General Arguments against Ethical Naturalism

Lewis Jed Brooks

Department of Philosophy

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Submission Date: November 2019.

Acknowledgements

I cannot really believe that I have finally finished this thesis. Thanks are due to my supervisors Jimmy Lenman and Yonatan Shemmer. Their advice, support and feedback has been extremely helpful and it is much appreciated. I am also grateful to many other members, both past and present, of the University of Sheffield's philosophy department. I will not mention many names for fear of forgetting to mention people I ought to, but particular thanks must go to Graham Bex-Priestly, Stephen Ingram, Miklos Kurthy, Alex Duval, James Lewis and Emma Bolton who were excellent sounding boards when I needed to run an embryonic or inchoate idea past them. I received financial support from the Arts and Humanities Research Council and training from the White Rose College of the Arts and Humanities; I am very grateful to both of these organisations for all they have done for me. Finally, I would like to thank my mum Carol and my partner Sophie without whose support I would not have been able to write this thesis.

Abstract

Ethical naturalists argue that normative properties are, in some sense, nothing over and above natural properties. Some philosophers think that all forms of naturalism must fail. They present, what Jonathan Dancy calls, 'Blockbuster arguments', which they think rule out every kind of naturalism. In this thesis I argue that these arguments do not succeed. In the first two chapters, I argue that both analytic and synthetic naturalists have good responses to G.E. Moore's open question argument. Derek Parfit, as well as other non-naturalists, have presented a number of supposedly separate arguments against naturalism which are not meant to rely on any considerations about meaning; I argue against these in chapters 3 and 4. Non-cognitivists argue that all descriptivist theories like naturalism fail to explain the necessary connection between normative judgements and motivation. I argue, in an oblique way, that the kind of motivational internalism needed to ground this argument is not a threat to naturalism in chapter 5. In chapter 6, I turn to what I take to be more serious problems for naturalists arising from Terry Horgan and Mark Timmons' Moral Twin Earth thought experiments. I argue first that tackling these thought experiments is a problem that all major metaethical theories share. The fact that everyone needs to address these problems makes more palatable the responses I then go on to develop. First, I argue that we might be able to accommodate these cases if we accept that people can disagree even if the semantic content of their utterances are not logically inconsistent. Second, I argue that we can debunk the intuitions that Moral Twin Earth cases are supposed to pump. I end by summarising what I have argued in the thesis.

Declaration

I, Lewis Brooks, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

List of contents

Introduction

A central metaethical question concerns the nature of normative properties. There is no easy way to define normativity and there is disagreement among philosophers over just which properties are the normative ones. For instance, some philosophers think that semantic properties like a word's meaning are normative.[1] Other philosophers think that being in a particular mental state, e.g. having a belief that P, is a normative matter.[2] I sidestep these issues by focusing primarily on paradigmatically normative properties such as moral and prudential properties like *goodness, badness, rightness, wrongness,* etc. I defend a naturalist view of normative properties against some influential arguments found in the metaethics literature.

Following David Enoch, I will understand ethical naturalism as the view that normative properties are, in a sense, nothing over-and-above natural properties.[3] This characterisation is deliberately vague because I wish, for the most part, to remain neutral over whether the naturalist should say that the relationship between normative and natural properties is one of identity, reduction, constitution, grounding, or something else.[4] Most of the time I will simply say that naturalism's central claim is that normative properties are identical to natural properties. I think that this claim can be defended against the objections I consider in this thesis, but if there is a better way to characterise naturalism then I would be happy to adopt it.

This characterisation is also vague in that it does not make clear what it means for a property to be natural. This is a large and difficult topic which I will largely skirt around as well. The standard account is that natural properties are those properties which could play explanatory roles in the natural sciences and psychology, and perhaps also the social sciences.[5] For instance,

---

[1] Kripke (1982) is the *locus classicus* for this position.
[2] e.g. Blackburn (1998, Appendix).
[3] Enoch (2011, ch.5). Graham Oddy (2005, Ch.1) does the same.
[4] I do discuss one way we could understand the relationship between normative and natural properties in chapter 1, but I do not rule out other ways that we could do this.
[5] Moore (1993) is an early proponent of this account.

since an object's mass will feature in the explanations of natural phenomena given by physicists, a property like, *has a mass of 1kg*, will count as a natural property on this account. I do not think that this will ultimately prove to be the best way to characterise natural properties. However, I also think that if ethical naturalism turns out to be true then normative properties will be among the properties that could figure in the natural and social sciences as well as psychology. Additionally, I think that the claim that normative properties are among the properties that figure in the natural and social sciences as well as psychology, is defensible against the arguments I am going to go on to consider. That being so, the standard account of what a natural property is will suffice for my purposes.

At first blush, naturalism appears to have several advantages over rival theories. I will mention just two. First, naturalism promises to vindicate many of our ordinary normative judgements with appeal only to a purely naturalistic ontology. This is in contrast to robust realist views, which have it that normative properties are *non-natural*, *sui generis* and *irreducibly normative*.[6] It is argued, rightly in my view, that such properties would be mysterious and objectionably 'queer'.[7] Relatedly, if normative properties are just natural properties then we might come to know that something is, say, morally right in the same kinds of ways that we learn which natural properties a thing has. This again contrasts with rival views, according to which we (sometimes) learn which normative properties a thing has using some kind of mysterious reasoning process or a special intuitive faculty.[8]

Second, naturalists have an easy time explaining why normative properties *supervene* on natural properties, or, why there could not be worlds which are the same in all natural respects but

---

[6] David Enoch (2011) offers, in my view, the best recent defence of this view.

[7] Mackie (1977, ch.1) is the *locus classicus* of this complaint.

