannot converge any further, and which contradict each other over a particular moral question. There may be no way to decide between the two. Is this really a problem for moral functionalism, though? Jackson does not believe so. He hopes that the different folk moralities will converge, but he also concedes that they might not, in which case the commonplace that morality is objective will have to be dropped.[46] So it seems possible that there are different, conflicting mature folk moralities. Moral functionalism is compatible with relativism, even though it does not necessarily imply it. The moral functionalist hopes, though, that convergence is always possible, and hence that moral objectivity is true.

The fact that moral functionalism is compatible with certain forms of relativism is not automatically a strike against it. Neither does Jackson seem that concerned with relativist implications elsewhere in his metaphysics. When he writes of colour properties, he goes for a 'thoroughly anthropocentric' view in which colour properties 'typically interact… with normal

---

[46] Jackson (1998), p. 137.

human perceivers in normal circumstances [in a particular way]'.[47] This leaves it open for aliens to have a completely different folk colour theory than humans have, and Jackson appears to see nothing wrong with this possibility.

Nevertheless, the amended objection should worry the moral functionalist. Zangwill tried to argue that moral functionalism implies a form of relativism because moral disagreement is impossible between disputants who have different moral commonplaces. There is thus no way to resolve the disagreement to show that one disputant's moral commonplaces are superior to those of another disputant. The problem with this objection is that it is difficult to show that such moral disagreement is impossible, as it is more likely that the disputants share some non-determinate moral commonplaces, and disagree over more determinate moral commonplaces. But this leaves open the possibility that different folk moralities may not converge after all, so Zangwill can still reach the same conclusion: there is no guarantee that Göring's folk morality will converge with ours, and hence no guarantee that either folk morality is superior to the other.

If Göring has his own mature folk morality, and there is no way that we can show that he is mistaken except by using our own mature folk morality, then we do not really have any way to morally argue against Nazi extermination camps without using our own mature folk morality. All Göring has to do is reject the determinate moral commonplaces that we hold, and he can rest content with his own. His folk morality might even be completely logically coherent. But his views are so morally perverse that surely this cannot be right. Something has gone catastrophically wrong with moral functionalism if it allows this possibility, and Jackson's only defence is his optimistic hope that mature folk moralities will converge. If this hope is unfounded, than Zangwill has a point. He is wrong to say that moral functionalism is committed to the assumption that there can be no genuine moral disagreement between us and Göring. Since we can share many non-determinate moral commonplaces with him, there can be. But the attempt to show that we must be able to share determinate moral commonplaces fails, leaving the way open for a relativism which accords the same validity to genocidal racism as to humanitarianism. We can have moral disagreements with Göring, but neither of our mature folk moralities is morally superior to the other. The reason why we cannot reach an agreement is

---

[47] Jackson (1998), p. 97.

because there is no justification for either mature folk morality that demonstrates that the other mature folk morality cannot also be justified.

The moral functionalist might choose to bite the bullet and accept this conclusion. It would make moral functionalism much less attractive, but this does not automatically show that it is false. Neither does the objection show that relativism is a necessary result of moral functionalism. If the moral functionalist can demonstrate that mature folk moralities necessarily converge, then in the end only one mature folk morality will be left standing and the threat of relativism dissipates. Or even if different mature folk moralities cannot completely converge, they may converge sufficiently to the extent that they exclude all folk moralities that endorse moral atrocities. This would still leave us with relativism, but a comparatively harmless kind. Zangwill's objection would become much weaker. The question now becomes how likely the possibility of complete or sufficient convergence is, though, and the moral functionalist will need to do much more work to make his case.

Zangwill's objection tries to show that we should reject moral functionalism outright. If we do so, then obviously it could not be used as a core-apriorist theory, so the core-aposteriorist may choose to press the objection as far as it can go. However, there is a different type of objection that is preferable. It does not claim that moral functionalism is false, but that it cannot be a core-apriorist theory. I will develop this objection in the next section.

7 – Moral functionalism's *a priori* commitments

The objection claims that the only way that we can show that moral functionalism's *a priori* commitments are true is by *a posteriori* investigation, and therefore its commitments are not really *a priori*. This implies that even if moral functionalism works, it cannot be a core-apriorist theory.

Return to Jackson's metaphysics. Consider the following argument:

a) Sixty percent of the earth is covered by H_2O.

b) Water is the stuff that plays the watery role.

c) H_2O is the stuff that plays the watery role.

Therefore

d) Sixty percent of the earth is covered by water.

(a) and (c) are *a posteriori*. Before we could know (c), we had to check that H_2O actually plays the watery role. (b) is *a priori*.[48] It is something that we can know *a priori* about water.

Consider now:

e) Ben performed an act that maximised utility.

f) The property of moral rightness is whatever plays the moral-rightness role.

g) The property of maximising utility plays the moral-rightness role.

Therefore

h) Ben performed a morally right act.

(e) and (g) are *a posteriori*, as (g) acts as a commonplace that we must discover *a posteriori*. (f) is *a priori*, which is what we should expect, since it is an A-extension proposition. Jackson's justification for (b) and (g) being *a priori* is that they are about reference fixers. As far as (b) goes, 'water' refers to what plays the watery role, the stuff that plays the role of a colourless, drinkable liquid that makes up the ocean, and so on. It does not indicate what actually plays the watery role or what the watery role consists of. In fact, there may be nothing that plays the watery role.

The same points arise with (f). (f) may be *a priori*, but note how little information it contains. It does not tell us a great deal about moral properties or how we come to know what moral properties are. But without it, Jackson claims, we could not discover what descriptive property is identical with the property of moral rightness. The *a posteriori* discovery rests on *a priori* assumptions about moral properties.

(b) and (f) are propositions that express the A-extensions of 'water' and 'moral rightness' respectively, and this is what makes them *a priori*. (c) and (g) are propositions that express C-extensions, and so they are *a posteriori*. If the core-aposteriorist wants to argue that moral functionalism fails as a core-apriorist theory, she must focus on A-extensions by claiming that

---

[48] Jackson (2003), p. 87.

they do not work as Jackson claims they do. We can set out the core-aposteriorist's options for criticising moral functionalism as follows:

1) Moral functionalism claims that A-extension propositions are *a priori*, and C-extension propositions are *a posteriori*.

2) The moral functionalist wants to claim that we use A-extension propositions to justify C-extension propositions. We can only accept the C-extension propositions that a particular descriptive property is a moral property if we accept the A-extension proposition that the moral property is whatever plays the appropriate 'moral property' role.

3) So the core-aposteriorist has three options:

   a) Moral A-extension propositions are non-determinate and so they contravene *Sufficient Determinacy*.

   b) Moral A-extension propositions are false and so cannot help justify any C-extension propositions.

   c) Moral A-extension propositions are really *a posteriori*, contrary to what moral functionalism claims.