[8] I add 'sometimes' in parentheses because rival accounts can allow that *much of the time* we learn what normative properties a thing has *derivatively*, i.e. by making inferences based on our knowledge of the other normative properties we know the thing to have. There must, however, be some way for us detect at least certain normative properties *non-derivatively* on these views, otherwise the process of normative deliberation could never get started. That *non-derivative normative deliberation* could not be understood in purely naturalistic terms.

which are different in normative respects. This normative supervenience claim is almost universally accepted; it just seems obvious that there could not be a world which was identical in every single natural respect to ours but in which, say, it was morally permissible to torture people for fun.[9] Since every set of properties trivially supervenes on any set of properties of which it is a subset, naturalists can easily explain why normative properties supervene on natural properties. Rival theories meanwhile must accept that supervenience is simply a brute metaphysical fact or, to quote Simon Blackburn, "an opaque, *isolated, logical fact, for which no explanation can* be proffered."[10]

Naturalism appears to be a promising theory, but many philosophers think that there is little hope of its proving to be true. In my view, establishing that naturalism is the correct view would require that we identify just which natural properties are identical to which normative properties, in a way that is consistent with, and make good sense of, the kinds of things we do and say using normative terms and normative concepts. Some philosophers disagree. They think that it is not necessary to examine particular identifications of normative properties with natural properties, since we know in advance that normative properties could not be natural properties. They present, what Jonathan Dancy calls, "blockbuster arguments", which are designed to refute every kind of naturalism.[11]

In this thesis, I look at these general anti-naturalist arguments and argue that they do not succeed. In the first two chapters, I argue that naturalists have good responses to G.E. Moore's open question argument. Derek Parfit, as well as other non-naturalists, have presented a number of arguments against naturalism which are not meant to rely on any considerations about meaning; I refute these arguments in chapters 3 and 4. Non-cognitivists argue that all descriptivist theories like naturalism fail to explain the necessary connection between normative

---

[9] Though see Sturgeon (2013) for some of the complexities and complications associated with characterising normative supervenience claims. I gloss over these issues for brevity's sake.
[10] Blackburn (1993, p.119), his italics. Dreier (2015) shows that this problem affects quasi-realist views as well.
[11] Dancy (2006)

judgements and motivation. I argue, in an oblique way, that the kinds of internalism needed to ground this argument are not a threat to naturalism in chapter 5. In chapter 6, I turn to problems for naturalists arising from Moral Twin Earth thought experiments. I argue first that tackling these kinds of thought experiments is a problem that all major metaethical theories share, for reasons that have not been appreciated in the literature. The fact that everyone needs to address these problems makes more palatable the responses I then go on to develop. First, I argue that we can accommodate these cases if we accept that people can disagree even if the semantic content of their utterances are not logically inconsistent. Second, I argue that if the first strategy does not work then we can always debunk the intuitions that Moral Twin Earth cases are supposed to pump. Having refuted the main general arguments facing naturalism, I provide, by way of conclusion, a brief summary of the thesis.

Chapter 1: The Open Question Argument Part 1

The most famous 'blockbuster argument' against naturalism is G.E. Moore's open question argument.[12] For many years, it was widely believed that Moore's argument had defeated naturalism for good, or at least refuted many significant versions of naturalism. Nowadays, his arguments are more often seen as flawed, particularly in Moore's original formulations of them. At the same time, it is also widely thought that Moore's arguments contain some important insight, development of which proves fatal to at least some forms of naturalism. In this chapter, I argue that the main responses that naturalists have given to even the strongest forms of the open question argument stand up to scrutiny.

§1

At the beginning of *Principia Ethica,* Moore sets out to find the meaning of 'good'.[13][14] He thinks that this is an important task because he takes 'good' to be the notion on which all ethical theory depends.[15] It is Moore's contention that 'good' cannot be defined. By saying that 'good' cannot be defined, Moore means that it cannot be *analysed*.[16] According to Moore, when we analyse some notion we describe the various parts of the object to which it refers.[17] For example, to analyse 'horse' we can describe the composition of a horse, which Moore does as follows:

> "[I]t has four legs, a head, a heart, a liver, etc., etc., all of them arranged in definition relations to one another".[18]

By describing horses in this way we can inform someone of the meaning of 'horse'. Terms like 'good' or 'yellow', however, denote only simple properties. The property *goodness* has no parts

---

[12] Moore (1993).

[13] ibid, p.58.

[14] I will put inverted commas around words to denote linguistic expressions (like 'good'), italicised words to denote properties (like *goodness*) and chevrons around words to denote concepts (like <good>).

[15] ibid. p.192

[16] ibid. p.5, p.61

[17] ibid. p.59.

[18] ibid. p.60

that we can enumerate in order to describe to someone who is unfamiliar with it what it is.[19] Because of this there is very little that we can say about *goodness*. Moore writes:

> "If I am asked 'What is good?' my answer is that good is good and that it is the end of the matter."[20]

Any attempt to analyse 'good' just mistakes *goodness* for some other, non-simple property.

It might be the case that all objects which are good, necessarily also share some natural property. For example, everything that is good might also cause pleasure, and if that were the case then it would be correct to say, 'Pleasure is good'. However, that would not mean that the property *goodness* was identical to the natural property *pleasurable*, nor would it mean that the term 'good' could be analysed as 'pleasure'. Moore thinks that naturalists get confused because claims like 'Pleasure is good' have the same basic form as identity claims like 'Water is H_2O'.[21] Because they fail to disambiguate the 'is' of identity and the 'is' of predication, they confuse *goodness* itself with some property that is distinct from *goodness*, but necessarily co-extensive with it. Moore famously calls this kind of mistake, "the naturalistic fallacy".[22]

Moore argues in a few different, but related, ways that 'good' cannot be analysed and that *goodness* is a simple property. Some of these arguments are variations on arguments I talk about in later chapters so I pass over them here. For instance, at one point Moore seems to think that we can tell that 'good' is distinct in meaning from any natural predicate by introspection. He writes:

> Everyone does in fact understand the question 'Is this good?' When he thinks of it, his state of mind is
> different from what it would be, were he asked 'Is this pleasant or desired, or approved?' It has a distinct
> meaning for him, even though he may not recognise in what respect it is distinct. Whenever he thinks of

---

[19] Moore refers to the property *goodness* as just 'good', but I do not follow him in this for clarity's sake. Additionally, Moore sometimes puts inverted commas around the word 'good' to denote the property *goodness* but sometimes he does not, so some of the passages I quote may be misleading, but I think I have provided enough context to avoid this.

[20] ibid. p.58

[21] ibid. p.66

[22] ibid. p.62. Confusingly, Moore also uses the expression 'the naturalistic fallacy' to refer to at least four other kinds of (what he considers to be) errors (Baldwin (1993), pp.16-20). Subsequent writers have used it in yet more ways, for example, Steven Pinker (2002, ch.8), identifies the 'naturalistic fallacy' with, "the belief that what happens in nature is good". In what follows, I ignore these other ways that 'the naturalistic fallacy' has been understood.

> 'intrinsic value' or 'intrinsic worth' or says that a thing 'ought to exist', he has before his mind the unique
> object – unique property of things by which I mean 'good'.[23]

The claim we can tell that *goodness* is distinct from natural properties by introspection is a central part of the normativity argument against naturalism. I return to this argument in chapter 3.