The options for criticising moral functionalism are the same as the options for criticising Michael Smith's moral realism, and as with Smith's moral realism, it is easiest for the core-aposteriorist to choose option (c). A-extension propositions are really *a posteriori*, and so moral functionalism cannot show that *a priori* determinate core moral knowledge is possible. Explaining why options (a) and (b) are difficult to pursue helps to show why option (c) is preferable.

Option (a) implies that moral A-extension propositions cannot be used to justify determinate moral propositions, and if we add more content to them so that they can justify such propositions, the extra content turns them into *a posteriori* propositions. Take the following moral A-extension proposition:

MRP: The property of moral rightness is whatever plays the moral-rightness role

What exactly is the moral-rightness role? We are not told, and many people may have very different ideas about what the role contains, even if everyone agrees that a moral-rightness role exists. So the possibility opens up that even if MRP is true, it may be insufficiently determinate to help justify any particular determinate moral proposition. Suppose that the moral functionalist wants to use the following argument:

4) The property of moral rightness is whatever plays the moral-rightness role.

5) The property of maximising utility plays the moral-rightness role.

Therefore

6) The property of moral rightness is the property of maximising utility.

(4) is MRP, an A-extension proposition. (5) is *a posteriori*, discovered by examining people's moral commonplaces. Assume that the term 'maximising utility' has very precise application conditions. Does the argument (4)-(6) work?

The term 'the moral-rightness role' currently does not have very precise application conditions. People can reasonably disagree on what the moral-rightness role contains, and so the argument (4)-(6) will only work if 'the moral-rightness role' contains sufficiently precise application conditions so that (5) is justified. Unless we do this, then two disputants can discover everything that they can *a posteriori* about the property of maximising utility, and agree on all the *a posteriori* discoveries that they make, but still disagree on what exactly the moral-rightness role contains.

As it stands, then, (4) and (5) are not sufficiently determinate. Concerning (4) in particular, we need something like

4*) The property of moral rightness is identical with whatever plays the following role: [description of what the moral-rightness role contains]

(4) is altered to become (4*) by being made more determinate, and (4*) is a determinate proposition. The core-apriorist may claim that (4*) is still an *a priori* proposition, but we can reasonably wonder whether it really would be. The core-aposteriorist may now try to claim that

making (4) more determinate must involve discovering *a posteriori* what the content of the moral-rightness role is. One possible way to do this is to focus on the fact that folk morality is built up out of the commonplaces that people accept, and that many of these commonplaces are *a posteriori* ones. The core-aposteriorist may hold that we need to discover *a posteriori* the various moral commonplaces people hold to determine what the content of the moral-rightness role is.

Of course, the core-apriorist could simply refuse to accept this move. He could continue insisting that we have to work out *a priori* what the content of the moral-rightness role is. If we can do so, then (4) and (4*) are both *a priori*. The only way that the core-aposteriorist can refute this claim is to actually investigate what the content of the moral-rightness role is, and this may take a lot of effort. The alternative is to declare a stalemate, which is not enough if the core-aposteriorist wants to show that moral functionalism cannot support the CAT.

Option (a) may look attractive, but it turns out to require a lot of work to see whether it is actually successful. What about option (b), that moral A-extension propositions are all false?

Moral A-extension propositions may be false because we are mistaken about the nature of moral properties, and hence about what 'moral property' roles should be. Suppose a scientist finds a new chemical that appears to dissolve plastic. For something to be this chemical, he believes, it has to play a role that includes the ability to dissolve plastic. It comes to be generally accepted that if something cannot dissolve plastic, it cannot be the chemical. But it is then discovered that the scientist has made a mistake, and the chemical does not dissolve plastic. The chemical role includes the ability to dissolve plastic, but the chemical itself cannot dissolve plastic, and so the chemical is not identical with whatever plays the chemical role.

This is not the only way in which we might be mistaken about the relation of X to the X role. Take the example of water. We say that water is transparent, drinkable, found in rivers, and so on, and to Jackson these properties constitute the watery role. It was not necessarily the case that anything has all these properties, though, or that there was only one thing that did. We might have examined oceans and found out that what we call 'water' is actually a mixture of H_2O and XYZ. Two things play the watery role, but can they both be water?

Another possibility is that the thing that plays the X role also has other properties that X itself does not have, and so the thing that plays the X role cannot be X. D.H Mellor argues that H_2O is

not identical to water, despite playing the watery role. Even if we allow that ice and steam are water, 'no single H_2O molecule can be water, since it instantiates hardly any of water's laws, having no solvent powers, density, freezing or boiling points, or latent heats'.[49] Of course, Mellor does not deny that water exists or that it is made up of H_2O. What he denies is that H_2O is identical to water, even though it plays the watery role.

It is tempting to think that we can make similar objections about the truth value of moral A-extension propositions. Take MRP again. We can try to pursue option (b) by questioning the truth of MRP in three possible ways.

First, it may be the case that it is impossible for anything to play the moral-rightness role. This could happen, for example, if the role is logically inconsistent. If the property of moral rightness exists, MRP cannot be true. Not only that, but the moral functionalist is left trying to find another way in which moral agents can recognise moral properties. MRP offered a straightforward way to do so – check what descriptive properties play a certain role, and those properties are moral properties. Now another method must be found.

This throws up two problems for the moral functionalist. The first problem is metaphysical – if moral properties are not the properties that play a particular moral role, then what are they? The second problem is epistemological – how can agents recognise moral properties if the properties are not what play a particular role? However, the moral functionalist need not worry yet, because it first has to be shown that it is impossible for anything to play the 'moral property' roles. The moral functionalist will of course deny that it is impossible, because we seem to identify moral roles all the time with the aid of the moral commonplaces that we accept. The 'fairnessy' role is the role of being even-handed, unaffected by biases, and so on. Where does the logical inconsistency lie in this? And if we can identify an inconsistency, why can we not just alter our conception of the role to exclude it, thus getting closer to a mature folk morality?

A second way that MRP might be false is if there is no special connection between the property of moral rightness and the moral-rightness role. Suppose that we discover a new non-moral property, Y, and we ascribe a number of characteristics to it that come to define the Y role. We then discover that Y has close to none of these characteristics. The property of Y is not what plays the Y role. Could this not be the case for the property of moral rightness? The

---

[49] Mellor (2003), p. 224.

trouble with this objection is that it once again does not allow for correction and refinement. As we find out more about Y, we refine our understanding of what the Y role is, so once we have realised our mistake we change our conception of the Y role to reflect the characteristics that Y actually has. Our conception of the Y role changes so that whatever plays the Y role is indeed identical to Y. This does not seem impossible, so the moral functionalist can say that same thing about moral properties. We may be currently mistaken about what the moral-rightness role is, but we can change our conception of it so that whatever plays that role is identical to the property of moral rightness.