Moore also questions how philosophers with different ethical views could disagree if naturalism were true.[24] If one philosopher uses 'good' to mean 'pleasure' and another uses it to mean 'the object of desire' then it is puzzling how these philosophers could disagree over whether the claim 'good is pleasure' is true. One possibility Moore raises is that the philosophers are disagreeing about how the term 'good' is typically used, but if this is the case then they are not really having an ethical disagreement at all. Alternatively, one philosopher might just mean by 'good is pleasure' that pleasure is identical to pleasure, but if that is the case then there is no possibility of sensibly disagreeing with their utterance. 'Pleasure is pleasure' is trivially true, so if 'Pleasure is good' were identical to it, it too would be trivially true. Either way, the naturalist cannot make sense of ethical disagreement. I return to issues around naturalism and disagreement in chapter 6.

Moore's most famous argument against naturalism and the one I focus on in this chapter is, what has come to be known as, the open question argument. This argument focuses on the difference in *cognitive significance* of normative and natural terms/concepts. Moore argues that if 'good' could be analysed as 'pleasure' then the claim 'Pleasure is good' would have the same meaning as 'Pleasure is pleasure'.[25] Moore argues that these claims cannot have the same meaning because someone could doubt that the former claim is true but no one could doubt that the latter is true.[26] Since these claims cannot have the same meaning, the property *goodness* cannot be the same as the natural property *pleasurable*. The same applies to any other way of defining 'good'. If

---

[23] Moore (1993, p.68)
[24] ibid. pp.62-4
[25] ibid. p.64
[26] ibid. p.68

'good' could be analysed as some naturalistic expression, 'N', then 'N is good' would have the same meaning as 'N is N', but for no value of N is this the case. No one can doubt that whether any claim in the form 'N is N' is true, but it is always possible to doubt whether 'N is good' is true. Therefore, *goodness* cannot be a natural property.

Shortly after introducing this argument, Moore makes the same point in a slightly different way. He writes:

> "[W]hen we think we have a definition [of 'good'], we cannot logically defend our ethical principles in any way whatever, we shall also be much less apt to defend them well, even if illogically. For we shall start with the conviction that good must mean so and so, and shall therefore be inclined either to misunderstand our opponent's arguments or to cut them short with the reply, This is not an open question: the very meaning of the word decides it; no one can think otherwise except through confusion."

Moore here is pointing out that if 'good' could be analysed as, say, 'pleasure' then a question like, 'Is pleasure good?' would be *closed*, meaning that its answer could not be doubted except by someone who was conceptually confused. The question would be closed because it would follow trivially from the fact that 'pleasure' and 'good' have the same meaning that the answer to the question was 'Yes'. We may as well be asking, 'Is pleasure pleasure?'.

Moore thinks that every question of the form, 'Is N good?', will be *open*. For any possible analysis, N, we might offer it will always be possible for someone to doubt, without conceptual confusion, that the answer to, 'Is N good?', is positive. Since, 'Is N good?', is always an open question, it cannot be the case that 'good' can be analysed in natural terms. Since 'good' cannot be analysed into natural terms, *goodness* cannot be a natural property. The same argument applies, *mutatis mutandis*, to all other normative and natural terms.[27] Therefore, normative terms in general cannot be analysed in natural terms and that means that normative properties cannot be natural properties, or so the argument goes. In this chapter, I will focus on naturalist accounts of 'good'

---

[27] Moore targets the view, which he attributes to Jeremy Bentham, that 'right' has the same meaning as 'conducive to general happiness' (ibid., pp.69-72). In arguing for this view, Bentham commits the naturalistic fallacy, according to Moore.

in particular and put aside other normative terms. We could summarise Moore's argument as it applies to naturalist accounts of 'good' as follows:

(1) For every possible naturalistic analysis, N, of 'good', the question, 'Is N good?' will be open.

(2) If, 'Is N good?', is open, then 'N' cannot be an analysis of 'good'.

(3) If 'N' is not an analysis of 'good', then 'N' and 'good' do not refer to the same properties.

(4) Therefore, 'good' cannot refer to any natural property.

§2 - Partners in Guilt

Before going on to look at Moore's argument in more detail it is worth pointing out that the argument does not just apply to naturalist theories. In fact, it seems to apply to most metaethical theories. This seems to be a reason to doubt the soundness of Moore's argument.

Moore recognises that his argument does not just apply to views which say goodness is identical to some *sensible* property, i.e. a property we can detect through the senses. In the fourth chapter of *Principia Ethica,* Moore applies his argument to what he calls "metaphysical" theories. According to these theories, the term 'good' refers to some supersensible property. An example might be a form of theological voluntarism which holds that 'good' refers to the property of *being approved of by God*.[28] Philosophers who defend metaphysical theories like these are just as guilty of committing the naturalistic fallacy as naturalists. Moore writes,

> "Just as, when Mill supposed that 'to be good' *meant* 'to be desired', the question 'What is good?' could be and must be answered solely by an empirical investigation of the question what was desired; so here, if to be good means to have some supersensible property, the ethical question can and must be answered by a metaphysical enquiry into the question, What has this property? What, then, remains to be done in order to

---

[28] Assuming that God's will cannot be known by naturalistically respectable means. Moore uses as his examples the ethical theories put forward by the Stoics, Spinoza and Kant.

> destroy the plausibility of Metaphysical Ethics, is to expose the chief errors which seem to have led metaphysicians to suppose that to be good *means* to possess some supersensible property?"[29]

It is easy to see why Moore thinks this, since we can easily apply his open question argument to metaphysical theories. If *goodness*, the argument would go, were the same as *being approved of by God* then the predicate 'is approved of by God' would have the same meaning as 'good'. But these predicates cannot have the same meaning since we cannot doubt the claim, 'That which is approved of by God is approved of by God', whereas we can doubt the claim 'That which is approved of by God is good'. So, *goodness* cannot be the same as *being approved of by God*.

But if to be good is not to possess some supersensible property, nor is it to possess some natural property, just what is it? Here things get a bit more complicated. The only remaining option, as Moore sees it, is that normative properties do not exist but, somehow, still *are*. He compares normative properties to the number two, of which he says, "Two is somehow, although it does not exist."[30] The same is true of *goodness,* according to Moore. It does not exist, nor is it real, and yet it still *is* and various natural objects instantiate it.

Contemporary non-naturalists put forward accounts which are a bit more clear. They argue that the open question argument rules out all *reductionist* theories but leaves us with the view that normative properties could exist, if their nature was *sui generis* and *irreducible*. I think that this is the kind of account of normative properties that Moore had in mind.[31] However, this account is also a bit unclear because it is difficult to spell out just what it means for normative properties to be *sui generis* and *irreducible*. One solution is to say that normative terms are *sui generis* and *irreducible* in the sense that *they cannot be referred to using any non-normative terms*. This distinguishes the view from "metaphysical" theories because on those theories *goodness* can be referred to using non-

---

[29] ibid. p.174
[30] ibid. p.169
[31] This way of understanding talk of the existence of universals is along the same lines that Bertrand Russell talks about it (e.g. Russell (1990, ch.9)). Moore plausibly has the same view of the existence of universals as Russell.

normative expressions like "approved of by God".[32] But so understood most non-naturalists views will be as vulnerable as naturalism is to the open question argument, or so I shall argue very shortly.

Other alternatives to naturalism are views that say that Moore is correct to think that normative *concepts* do not reduce to natural concepts but wrong to think that they successfully refer to any property. Error theorists think that we *attempt* to refer to some irreducible normative properties with our normative terms but fail because no such properties exist.[33] Non-cognitivists, meanwhile, argue that normative terms do not try to refer at all but instead express something like a desire.[34] It does not take too much to see that Moore's argument applies, *mutatis mutandis*, to these theories too.

While these theories do not reduce normative concepts to any *non-normative* concepts; they do tend to analyse some normative concepts in terms of other normative concepts.[35] This is especially clear on Moore's own theory. Consider these remarks:

> "Every judgement in practical Ethics may be reduced to the form: This is a cause of a good thing."[36]

> ""right" does and can mean nothing but 'cause of a good result' and is thus identical with 'useful'"[37]

> "'I am morally bound to perform this action' is identical with the assertion 'This action will produce the greatest amount of good in the Universe'"[38]

---

[32] A complication with this solution is that it seems like we should be able to refer to *sui generis*, normative properties with *non-normative* expressions like: "the property that the Germany-language term 'gut' denotes". To solve this we need to finesse the account to say that normative properties are those properties that cannot be ascribed by non-evaluative expressions unless those expressions contain a non-evaluative phrase which denotes a normative property without ascribing it. An expression like "has the property that the German-language term 'gut' denotes" denotes a property without ascribing it, whereas "is approved of by God" both ascribes and denotes a property. This problem and its solution are lifted from Streumer (2011).

[33] Mackie (1977), Joyce (2001), Lillehammer (2004), Olson (2014).
[34] Ayer (1936), Stevenson (1944, 1975), Hare (1952, 1963), Blackburn (1984, 1993, 1998), Gibbard (1990, 2003). The possibility of quasi-realism complicates matters in ways I do not go in to here.
[35] An exception would be theories which deny that there are conceptual or analytic truths involving normative concepts. The fact that there are possible exceptions does not undermine my point however, since if there are no conceptual or analytic truths Moore's argument does not get off the ground and naturalists are off the hook.
[36] Moore (1993, p.196)
[37] ibid.
[38] ibid.

> "a virtue may be defined as a habitual disposition to perform certain actions, which generally produce the best possible results."[39]

Value fundamentalists, like Moore, think that we can understood 'right' and 'ought' in terms of 'good' and other evaluative terms/concepts. The open question argument can be used against this theory. We could argue that the claim, 'I am morally bound to perform this action', could not be identical in meaning to a claim, 'This action will produce the greatest amount of good in the Universe' on the same grounds that Moore uses to rule out naturalism. It seems impossible to doubt whether the claim, 'I am morally bound to perform actions that I am morally bound to perform', is true, but it is possible to doubt whether the claim, 'I am morally bound to perform actions which produce the greatest amount of good in the Universe', is true. Deontologists, for instance, do not typically believe that we are always morally bound to perform whatever action promotes the most good. According to Moore's own argument then, Moore's theory of 'right' cannot be correct.[40]

It is not just a problem for value fundamentalists like Moore, the open question argument also applies to the current orthodoxy, reasons fundamentalism. According to this view, concepts like <good> and <right> are fundamentally to be understood in terms of reason-concepts like <has most reason to>.[41] This is common ground between many non-naturalists, error theorists and non-cognitivists nowadays. We could argue against this theory in a Moorean vein. We could argue that a claim like "That was the best thing to do" could not possibly have the same meaning as "The balance of reasons favoured doing it" since it is possible to doubt whether the best thing to do is what the balance of reasons favours, but not possible to doubt whether the best thing to do is the best thing to do.

---

[39] ibid. p.221
[40] Something that Russell (1903) points out and Moore concedes (1968). cf. Hare (1954, pt. 2).
[41] See e.g. Parfit (2011a, p.38). Scanlon (2000, p. 97. 2014, p.2)

The same argument can be repeated for most accounts of normative concepts. They mostly involve, in some form or another, analysing some normative concepts in terms of others or the identification of one normative concept with another, or the reduction of one normative concept to another, where those analyses, reductions and identity claims are possible to doubt. The fact that there is so much disagreement amongst moral philosophers regarding whether reasons fundamentalism, or evaluative fundamentalism, or some other theory is correct suggests as much. So, most ethical theories are potentially vulnerable to the open question argument. 'Is X good?' is plausibly an open question whether X refers to a natural predicate like 'that which maximises happiness' or to a normative predicate like 'that which I have a reason to do'. So, the naturalist is in a similar position to the anti-naturalist. This should give the naturalist grounds to be optimistic that something has gone wrong with the open question argument; it seems more likely that the open question argument is flawed than that there is no correct account of normative language.

This *partners in guilt/partner in innocence* response would, however, not be much good if the anti-naturalist could respond to the open question argument in ways that the naturalist cannot. Since the open question argument is, as I have reconstructed it, valid, the way to reply to it is to reject one of the premises. So, the main replies to the argument would be:

> *Reject Premise 1*: Argue that, 'Is X good?', is in fact closed for some values of 'X'.
>
> *Reject Premise 2*: Argue that even if, 'Is X good?', is in fact always open that does not show that good cannot be analysed as 'X'.
>
> *Reject Premise 3*: Argue that even if 'good' cannot be analysed as 'X' that does not show that the property *goodness* is not identical to the property referred to by 'X'.

However, each of these lines of argument can plausibly be used to defend naturalism against the open question argument. Whether or not these responses are as plausible when applied to

naturalism as to other theories remains to be seen. I move on now to look at each of these replies in turn.