A third way that we could argue for option (b) is to pursue a variation of Mellor's metaphysical objection. Something plays the moral-rightness role, and there may be such a thing as moral rightness, but the property that plays the moral-rightness role is not identical to the property of moral rightness. Moral-Mellor may say about MRP:

Assume that there is a property that plays the moral-rightness role. Let us say that the property of maximising welfare plays it. That certainly does not mean that such a property is identical to the property of moral rightness, because it may not play that role at all times and under all conditions (Remember that H_2O plays the watery role, but it does not always do so.). And we certainly cannot assume ahead of investigation that there is any single descriptive property that is identical to the property of moral rightness, or that there is a collection of descriptive properties that constitute the property of moral rightness.

Call this 'the Moral-Mellor objection'. It is consistent with the assumption that we can have moral knowledge. We may be able to know the proposition 'Maximising the welfare of the homeless is morally right', even though we do not identify the property of maximising utility with any particular moral property. The objection is also consistent with the idea that there are moral properties, that there are moral roles that descriptive properties play, and that we can know about them. What the Moral-Mellor objection rejects is the moral functionalist's assumption without sufficient investigation that moral A-extension propositions are true, and that particular moral properties are identical to particular descriptive properties.

At this point, the moral functionalist can reasonably object that despite the Moral-Mellor objection, we cannot simply assume that moral A-extension propositions are false. That would just state by fiat that option (b) is correct. But we now have a way to argue that option (c), that moral A-extension propositions are really *a posteriori*, is correct instead. H_2O is identical to water, says Jackson. Not so, says Mellor. Mellor's objection appears to be *a posteriori*. H_2O, as a matter of *a posteriori* fact, does not always act like water; for example, occurrences of H_2O need not have any boiling point, and we discover this fact *a posteriori*. By observation we attribute various characteristics to occurrences of water and of H_2O, and as a matter of *a posteriori* fact these characteristics are not always the same.

So it appears that we discover whether 'Water is the stuff that plays the watery role' is true or false by *a posteriori* observation. H_2O plays the watery role, but that does not necessarily mean that it is identical with water. We cannot discover *a priori* that it is water, and so we cannot discover *a priori* that water is identical to the stuff that plays the watery role. The proposition is therefore an *a posteriori* proposition, and we can argue analogously about moral A-extension propositions. We can only discover if, say, MRP is true by investigating *a posteriori* what people take the moral-rightness role to be, seeing what plays that role, and then checking whether it always conforms with the content of 'the property of moral rightness'. The proposition thus becomes harmless to the core-aposteriorist. If the proposition is true, we can only discover that it is true by *a posteriori* investigation. If it is false, then obviously it cannot help justify any true moral proposition. This means that it might be the case that option (b) works, and that the A-extension propositions that moral functionalism relies on are false, but we can only discover that option (b) works by first pursuing option (c). We discover *a posteriori* that MRP is false.

The moral functionalist might argue that the analogy between moral and non-moral A-extension propositions breaks down at a critical point. He might claim that whilst we discover *a posteriori* that water is not identical to what plays the watery role, this is not the case for moral properties. With water, we make the identification of water with whatever plays the watery role *a priori*, and we make the identification of water with H_2O *a posteriori*. However, he continues, we make the identification of a moral property with whatever plays the 'moral property' role *a priori*, and we can also make the identification of the moral property with the descriptive property that plays that role *a priori*. The moral functionalist thus changes moral functionalism to claim that

moral C-extension propositions are *a priori*, not *a posteriori*. So even if moral A-extension propositions are justified by moral C-extension propositions, we can know moral A-extension propositions *a priori*, since all moral C-extension propositions are knowable *a priori*.

But the moral functionalist is barred from making this move. The less important reason for this is that in Jackson's general metaphysics, he states that C-extensions are *a posteriori*. However, the moral functionalist can refuse to accept Jackson's general metaphysics whilst still accepting moral functionalism, and so this need not trouble him. The second reason, though, is more serious. We ascertain moral C-extensions by examining certain moral commonplaces, and these commonplaces are discovered *a posteriori*, by observing how actual agents behave and what moral judgements they make. Moral functionalism is committed to holding on to as many of these commonplaces as possible, so the moral functionalist is committed to saying that C-extension propositions are knowable *a posteriori*. Moral functionalism would change massively if this claim were dropped, so the moral functionalist requires good independent reasons to do so. If no such reasons exist, it follows that our discovery of moral C-extensions is based on *a posteriori* discovery of moral commonplaces, and if we can only ascertain moral A-extensions by ascertaining moral C-extensions, then the discovery of moral A-extensions is based on *a posteriori* observations as well.

Moral functionalism thus turns out to have a significant difficulty when we use the *a posteriori* strategy against it, as we have good reason to suspect that we can only verify the truth of moral A-extension propositions by using *a posteriori* investigation. Note that the *a posteriori* strategy does not imply that moral functionalism is completely implausible, because that is not its aim. The *a posteriori* strategy simply seeks to show that core-apriorist theories are wrong to claim that we can use *a priori* methods alone to justify the normative content of all true determinate core moral propositions. If a core-apriorist theory gives up that claim when it is put under pressure by the *a posteriori* strategy, it may be able to survive, even though it will no longer be a core-apriorist theory. Moral functionalism may be able to jettison the claim that moral A-extension propositions are *a priori* without losing very much. However, this means that we can no longer read moral functionalism as a core-apriorist theory, and it cannot be used to support the CAT.

We have now reached the end of the discussion of the four core-apriorist theories that I wished to test the *a posteriori* strategy on. I mentioned above that the theories have something in common: though the *a posteriori* strategy argues against their status as core-apriorist theories, it does not directly refute them. If the theories give up their commitment to the *a priori*, they may be able to survive without any trouble from the core-aposteriorist. For example, if Michael Smith changed his moral realism to allow that *a posteriori* moral platitudes as well as *a priori* moral platitudes are necessary for the mastery of moral terms, the core-aposteriorist would have no way to criticise him by using the *a posteriori* strategy. In this section, I will say a little more about how the *a posteriori* strategy works and why it will often have this result.

Take moral functionalism. Moral functionalism depends on various moral A-extension propositions being true. The propositions have a template:

Template: The property of <moral concept> is whatever plays the <moral concept> role

Moral functionalism tells us that when we slot various moral concepts into the template, the result is moral A-extension propositions that we can know *a priori*. One example is MRP. To discover what precisely plays the moral-rightness role requires *a posteriori* investigation, but we do not need to conduct that investigation to know that MRP is true.

The core-aposteriorist's objection to this was not that MRP or other moral A-extension propositions were false, but that we could not know that they were true without doing *a posteriori* work. We need to find out what plays the moral-rightness role before we can know that MRP is true, and the moral functionalist has conceded that we find out exactly what plays the moral-rightness role by *a posteriori* investigation. So MRP is only knowable *a posteriori*.

In other words, knowing the template and slotting different moral concepts into it as needed is not enough to show that various moral A-extension propositions are true. We must do further work. If they are true, then the core-aposteriorist simply points out that we cannot know that they are true *a priori*. The only way that we can perform this further work is to do so *a posteriori*. Alternatively, it might turn out that we have to alter the template in some other way so that it

works correctly and so that it can justify the normative content of determinate moral propositions. But, the core-aposteriorist points out, we can only discover *a posteriori* how the template should be altered.