§3

Several philosophers have argued that Moore begged the question against the naturalist. Moore assumes that there is no naturalistic expression N for which the question. 'Is N good?', is closed.[42] This begs the question against analytic naturalists, since whether there is some analysis of 'good' on which the question, 'Is N good?', is closed is exactly what is at issue between Moore and the analytic naturalist (assuming for the moment that Moore is right that on any successful account of *goodness* questions like 'Is N good?' would be closed).

Moore's claim that questions of the form, 'Is N good?', are always open seems to be based on a generalisation from a small number of examples of possible naturalistic analyses. Moore considers whether 'good' might be analysed as 'pleasure', 'desired', 'approved of', 'tends to increase quantity of life' or 'being that which we desire to desire'. For each of these analyses, Moore claims that, 'Is N good?', is an open question. From that, Moore seems to infer that it will always be an open question whether something with certain natural qualities is good. If that generalisation were warranted then Moore could avoid the charge of begging the question against the naturalist since he would have independent grounds for believing that questions of the form, 'Is N good?', are never closed. The problem for Moore is that the generalisation does not seem warranted. Since Moore considers so few cases it seems hasty to infer that no analysis will work. We do not think in other cases of analysis - e.g. the analysis of 'free choice' or of 'knowledge' - that because five or so proposals have not been successful, no proposal could be successful.

Additionally, the proposed analyses that Moore examines are all quite simple. Moore looks at whether 'good' could be analysed by some single term or short phrase, but the analysis of 'good'

---

[42] Broad (1934), Frankena (1939, 1973), Sturgeon (2003), Finlay (2014)

might be much more complex than that. Frank Jackson, for example, offers an analysis of 'good' according to which it refers to the particular natural property that would make true a sufficient number of the normative claims that we would regard as platitudinous in particular idealised circumstances. On Peter Geach's analysis, 'good' picks out the relational property of meeting the standard which is most relevant in the conversational context. For example, if we call a toaster 'good' then on Geach's view, in ordinary conversational circumstances, we would say that that toaster met the functional standards we judge toasters by; that they toast bread evenly, safely, without consuming much electricity, etc. To infer that there is no analysis of good even of the complexity of the analyses that Jackson or Geach offer from the fact that 'good' cannot be analysed as 'pleasure' or 'desired', does not seem warranted.

To be fair to Moore, it should be mentioned that these more complex naturalistic analyses had not yet been proposed when he was writing and that he does quite a good job covering most of the major analyses that were on the table at the time.[43] It should also be mentioned that on none of the more complex analyses of 'good' that have been offered since Moore was writing does the question, 'Is N good?', appear closed to most who have considered them. Charles Pigden argues that the continued failure to find any analysis of 'good' on which questions like, 'Is N good?', are closed gives us good inductive evidence that there is no such analysis waiting to be found.[44] Since for every proposed analysis, N, that we have looked at so far the question, 'Is N good?', has been open we can be reasonably confident that for future proposed definitions we will look at the corresponding questions with also be open. Pigden admits that this weakens the anti-naturalist's position since we can never be wholly certain that there will not be some future case that runs counter to the trend observed so far. At the same time, we rely on and are confident about all

---

[43] See Nuccatelli and Seay (2007, p.12) for more detail.
[44] Pigden (2012).

kinds of results reached by induction, so the anti-naturalist has a large, though defeasible, advantage over the naturalist as the dialectic stands, or so Pigden argues.

The problem is that this is not a good induction, as Stephen Finlay points out.[45] The anti-naturalist is not merely saying that, since on analyses that we have looked at so far questions of the form, 'Is N good?', have been open, we know that on the *next* analysis we look at the corresponding questions will also be open, rather, they are arguing that *every* future question will be open. This inference is not licensed by the data. Since on the naturalist picture there can be at most one correct analysis of 'good', we would expect all but one of the indefinitely many questions of the form, 'Is N good?', to be open. The fact that questions of the form of, 'Is N good?', that we have so far considered have been open is not good inductive evidence that all such questions are open. That evidence equally supports the hypothesis that just one of the indefinitely many possible questions of the form, 'Is N good?' is closed and we have not yet come to consider it.

The anti-naturalist might point out at this point is that the analyses we have so far considered were not randomly selected; they were the product of a great deal of deliberation on the part of some very clever philosophers. Many philosophers have devoted a great deal of energy to finding analyses of normative terms and on none of them so far considered has the question, 'Is N good?', appeared to be closed. Some anti-reductionists have argued that the *best explanation* of this fact is that all possible analyses are in fact open.[46] This is the basis of the *abductive* version of the Open Question Argument. The idea is that we rule out analytic naturalism on the basis that it is much more plausible that there is no naturalistic analysis of 'good' than that there is an analysis which we just have not found yet, given our apparent failure to find any such analysis heretofore.

---

[45] Finlay (2014), Lenman & Lutz (2018).
[46] Smith (1994), Strandberg (2003), Ridge (2008)

Finlay argues that this argument faces its own induction problem since we cannot be confident that this is the best explanation until we consider all the other possible explanations.[47] I do not agree with Finlay on this point. Presumably, we have never considered *every* possible explanation of any data. There are, I take it, indefinitely many possible explanations, many of them outlandish and grossly implausible, which could, in principle, be proposed for any data. Nevertheless, we feel confident that in many cases we have secured the best available explanation. It would surely prove too much to rule out abductive arguments on the grounds that we cannot consider every possible explanation of any data.[48]

However, I do not think the hypothesis that there is no naturalistic analysis of 'good' offers the best explanation of the fact that on the analyses we have considered so far these questions have been open, for two reasons. First, it is doubtful that successful analyses will always be obviously correct to all competent language users. This means that there could be a naturalistic analysis of 'good', even if 'Is N good?' is always an open question. Taking this line is to reject the second premise of Moore's argument so I will return to this issue in the next section.

Second, even if we think that on any correct analysis A of some concept C, the question 'Is A C?' will be closed, the fact that we have not found such an analysis for 'good' yet should come as no surprise. As Finlay noted, naturalists expect that at most one instance of, 'Is N good?', will be closed, so the fact that on all the instances we have considered so far the question seems open is not in need of much explaining. Unless that is, we have particularly good reason to expect that we ought to have found the right analysis by now. I do not see that we do have good reason to

---

[47] Finlay (2014, p.14)

[48] Finlay's objection against abduction also applies to his own metaethical theory. Further down the page I just referenced Finlay writes, "My case for the end-relational theory will be that it provides the best explanation of our practices of normative speech and thought". We might ask Finlay how he can be confident that his theory does offer the best explanation given that the possibility remains that there is a better explanation of our practices of normative speech and thought that may be found. Of course, I am not suggesting that that is in fact a good objection to Finlay's theory, rather, I am highlighting that his objection to the abductive version of the Open Question Argument would apply to all abductive arguments including the one he makes for his own theory.

expect this. Though some very clever philosophers have tried to find analyses of normative terms, they do not number very many people and they have not been looking for all that long. There have been, at most, fairly few analyses A of philosophically interesting concepts C on which the question 'Is A C?' would be closed (some would argue that there have not been any). The debate concerning how to analyse <knowledge>, for example, is seemingly interminable and the literature on it is voluminous, but most philosophers do not take that as a reason to think that knowledge cannot be analysed.[49] Even analyses of seemingly humdrum concepts like <bachelor>, <chair> or <paints> are open to doubt.[50] The analysis of <bachelor> as being equivalent to <unmarried adult male> is often treated as the paradigmatic example of conceptual analysis but, as many philosophers have pointed out, cases like that of widowers and clergymen are possible counterexamples to it. If philosophers have not found an analysis of even <bachelor> on which the question 'John is X but is he a bachelor?' is not open, it is not surprising that we have not found a naturalistic analysis of <good> - which is potentially much more complex than <bachelor> - which makes, 'Is N good?', closed.

A bolder way that the analytic naturalist can respond is to deny that there is anything to explain by showing that we have in fact already found successful analyses of normative terms. Finlay, Paul Ziff, Judith Jarvis Thomson, Frank Jackson, Michael Zimmerman and others have put forward analyses of normative terms on which it is not outrageous to suggest that, 'Is N good?', is closed. It does not seem that we can dismiss out of hand that, after due consideration and philosophical deliberation, we will see that these accounts do find some natural term which is more or less what we mean when we say that something is good and that anyone who, after due reflection, doubts the analysis is, to some extent or other, at least a little bit confused about how the concept <good> works. I do not want to defend any particular analysis here but it is worth

---

[49] Though see Fricker (2007), who suggests that <knowledge> may not be analysable, partly on the basis that attempts to analyse it so far have been futile and unedifying.
[50] Quine (1951), Fodor (1981, 1998a, 2004), Fillmore (1982), Margolis and Laurence (1999, 2003, 2004).

pointing out that anti-naturalists do not usually take any time to show that on these particular analyses questions of the form, 'Is N good?', are open. Instead, they tend to just assume that, 'Is N good?' will always be open based on a very small number of simple toy examples, such as the view that 'good' means 'pleasant'. I am not claiming that any of the more complex analyses are correct, just that the open question argument does not show that all of these accounts are not. We cannot be confident that on all these accounts 'Is N good?' will be open until we examine them. So, the open question argument has not proven to be the "blockbuster argument" by which we can know that all naturalistic accounts fail without examining the particular details of individual analyses.[51]

To summarise, the naturalist is entitled to challenge Moore's claim that all questions of the form, 'Is N good?', are open. We have found very few, if any, analyses of philosophically interesting concepts, which are not open to doubt. If we are to remain open to the possibility that such analyses may still be found then there is no reason for us to think that the analysis of 'good' will be an exception. If we are happy to go along with Moore's assumptions regarding meaning then we could test naturalistic analyses according to whether they make, 'Is N good?', closed. We do not have sufficient reason yet to conclude that every naturalistic analysis will fail this test. It may be that one such analysis has already been presented which passes the test and even if not, for all that Moore says, there remains the possibility that some future analysis could.

§4

That said, I am, as it happens, doubtful that if there is an analysis of 'good' (or of 'knowledge', 'free choice', 'truth' and most other philosophically interesting terms) it will not be open to doubt. The analytic naturalist can argue that this is not a problem because successful analyses can

---

[51] Dancy (2006).

be non-obvious, at least in some cases. Adopting this approach is to reject the second premise of Moore's argument.

Many philosophers have linked Moore's argument to *the paradox of analysis*.[52] The paradox of analysis says that no informative analysis is possible. If, on the one hand, the analysans and analysandum have the same meaning then the analysis will be uninformative, but if, on the other hand, the analysans and analysandum do not have the same meaning then the 'analysis' is not really an analysis at all (though it may truly state that two things are identical).[53] Moore makes the same basic move with his open question argument; either, 'pleasure is good', is equivalent to the trivial claim, 'pleasure is pleasure', or, 'pleasure is good' relates two distinct concepts, undermining analytic naturalism. That two expressions could have the same meaning without it being obvious to competent speakers might seem *prima facie* surprising. It seems plausible that competence with a term consists in knowing its meaning. If X means F and Y also means F, and I know the meaning of both X and Y, then surely it should also be obvious to me that X and Y both mean the same thing.[54]

At this point, some naturalists go in for a kind of *partners in innocence* response.[55] We know that there are unobvious synonyms and unobvious analyses. It is everyone's problem to explain how this can be so. Whatever the correct account of unobvious synonymy and analyticity turns out to be, the naturalist can just plug it into their story. For all that Moore says, normative terms and natural terms might be among the terms which stand in a non-obvious analytic relation to one another. If the correct analysis, 'N', of 'good', is synonymous with 'good' but not obviously so, then it is no surprise that the question, 'Is N good?', would appear open.

---

[52] e.g. Black (1944), Langford (1968), O'Connor (1982).
[53] Black (1944, 1945), White (1945, 1948), Church (1946), Feyerabend (1956) and Lewy (1976) offer useful discussions of the paradox.
[54] Michael Dummett (1978, p.131) claims that this is an "undeniable" feature of meaning
[55] e.g. David Lewis (1988) and Smith (1994). Schroeder (2006) does not commit himself to a view of normative *concepts* but he suggests a similar response to the open question argument.

A classic example of unobvious synonyms are the terms 'furze' and 'gorse', which both refer to the same plant.[56] These terms are synonymous but, so the thought goes, someone could be competent with both terms and not know this. They might know that both terms refer to shrubs with yellow flowers, which are native to Western Europe and North Africa and which looks such-and-such a way, etc., while being unsure whether they refer to exactly the same kind of shrub or merely similar-looking shrubs. The upshot is that the question, 'I know that X is Furze, but is it Gorse?', could be open to someone who we would class as competent with the use of both 'Furze' and 'Gorse'. If that is correct then the same might be said of normative and natural terms.