This point generalises. The four core-apriorist theories we have examined all depend on some type of template that they claim can be ascertained *a priori*, but in fact the template is only useful if it is developed with *a posteriori* information. To make this claim, I need to define what I mean by 'template', since so far in this section the only template we have seen relates to moral functionalism. A template for a core-apriorist theory is either a general formula that generates core moral propositions (*Template* is one example) or the set of core moral propositions in the theory. This second option is a catch-all for core-apriorist theories that have core moral propositions that have nothing in common other than that they are ascertainable *a priori*.

Take the SLS theory and the self-evident axiom *Aim*. The SLS theory appears to have a template, which is Sidgwick's criteria to determine a set of self-evident axioms (Of course, the SLS theory claims that this task can be performed *a priori*.). However, *Aim* cannot be used to justify determinate moral propositions, since it contravenes *Sufficient Determinacy*. The answer was to add more content to it by selecting a particular first-order moral theory, such as rule utilitarianism. But there are *a posteriori* grounds on which to criticise and defend rule utilitarianism, such as the difficulty of being a rule utilitarian if the rules are too numerous, and so there are also *a posteriori* grounds on which to criticise and defend the altered version of *Aim*. Filling in the template with *a posteriori* information thus alters the template. In order to use *Aim* successfully as part of a justification of a determinate moral proposition, it must be interpreted in a particular way, with the help of a first-order moral theory. The first-order moral theory we select will alter how *Aim* works; we saw that a reading of *Aim* based on direct act utilitarianism looks very different to a reading of *Aim* based on rule utilitarianism. Furthermore, some readings will look completely implausible. Those based on direct act utilitarianism are one example. Very few people are direct act utilitarians, partly because of the sheer practical difficulty of being one successfully. It is an immense task to try to work out the consequences of every action and whom they affect positively and negatively.

This is similar to what we just saw with moral functionalism. *Aim* is generated by the SLS theory's template, but to use the template so that *Aim* becomes determinate, we also need to

add something else to it, a first-order moral theory.[50] For example, say I intend to drink-drive, because I want to enjoy myself at a party and I want to get home afterwards. I check *Aim*, and I discover that I am not aiming at good generally, as I am only aiming at my own good. So I should not drink-drive. However, in order to use *Aim* to say that I should not drink-drive, we need to use the 'other resources' of a first-order moral theory in order to interpret *Aim* in a plausible way. As I argued in chapter 3, the choice of the first-order moral theory is partially defended on *a posteriori* grounds. This does not mean that the template for the SLS theory is wrong. It means instead that if the template generates core moral propositions, we need to plug *a posteriori* information into it. Otherwise we cannot use the core moral propositions (which include *Aim*) to discover that I should not drink-drive.

Some parts of a template may still be *a priori*. The template for Michael Smith's moral realism is made up of the five types of moral platitudes that allow us to have mastery of moral terms. We saw that Smith's moral platitudes concerning objectivity and supervenience were the most likely platitudes to withstand criticism from the core-aposteriorist. It is very difficult to see how they could only be known *a posteriori*. Other moral platitudes that Smith relies on are much more vulnerable, though, and that is all that the core-aposteriorist needs. If there is a single determinate core moral proposition that is only knowable via an *a posteriori* or mixed method, then the Central Claim is true. Nevertheless, the part of the template that generates platitudes of objectivity and supervenience is unaffected, because such platitudes can still be known *a priori*.

With Matthew Kramer's moral realism as a moral doctrine, it is more difficult to discover a template because he does not explicitly give one. He does not list what the basic moral principles are, choosing instead to focus on only two of them, moral objectivity and moral supervenience. However, the *a posteriori* strategy can still be used against his metaethical theory's status as a core-apriorist theory. We have no particular reason to think that we can ascertain *a priori* everything that is morally possible or impossible, so we have no particular reason to think that we can ascertain *a priori* all true core moral propositions, or all true basic moral principles. Some of the basic moral principles, at least, may only be ascertainable *a*

---

[50] Actually, we may not need something as structured as a complete first-order theory such as utilitarianism. Maybe the theory could just be a collection of various moral norms. The point is that we need something that we can use to alter the template so that *Aim* becomes determinate. If a simple collection of various moral norms can do that, then that is fine.

posteriori, which is far from saying that there are no basic moral principles. The *a posteriori* strategy need say nothing about whether there are or not.

It now appears that we can give a generalised story about the *a posteriori* strategy, which goes as follows: the core-apriorist claims that there is a set of core moral propositions, and that the normative content of these propositions is knowable entirely *a priori.* The propositions are based on a template. We can use these core moral propositions to ascertain the normative content of all true non-core moral propositions. The *a posteriori* strategy does not claim that the template is wholly wrong, and therefore it does not need to claim that moral propositions endorsed by a core-apriorist theory are false. The template may be hopelessly mistaken, but that is a subject for a separate investigation. Instead, the *a posteriori* strategy targets the claim that the template generates all the determinate core moral propositions that justify the normative content of all true determinate non-core moral propositions. The strategy tries to show that the template cannot do this if it is only constructed *a priori*, because some determinate core moral propositions can only be known via an *a posteriori* method or a mixed method. If the template is to generate all the determinate core moral propositions that are needed, we need to alter the template appropriately, which requires the addition of *a posteriori* information. The core-apriorist's challenge is to create a template that does not have this problem.

We have now tested the *a posteriori* strategy to see how it operates and how successful it is against various core-apriorist theories. In the next chapter, I will briefly sum up the results and conclude that the strategy is a useful tool for the core-aposteriorist to use. I will also ask where the core-aposteriorist and the core-apriorist should go from here. If the *a posteriori* strategy works, what does this imply for metaethical theories?

Chapter 6 - Conclusion

1 – The Central Claim and the CAT

In this section, I will sum up the results of the last four chapters, and I will conclude that we have good reason to prefer the Central Claim to the CAT. The four metaethical theories we examined in chapters 3-5 all have problems when we read them as core-apriorist theories, because they appear either to covertly rely on *a posteriori* information, or they cannot justify the true normative content of some determinate core moral propositions.

The overall aim of this thesis was to defend the claim that there are at least some determinate core moral propositions that have normative content that can only be ascertained via *a posteriori* or mixed methods. In other words, the core-aposteriorist, who supports this claim, wants to prove that the Central Claim is correct. If the Central Claim is correct, then both parts (a) and (b) of the CAT, which is defended by the core-apriorist, cannot be correct.

In chapter 2, we saw one attempt to prove the Central Claim positively, by arguing that since we can only discover *a posteriori* what our moral cognitive abilities are, we can only discover *a posteriori* which determinate core moral propositions are true. This attempt was a failure, though, as there was no way to get the conclusion from the premises. I then developed a second strategy, the *a posteriori* strategy, which is a negative approach. Rather than developing particular *a posteriori* or mixed methods that must be used to attain determinate core moral knowledge, it tries to show that some of that knowledge cannot be attained via an *a priori* method. Since moral agents can still attain such knowledge, though, this implies that it can only be attained via an *a posteriori* or mixed method.

One reason why it is worth seeing if we can defend the Central Claim is because, as I noted in chapter 1, the CAT has been very popular historically amongst philosophers. Core-apriorists include philosophical heavyweights such as Kant, W.D. Ross and Derek Parfit. Conversely, core-aposteriorists have been comparatively few in number, and have not received as much attention. This position has been changing in recent years, but it is easy to see why philosophers have tended to be core-apriorists. Moral truths are normative, and historically it has been regarded as difficult or impossible to discover how we can know their normative

content *a posteriori*. *A posteriori* truths do not appear to be in the business of being normative truths. You can run as many scientific *a posteriori* experiments as you like on the development of morality in children, or how we actually make moral judgements, or at what point a species can be said to have developed a moral sense, but this will never tell you how to ascertain what is morally permissible. Even if you discover that all your test subjects reject something for moral reasons, they could just be wrong. For example, if humans usually make moral judgements emotionally rather than rationally, as Jonathan Haidt claims, this does not show that they are correct to do so or that they reach correct moral conclusions.[1]

This is not to say that core-aposteriorists necessarily make mistakes like this. However, it is sometimes difficult to see exactly what they are doing when they use *a posteriori* research from areas such as psychology, anthropology, sociology, and so on. The risk is that they accidentally appeal to such evidence illegitimately. For example, as chapter 2 noted, there is a real danger of accidentally conflating two distinct claims, which must be kept separate or the core-aposteriorist will be in trouble. They are

1) We can ascertain *a posteriori* what methods to use to attain [some type of] moral knowledge.
2) We can attain via *a posteriori* or mixed methods [some type of] moral knowledge.

The phrase '[some type of]' can be replaced with what the ethicist chooses; the core-aposteriorist will course replace it with 'determinate core'. The attempt in chapter 2 to use *a posteriori* research on cognitive abilities in order to reach an *a posteriori* method to attain determinate core moral knowledge is an example of what happens when one tries to use claim (1) as part of an argument for claim (2). This seemed to be, for example, part of what Flanagan, Sarkissian and Wong were attempting to do. There is in fact no necessary link between the claims, though it may be tempting to believe that there is. Claim (1) is easier to argue for. If actual agents can attain moral knowledge, we can observe which processes they use when they do so. However, for such an investigation to work, we have to already know which methods to attain moral knowledge are successful and which are not, and that may not be an *a posteriori*

---

[1] Haidt (2001).

matter. The *a posteriori* observations that we make of actual agents when they make moral judgements are not sufficient to determine what determinate core moral knowledge consists of, so we cannot say whether we attain such knowledge via *a priori*, *a posteriori* or mixed methods.[2]

The core-aposteriorist does not need to take any stance on (1), and she certainly does not have to claim that endorsing (1) is necessary for endorsing (2), or *vice versa*. However, (2) is much more controversial than (1), because of the perceived divide between the normative and the *a posteriori*. The core-apriorist and the core-aposteriorist may both agree that (2) is correct, but they would differ over how far it goes. The core-aposteriorist believes that there is some determinate core moral knowledge that can *only* be known via an *a posteriori* or mixed method, whilst the core-apriorist does not.

By using the *a posteriori* strategy, the core-aposteriorist does not defend her position by producing an example of a determinate core moral proposition and then demonstrating that it can only be known *a posteriori*. Rather, she uses the *a posteriori* strategy to show that, for some determinate core moral propositions, if we do know the proposition, we cannot know it via an *a priori* method. If we can use an alleged *a priori* method to ascertain the truth of the normative content of the propositions, that is because the method is not really an *a priori* method. If it really were an *a priori* method, we would not be able to use it to know the proposition, as its use would entail contravening *Sufficient Determinacy*.

So, concludes the core-aposteriorist, we must use *a posteriori* or mixed methods instead. It is true that the strategy does not automatically refute every core-apriorist theory. It cannot do so, because we can only use it against a core-apriorist theory once we know what the details of that theory are. We had to know what moral functionalism's *a priori* commitments were before we could use the *a posteriori* strategy to argue that moral functionalism cannot be a core-apriorist theory, and its *a priori* commitments are different to those of, say, the SLS theory. The core-aposteriorist's hope is that if the *a posteriori* strategy is consistently successful against a variety of core-apriorist theories, this will make the CAT increasingly implausible, and so the Central Claim will become more plausible. As we saw in this thesis, the *a posteriori* strategy causes a lot of trouble for the core-apriorist theories that we looked at, and so we have

---

[2] This point reminds us of Selim Berker's (2009) objection that Joshua Greene's (2007) *a posteriori* research actually has no normative implications. Berker claims that Greene's normative conclusions do not rely on the research that Greene cites at all. The research is thus normatively redundant.

evidence that the CAT is not as well-founded as the core-apriorist claims. Consequently, we have good evidence in favour of the Central Claim.

Furthermore, the core-apriorist cannot avoid this result by choosing between moral intuitionism and moral coherentism. The Central Claim does not presuppose that there are no self-evident principles or propositions; depending on the core-apriorist's particular intuitionist theory, the core-aposteriorist could agree that the *a priori* self-evident propositions that the core-apriorist identifies really are self-evident. What she denies is that the set of *a priori* self-evident propositions that the core-apriorist identifies is identical to the set of determinate core moral propositions. As for moral coherentist core-apriorist theories, even if the core-apriorist successfully gives us a maximally coherent set of determinate core moral propositions, the core-aposteriorist can, for example, argue that some of these propositions are only knowable *a posteriori*. Neither moral intuitionism nor moral coherentism will protect a core-apriorist theory from the *a posteriori* strategy. The core-apriorist must look elsewhere for a defence.

In section 1 of chapter 1, I claimed that my defence of the Central Claim does not imply that mixed methods are impossible or even unlikely. Though I have typically referred to *a posteriori* methods in this thesis rather than mixed methods, this is not to say that my arguments have any force against the possibility of mixed methods, and we can now see why. Suppose that we are testing whether the normative content of a particular determinate core moral proposition can only be ascertained via a mixed method. If so, it can only be ascertained with a mix of *a priori* and *a posteriori* premises. Might the use of the method contravene *Sufficient Determinacy*? Following *Sufficient Determinacy*, the proposition can only be determinate if the conjunction of the premises is determinate. This need not cause a problem. If the conjunction of the premises is non-determinate, we can in theory add more content to the premises until the conjunction becomes sufficiently determinate. Whether this extra content is knowable *a priori* or *a posteriori*, the mixed method remains a mixed method, since it still relies on both *a priori* and *a posteriori* premises. It thus does not become an *a posteriori* method or an *a priori* method. So we cannot reject mixed methods by appealing to *Sufficient Determinacy*.