Other cases concern singular terms and co-referring descriptions. An example from C.H. Langford is of the analysis of 'cube' as 'regular polyhedron with twelve edges'.[57] 'Regular polyhedron with twelve edges' seems to give a good analysis of 'cube' by reducing it to simpler parts, but someone could conceivably doubt that a cube is a regular polyhedron with twelve edges.

One might object to this by challenging the examples. It could be argued that competence with both 'furze' and 'gorse' requires possessing enough information such that one could not fail to realise that they were synonymous. This seems to set the bar for competence very high since we could possess rather a lot of information about the terms while still being open to the possibility of their referring to two different natural kinds rather than one. I do not think it matches what we would intuitively believe 'competence' to require, but we do not need to get bogged down in that issue. If the bar for competence with 'Furze' and 'Gorse' is set so high that one could not be competent with both without realising that they are synonyms, then the naturalist can argue that we are in a similar position concerning normative and natural terms. It might be that we possess

---

[56] Kripke (1979)
[57] Langford (1968)

enough information about both the normative terms and natural terms such that we are confident when they should apply, have a reasonable idea of what follows from them, etc. but that we are shy of knowing that they are synonyms and so technically not competent with them.

However, I do not think that these examples are ultimately very helpful for the naturalist. There seem to be limits on what competent speakers can believe about 'cube' or about 'furze' and 'gorse' for which we do not find analogous limits in the case of normative terms. If someone thinks that 'X does not have 12 edges and X is a cube' then we will start to doubt their competence with the term 'cube'. If they still believe that claim after being given time and opportunity to reflect upon the matter then our belief that they do not really understand what a cube is will grow much stronger. Similarly, one can be unsure whether all furze are gorse and be competent with the terms 'Furze' and 'Gorse', but it seems less likely that someone might be competent with both terms while thinking 'X is Furze and X is not Gorse'. If someone confidently thinks that furze shrubs are not gorse, then it seems unlikely that they really mean by 'Gorse' what they mean by 'Furze'.

By contrast, it seems possible that someone could think that some action was good, while also thinking that it does not, say, cause pleasure. The same seems to apply, *mutatis mutandis*, to other cases of normative and natural terms. Seemingly, we can be competent with normative terms while also believing that actions to which those terms apply either have or do not have particular natural properties, even after given chance for reflection. [58] The examples of 'cube' and 'furze' and 'gorse' then do not seem to be perfect analogies for the naturalist's purposes. There seem to be limits on what competent speakers can intelligibly believe about 'cube', 'furze' and 'gorse', at least after reflection, that do not apply to normative terms.

---

[58] Strandberg (2004) makes a similar point.

Better examples for the naturalist's purposes are cases of philosophical analysis. Concepts like 'knowledge', 'causation', 'free will', etc. are difficult to analyse. An analysis like, 'knowledge is true justified belief with no false lemma', seems neither obviously correct nor obviously incorrect to speakers who are competent with the term 'knowledge', even after reflection. The fact that proposed analyses are not obviously correct or incorrect suggests that if there are analyses of these concepts then they will be nonobvious. The naturalist can claim that the same is true of normative terms. Whichever way is found to explain how non-obvious analysis is possible could be appropriated by the naturalist to respond to the open question argument.

§4.1

A worry about the partners in innocence response it that it is by no means obvious that everyone must explain how analysis is possible. There are powerful arguments against the idea of conceptual analysis, at least as it has traditionally been understood. The failure to find any uncontroversial cases of analysis is used as evidence that conceptual analysis of philosophically interesting terms is bunk.[59] A second worry, is that even if we can show that there are analyses of terms like 'knowledge' there is no guarantee that the solution can be applied to normative terms. It would be good if analytic naturalists went beyond a simple partners in innocence response and said something about how analysis works.

One solution to the paradox of analysis is to disambiguate between two ways which someone could be said to know the meaning of an expression. One way is to have what we might call *conscious knowledge* of the meaning of an expression whereby we consciously associate with the expression, and would be able to report if asked, some description, or prototypical features, or (depending on your theory of meaning) whatever other semantic value might constitute the meaning of the expression. The other way of knowing the meaning of an expression might be a

---

[59] Fodor (1998a), Margolis and Laurence (1999, 2003, 2004).

case of *tacit knowledge* or of *knowledge-how*. This could consist of some kind of disposition or pattern of usage, which demonstrates that we are competent at using the expression, but which does not require that we would be able to report anything like a definition or analysis of the expression. Because the latter kind of knowledge does not require conscious acknowledgement this gives us a way to answer the paradox of analysis. If we only know the meaning of the analysans in some kind of tacit sense it becomes possible that an analysis might be informative to us, since something's being informative to us partially consists in its becoming apparent to us on the conscious level. [60]

This solution can be adapted so as to apply to the open question argument. The reason that, 'Is N good?', appears open to competent language users, might be that we only associate 'N' with 'good' in some tacit or unconscious way. In other words, 'N' and 'good' might only have the same meaning in virtue of the fact that we have the same kinds of dispositions in how we use these expressions. Since these dispositions do not require that we recognise that synonymous expressions are synonymous we can explain why 'Is N good?' looks open to us even though 'N' is an analysis of 'good'. Stephen Finlay takes a view along these lines. [61]

A second solution to the paradox of analysis is to deny that in successful analyses the analysans and analysandum are always synonyms. This makes some intuitive sense. Identity of meaning is a symmetrical relation whereas we tend to think of analysis in asymmetrical terms. For instance, we might say that 'justified true belief with no false lemma' is an analysis of 'knowledge', but we would not offer 'knowledge' as an analysis of 'justified true belief with no false lemma'. If analysis just uncovered synonyms then we would need to explain why analysis is asymmetrical.

This suggestion makes the problem of explaining how there can be informative analyses tractable. If 'knowledge' can be analysed as, but does not have the exact same meaning as,

---

[60] See Rey (1993) and Peacocke (1992).
[61] Finlay (2014, ch.1).

'justified true belief with no false lemma,' then it makes sense that we could be competent with the term 'knowledge' but not already know how it is to be analysed.

However, if analysis is not the uncovering of synonymity, what is it? According to Jeffrey King, analysis discovers the structure of the properties referred to by the various terms or concepts that are being analysed.[62] If we discover that 'bachelor' can be analysed as 'unmarried adult man' then we discover that the property *being a bachelor* has as its components the properties *being unmarried*, *being an adult* and *being a man*. These properties may in turn be composed of other properties. There is not space for a defence of this kind of account of properties but it is worth noting one of the main advantages of this account; it offers a straightforward explanation of property entailment.[63] If *being a bachelor* has as one of its components the property *being a man* then it is easy to see why everything that is a bachelor is necessarily a man.