Nothing I have argued in this thesis implies that mixed methods cannot exist, or that any method to ascertain the truth of the normative content of a determinate core moral proposition must be either an *a priori* method or an *a posteriori* method. It might be that a particular

purported mixed method really is an *a posteriori* method, which would mean that all the purported *a priori* premises that it uses must actually be *a posteriori*. But we have no reason to assume from the outset that they must be. The *a posteriori* strategy does not show that they are. Therefore, since *a posteriori* methods use *only a posteriori* premises, mixed methods cannot be *a posteriori* methods. Neither, of course, can mixed methods be *a priori* methods, which use only *a priori* premises.

Consequently, we cannot use the *a posteriori* strategy to argue against mixed methods. All we can do is examine the core-aposteriorist's actual metaethical theory. If the theory is correct and excludes mixed methods to ascertain the normative content of any determinate core moral propositions, then we have no reason to suppose that mixed methods exist. If, however, the core-aposteriorist's theory holds that we can only ascertain the normative content of some determinate core moral propositions with mixed methods, we will have to investigate the details of her theory to discover whether her use of mixed methods is valid. Perhaps the methods fail to work as the core-aposteriorist wants them to work, or perhaps premises that she takes as knowable *a priori* are actually only knowable *a posteriori*, but that will depend on the content of the particular theory itself, and as I have not examined any such theories in this thesis, I cannot claim that mixed methods should be rejected. So we should continue including mixed methods in the Central Claim.

Two questions now arise. The first is that if we give up the idea that the CAT is correct, what do core-apriorist theories actually lose? The second question is how we can use the core-aposteriorist's results to defend or amend particular core-aposteriorist theories.

2 – Where next for the core-apriorist and the core-aposteriorist

I will start with the first question. A core-apriorist theory can be much more than just a claim about the *a priori* status of determinate core moral propositions. It may include other epistemological claims, metaphysical claims, semantic claims, and so on. How badly damaged a core-apriorist theory would be overall, if the CAT were false, would depend on how heavily the theory relied on it. I pointed out at the end of chapter 6 that none of the four core-apriorist

theories we examined with the *a posteriori* strategy need fail if the Central Claim is correct. They simply need to be altered.

For example, suppose that Zangwill's objection that moral functionalism is relativist were a fatal one, or that Clark's 'horrified assertion' objection against Kramer were. In both cases, the metaethical theory targeted would simply fail to work, and this would suit the core-aposteriorist's purposes. If all core-apriorist theories uniformly fail, then the core-apriorist is no longer a serious opponent.[3] But by using the *a posteriori* strategy, the core-aposteriorist's aim becomes much more focussed. All she needs to do is refute the core-apriorist's commitment to the CAT. If the core-apriorist gives up this commitment, then the core-aposteriorist has made her point. She has no reason to refute the core-apriorist's theory wholesale. Obviously, the core-apriorist's theory will have to change if he gives up the CAT, but if he can make these changes successfully, the core-aposteriorist will not protest.

Neither is it necessarily the case that *all* true determinate core moral propositions are only knowable via an *a posteriori* or mixed method. Sometimes they may not be; for example, we saw that *Prudence*, one of the axioms that the SLS theory relies on, may be knowable *a priori*. Similarly, it was difficult for the core-aposteriorist to argue that Michael Smith's moral platitudes concerning objectivity and supervenience must only be knowable *a posteriori*. So we do not completely exclude the possibility of *a priori* determinate core moral knowledge. What we deny is that the normative content of all determinate core moral propositions is ascertainable *a priori*. Sometimes the normative content can only be ascertained via an *a posteriori* or mixed method.

Ultimately, then, the results of the thesis are only really going to trouble the core-apriorist if he wants to retain the CAT. He could decide that it does not matter to his overall metaethical theory if determinate core moral knowledge can only be attained *a posteriori*, or he could argue that a limited subset of such knowledge can be attained *a priori* (Perhaps knowledge about moral supervenience can only be attained *a priori*, for example.). However, what he cannot do is dismiss the possibility that one can only ascertain the normative content of a particular determinate core moral proposition via an *a posteriori* or mixed method simply because the method is *a posteriori* or mixed. The mere fact that it is is not a blow against it.

---

[3] It would not prove that the core-aposteriorist was right; maybe the moral error theorist would be, for example. Nevertheless, the challenge to the Central Claim would no longer come from the core-apriorist.

This brings us to the second issue of where the core-aposteriorist should go from here. I have made no claims about what precisely the correct *a posteriori* or mixed method to ascertain a particular determinate core moral proposition would be. This was intentional. If I simply worked out an *a posteriori* or mixed method to ascertain the normative content of determinate core moral propositions, the danger would be that the core-apriorist could claim that we can do so with an *a priori* method as well. This would do nothing to help the core-aposteriorist. The *a posteriori* strategy gets around this difficulty not by providing an *a posteriori* or mixed method, but by showing that *a priori* methods cannot always work. There will always be some determinate core moral propositions that moral agents can know, but that they cannot know via an *a priori* method. This means that *a posteriori* and mixed methods are necessary for agents to know the propositions, even though the *a posteriori* strategy does not say anything about what the content of these methods are. The task now for the core-aposteriorist is to provide her own methods to ascertain the normative content of determinate core moral propositions. There are many different ways that such a theory could go, and I do not propose to start the task here.[4] However, we can see that the core-aposteriorist can use the *a posteriori* strategy independently of her own theory. Whatever the content of her own theory is, the *a posteriori* strategy can help to lay the groundwork for it. For example, the core-aposteriorist can use it in the following argument:

1) It is possible for actual agents to have moral knowledge (and thus be moral agents).

2) So they must have some method to attain moral knowledge.

3) It has often been theorised that moral agents can ascertain the normative content of core moral propositions *a priori*, and this has been seen as evidence that the CAT is correct.

4) But the CAT is not correct, because the *a posteriori* strategy provides good reason to believe that the Central Claim is correct.

5) So there must be some determinate core moral propositions that are only knowable via an *a posteriori* or mixed method.

---

[4] A couple of alternatives that the core-aposteriorist could endorse are Richard Boyd's moral realism (1988) and Peter Railton's moral realism (1986).

This argument makes no reference to the content of the specific metaethical theory that the core-aposteriorist endorses. It works independently to eliminate the core-apriorist's objections that core-aposteriorist theories cannot work, and it can do this before the core-aposteriorist introduces her own metaethical theory. The argument thus functions as a pre-emptive defence. It does not, of course, show that the core-aposteriorist's theory identifies the correct methods to attain determinate core moral knowledge, but the *a posteriori* strategy was never designed to do this. It merely helps to show that *a posteriori* and mixed methods are necessary in order to ascertain the truth of the normative content of some determinate core moral propositions, not to show which particular methods are correct.

Ultimately, the thought that there are *a posteriori* or mixed methods to ascertain the normative content of some determinate core moral propositions should not be an implausible or worrying one. Accepting that *a posteriori* or mixed methods exist does not entail that no *a priori* methods exist, and it certainly does not mean that philosophers must abandon ethical study to the scientists, anthropologists and sociologists. Rather, it means that when we examine how we attain moral knowledge, the area to explore is wider than originally thought. We do not need to limit ourselves to *a priori* methods. Instead, we must explore *a posteriori* and mixed methods as well. We cannot ignore such methods simply because they appeal to the *a posteriori*.

Bibliography

Adams, Robert Merrihew, (1976), 'Motive Utilitarianism', *Journal of Philosophy* 73(14), pp. 467-481

Aharoni, Eyal, Sinnott-Armstrong, Walter and Kiehl, Kent A., (2012), 'Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction', *Journal of Abnormal Psychology* 121(2), pp. 484-497

Audi, Robert, (2004), *The Good in the Right: A Theory of Intuition and Intrinsic Value*, Princeton: Princeton University Press

Berker, Selim, (2009), 'The Normative Insignificance of Neuroscience', *Philosophy and Public Affairs* 27(4), pp. 293-329

Blackburn, Simon, (1973), 'Moral Realism', reprinted in his (1993)

Blackburn, Simon, (1984), *Spreading the Word*, Oxford: Oxford University Press

Blackburn, Simon, (1985), 'Errors and the Phenomenology of Value', reprinted in his (1993)

Blackburn, Simon, (1988), 'How to Be an Ethical Anti-Realist', reprinted in his (1993)

Blackburn, Simon, (1993), *Essays in Quasi-Realism*, Oxford: Oxford University Press

Blair, R.J.R., (1995), 'A cognitive developmental approach to morality: investigating the psychopath', *Cognition* 57, pp. 1-29

Boolos, George, (1999), 'Must We Believe in Set Theory?', reprinted in his (1998), *Logic, Logic and Logic*, Cambridge, Massachusetts: Harvard University Press

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

Brandt, Richard, (1979), *A Theory of the Good and the Right*, New York: Oxford University Press

Brink, David, (1985), *Moral Realism and the Foundations of Ethics*, New York: Cambridge University Press

Broome, John, (2004), *Weighing Lives*, New York: Oxford University Press

Burgess, John P. and Rosen, Gideon, (1997), *A Subject with No Object: Strategies for Nominalistic Interpretation of Mathematics*, Oxford: Oxford University Press

Clark, Samuel, (2011), 'Review of *Moral Realism as a Moral Doctrine*', *Philosophical Quarterly* 61(243), pp. 425-427

Clarke-Doane, Justin, (2014), 'Moral Epistemology: The Mathematics Analogy', *Noûs* 48(2), pp. 238-255

Clarke-Doane, Justin, (2015), 'Justification and Explanation in Mathematics and Morality', in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol 10*, Oxford: Oxford University Press

Clipsham, Patrick, (2014), 'Does Empirical Moral Psychology Rest on a Mistake?', *Philosophical Studies* 170, pp. 215-233

Copp, David, (1997), 'Belief, Reason and Motivation: Michael Smith's *The Moral Problem*', *Ethics* 108(1), pp. 33-54

Cowie, Christopher, (2014), 'Why Companions in Guilt Arguments Won't Work', *Philosophical Quarterly* 64(256), pp. 407-422

Cowie, Christopher, (2016), 'Good News for Moral Error Theorists: A Master Argument Against Companions in Guilt Strategies', *Australasian Journal of Philosophy* 94(1), pp. 115-130

Crisp, Roger, (2006a), *Reasons and the Good*, Oxford: Oxford University Press

Crisp, Roger, (2006b), 'Hedonism Reconsidered', *Philosophy and Phenomenological Research* 73(3), pp. 619-645

Cuneo, Terence and Shafer-Landau, Russ, (2014), 'The moral fixed points: new directions for moral nonnaturalism', *Philosophical Studies* 171, pp. 399-443

Darwin, Charles, (2004 [1871]), *The Descent of Man, and Selection in Relation to Sex*, London: Penguin

Davidson, Donald, (1984), 'On the Very Idea of a Conceptual Scheme', in his *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press

Dawkins, Richard, (2006), *The Selfish Gene*, 30th anniversary edition, Oxford: Oxford University Press

De Brigard, Felipe, (2010), 'If you like it, does it matter if it's real?', *Philosophical Psychology* 23(1), pp. 43-57

De Lazari-Radek, Katarzyna and Singer, Peter, (2014), *The Point of View of the Universe*, Oxford: Oxford University Press

Dworkin, Ronald, (2006), *Justice in Robes*, Cambridge, Massachusetts: Harvard University Press

Dworkin, Ronald, (2011), *Justice for Hedgehogs*, Cambridge, Massachusetts: Harvard University Press

Enoch, David, (2011), *Taking Morality Seriously: A Defence of Robust Realism*, Oxford: Oxford University Press

Erdur, Melis, (2016), 'A Moral Argument Against Moral Realism', *Ethical Theory and Moral Practice* 19, pp. 591-602

Field, Hartry, (1989), *Realism, Mathematics and Modality*, Oxford: Wiley-Blackwell

Fine, Kit, (1995), 'Ontological Dependence', *Proceedings of the Aristotelian Society* 95, pp. 269-290

Flanagan, Owen, Sarkissian, Hagop and Wong, David, (2008), 'Naturalizing Ethics', in Sinnott-Armstrong (2008)

Gettier, Edmund, (1963), 'Is Justified True Belief Knowledge?', *Analysis* 23(6), pp. 121-123

Gibbard, Allan, (2003), *Thinking How to Live*, Cambridge, Massachusetts: Harvard University Press

Gibbard, Allan, (2006), 'Moral Feelings and Moral Concepts' in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics, Volume 1*, Oxford: Oxford University Press

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

Greene, Joshua, (2014), 'Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics', *Ethics* 124, pp. 695-726

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

Hills, Alison, (2003), 'The Significance of the Dualism of Practical Reason', *Utilitas* 15(3), pp. 315-329

Hooker, Brad, (2002), *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*, Oxford: Oxford University Press

Hume, David, (1998 [1751]), *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp, Oxford: Oxford University Press

Hume, David, (1999 [1748]), *An Enquiry Concerning Hume Understanding*, ed. Tom L. Beauchamp, Oxford: Oxford University Press

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

Hurley, Susan, (1989), *Natural Reasons*, Oxford: Oxford University Press

Jackson, Frank, (1986), 'What Mary Didn't Know', *Journal of Philosophy* 83(5), pp. 291-295

Jackson, Frank, (1998), *From Metaphysics to Ethics*, New York: Oxford University Press

Jackson, Frank, (2003), 'From H_2O to water: The relevance to a priori passage', in Lillehammer and Rodriguez-Pereyra (2003)

Jackson, Frank and Pettit, Philip, (1995), 'Moral Functionalism and Moral Motivation', *Philosophical Quarterly* 45(178), pp. 20-40

Joyce, Richard, (2016), 'The Evolutionary Debunking of Morality', in Joel Feinberg and Russ Shafer-Landau (eds.), *Reason and Responsibility: Readings in Some Basic Problems of Philosophy*, 16th edition, Boston, Massachusetts: Cengage Learning

Kant, Immanuel, (1996 [1785]), *Groundwork of the Metaphysics of Morals*, in his *Practical Philosophy*, translated and edited by Mary J. Gregor, Cambridge: Cambridge University Press

Kant, Immanuel, (1998 [1781]), *Critique of Pure Reason*, translated and edited by Paul Guyer and Allen Wood, Cambridge: Cambridge University Press

Kelly, Daniel, Stich, Stephen, Haley, Kevin J., Eng, Serena J. and Fessler, Daniel M.T., (2007), 'Harm, Affect and the Moral/Conventional Distinction', *Mind and Language* 22(2), pp. 117-131

Kitcher, Philip, (1984), *The Nature of Mathematical Knowledge*, New York: Oxford University Press

Kramer, Matthew, (2009), *Moral Realism as a Moral Doctrine*, Chichester, West Sussex: Wiley-Blackwell

Kramer, Matthew, (2017), 'There's Nothing Quasi about Quasi-Realism: Moral Realism as a Moral Doctrine', *Journal of Ethics* 21(2), pp. 185-212

Kripke, Saul, (1972), *Naming and Necessity*, Oxford: Blackwell Publishing

Langton, Rae, (1992), 'Duty and Desolation', *Philosophy* 67, pp. 481-505

Lillehammer, Hallvard, (2010), 'Method of Ethics and the Descent of Man: Darwin and Sidgwick on Ethics and Evolution', *Biology and Philosophy* 25, pp. 361-378

Lillehammer, Hallvard, (2011), 'The Epistemology of Ethical Intuitions', *Philosophy* 86(2), pp. 175-200

Lillehammer, Hallvard, (2016), '"An Assumption of Extreme Significance": Moore, Ross and Spencer on Ethics and Evolution', in Uri D. Leibowitz and Neil Sinclair (eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability*, Oxford: Oxford University Press

Lillehammer, Hallvard and Rodriguez-Pereyra, Gonzalo (eds.), (2003), *Real Metaphysics: Essays in Honour of D.H. Mellor*, London: Routledge

Mackie, J.L., (1977), *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin

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

Mellor, D.H., (2003), '*Real Metaphysics*: Replies', in Lillehammer and Rodriguez-Pereyra (2003)

Mill, John Stuart, (1987 [1871]), 'Utilitarianism' in John Stuart Mill and Jeremy Bentham, *Utilitarianism and Other Essays*, London: Penguin

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

Nagel, Thomas, (1979), *Mortal Questions*, Cambridge: Cambridge University Press

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

Nichols, Shaun, (2004), *Sentimental Rules*, New York: Oxford University Press

Nozick, Robert, (1974), *Anarchy, State and Utopia*, Oxford: Blackwell Publishing

Olds, James and Milner, Peter, (1954), 'Positive Reinforcement Produced by Electrical Stimulation of Septal Area and Other Regions of Rat Brains', *Journal of Comparative and Physiological Psychology* 47, pp. 419–427

Parfit, Derek, (1984), *Reasons and Persons*, Oxford: Oxford University Press

Parfit, Derek, (2011), *On What Matters*, volume 2, Oxford: Oxford University Press

Price, Huw, (2004), 'Naturalism Without Representationalism', in Mario De Caro and David Macarthur (eds.), *Naturalism in Question*, Cambridge, Massachusetts: Harvard University Press

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

Prinz, Jesse, (2008), 'Is Morality Innate?', in Sinnott-Armstrong (2008)

Quine, Willard Van Orman, (1969), 'Epistemology Naturalized', in his *Ontological Relativity and Other Essays*, New York: Columbia University Press

Quine, Willard Van Orman, (1980 [1953]), 'Two Dogmas of Empiricism' in his *From a Logical Point of View*, second edition, revised, Cambridge, Massachusetts: Harvard University Press

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

Rawls, John, (1999), *A Theory of Justice*, revised edition, Oxford: Oxford University Press

Richardson, Henry S., (1994), *Practical Reasoning About Final Ends*, Cambridge: Cambridge University Press

Ross, W.D., (1927), 'The Basis of Objective Judgments in Ethics', *International Journal of Ethics* 37(2), pp. 113-127

Ross, W.D., (1939), *Foundations of Ethics*, Oxford: Oxford University Press

Ross, W.D., (2002 [1930]), *The Right and the Good*, Oxford: Oxford University Press

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

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

Sidgwick, Henry, (1876), 'The Theory of Evolution in its Application to Practice', *Mind* 1(1), pp. 52-67

Sidgwick, Henry, (1879), 'The Establishment of Ethical First Principles', *Mind* 4(13), pp. 106-111

Sidgwick, Henry, (1981 [1907]), *The Methods of Ethics*, seventh edition, Indianapolis: Hackett Publishing Company

Singer, Peter, (1974), 'Sidgwick and Reflective Equilibrium', *The Monist* 58(3), pp. 490-517

Singer, Peter, (2005), 'Ethics and Intuitions', *Journal of Ethics* 9(3/4), pp. 331-352

Sinnott-Armstrong, Walter (ed.), (2008), *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*, Cambridge, Massachusetts: MIT Press

Smith, Michael, (1994), *The Moral Problem*, Oxford: Blackwell Publishing

Sober, Elliott, and Wilson, David Sloan, (1999), *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, Massachusetts: Harvard University Press

Southwood, Nicholas, (2011), 'The Moral/Conventional Distinction', *Mind* 120(479), pp. 761-802

Sreenivasan, Gopal, (2001), 'Understanding Alien Morals', *Philosophy and Phenomenological Research* 62(1), pp. 1-32

Stratton-Lake, Philip, (2002), Editor's Introduction to Ross (2002)

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

Wheatley, Thalia and Haidt, Jonathan, (2005), 'Hypnotic Disgust Makes Moral Judgments More Severe', *Psychological Science* 16(1), pp. 780-784

Williams, Bernard, (1973), 'A Critique of Utilitarianism', in Bernard Williams and J.J.C. Smart, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press

Williams, Bernard, (1979), 'Conflicts of Values', reprinted in his (1981), *Moral Luck*, Cambridge: Cambridge University Press

Williamson, Timothy, (2002), *Knowledge and its Limits*, Oxford: Oxford University Press

Zagzebski, Linda, (1994), 'The Inescapability of Gettier Problems', *Philosophical Quarterly* 44(174), pp. 65-73

Zangwill, Nick, (2000), 'Against Analytic Moral Functionalism', *Ratio* 13(3), pp. 275-286