King's solution to the paradox of analysis suggests a way for the analytic naturalist to refute the open question argument. They could maintain that 'good' can be analysed as, say, 'causes pleasure', but that these expressions are not strictly synonymous. Since these expressions do not have identical meaning, a competent language users could think that, 'Is that which causes pleasure also good?', is an open question. Someone could understand the meaning of 'good' without necessarily recognising an analysis of it when they see it, since the analysans would have a separate meaning.

A potential problem with the analytic naturalist adopting King's account of analysis is that it might seem to go against the motivation for being an analytic naturalist in the first place. On King's account 'being alcohol' can be analysed as 'being composed of C2H5OH molecules' since the property referred to by 'being alcohol' is plausibly composed of the simpler properties referred to by 'being composed of C2H5OH molecules'. This kind of analysis can only be

---

[62] King (1998).
[63] For a full defence of this account see King (1998) and Schroeder (2007, ch.5).

uncovered by scientific investigation however. The analytic naturalist does not want to say that identity claims between normative and natural terms can only be uncovered by scientific investigation; they want to make sense of the idea that normative inquiry is, to a significant extent, an *a priori* process.

Fortunately, King makes some distinctions between different kinds of analysis which are helpful for the analytic naturalist. He distinguishes between *scientific analyses* (like the alcohol example), *trivial analyses* (like the analysis of 'brother' as 'male sibling') and *philosophical analyses* (like the analysis of 'knowledge'). These kinds of analysis are distinguished from one another along epistemic lines. Competence with a term that has a trivial analysis requires that one would recognise that the property picked out by the analysandum is composed of the properties picked out by the analysans. For example, to be competent with 'brother' requires that one would be able to recognise that, 'X is a brother but is x a male sibling?', is a closed question. Philosophical and scientific analysis meanwhile do not require this. However, philosophical analyses do require that competent language users *might*, at least in principle, be able to recognise that at least some of the component properties picked out by the analysans constitute the property picked out by the analysandum by reflecting sufficiently on their own usage of the term and on their intuitions about when the term correctly applies.

This, I take it, is roughly what analytic naturalists want to say about competence with normative terms. They think that we can arrive at an analysis of normative terms by reflecting on our normative thought and practices; a process which is complicated enough that one need not recognise a correct analysis as correct when one sees it. Because analyses can be non-obvious on this model, questions of the form, 'Is N good?', could be open even if 'N' was an analysis of 'good'. So, King's account of philosophical analysis is a good model for the analytic naturalist to

adopt when responding to the open question argument. It allows room for *a priori* normative

theorising but also gives a ready explanation of why intuitively open questions are open.[64]

As we have seen, there are at least two promising models of analysis on which non-obvious

analysis looks possible. Either of these can also be used to explain why intuitively open questions

are open. If, however, neither of these models proves to be any good, the naturalist can still

always retreat back to a partners in innocence response. Moore never managed to solve the

problem of analysis so he would have no way of barring this move.[65] We have no reason for

thinking that the solution to the paradox of analysis will not be open to the analytic naturalist.

Alternatively, if there is no solution to the paradox and analysis is not possible, then we could

reject the third premise of Moore's argument. If concepts were not, in general, analysable then

we would have no reason to maintain that distinct concepts cannot refer to the same properties.

I move on to consider the third premise of Moore's argument now.

§5

The third premise of Moore's argument stated:

> (3) If 'N' is not an analysis of 'good', then 'N' and 'good' do not refer to the same
>
> properties.

Perhaps the most popular response to Moore's argument has been to reject this premise.[66] We

might go about doing this a few ways, which I go through below.

§5.1 – Simple Natural Properties?

---

[64] Once we plug the story about philosophical analyses into King's account it is, I think, in essentials very similar to the first solution. Finlay acknowledges his debt to King so this is probably no surprise.

[65] Moore (1968) admits as much.

[66] Durant (1970), Harman (1977), Putnam (1981, p.208), Lycan (1988), Sturgeon (1985), Boyd (1988/1995), Brink (1989), Railton (1989, 1993), Schroeder (2007).

Moore argues that because 'good' is unanalysable, *goodness* must be a simple property. That, however, does not show that *goodness* is *non-natural* since, as Moore accepts, there are natural properties which are simple. Moore's argument does not block a kind of naturalism that says that 'good' is unanalysable but that *goodness* is a simple, natural property.

Moore recognises the possibility of this kind of reply but pledges to show that it cannot work later in *Principia Ethica*.[67] However, the argument he comes back with is unconvincing. Throughout the book Moore moves between different ways of understanding what counts as a 'natural', but he eventually settled on the idea that natural terms and natural properties are the kinds of terms or properties that are countenanced in the natural sciences and psychology.[68] The kinds of properties that natural sciences and psychology study, Moore says, are those that exist in time.[69] Moore argues that we can test whether *goodness* exists in time, and hence whether *goodness* is natural, in the following way:

> "Can we imagine 'good' as existing by *itself* in time, and not merely as a property of some natural object? For myself, I cannot imagine it, whereas with the greatest number of properties of objects – those which I call the natural properties – their existence does seem to me to be independent of the existence of those objects. They are, in fact, rather parts of which the object is made up than mere predicates which attach to it. If they were all taken away, no object would be left, not even a bare substance."[70]

This is a flawed test. I cannot imagine *any* property as existing by itself in time. The property *being six feet tall* is a natural property that many objects have but it is scarcely possible to imagine the property *being six feet tall* as existing by itself in time, separate from any object with a particular height. I can also make little sense of the idea that we might take away the property *being six feet tall* from an object and in doing so remove one of that object's parts but leave the rest of its properties. So, Moore's test is no help in distinguishing natural from non-natural properties. Consequently, he has no basis for saying that *goodness* could not be a simple, natural property.

---

[67] Moore (1993, p.65).
[68] ibid., pp.13-5, p.92.
[69] ibid, p.93
[70] ibid.

Nicholas Sturgeon has argued that *goodness* is a natural property but that it cannot necessarily be denoted by, or analysed in terms of, purely non-normative terms.[71] If *goodness* is causally efficacious (as Sturgeon thinks it is) then it is broadly a natural property, but there is no guarantee that all natural properties can be denoted by a term from the natural sciences or psychology. It may be that *goodness* is too multiply realisable in scientific or psychological terms to be denotable by anything other than 'good' and equivalent ethical terms. Moore has no way to refute Sturgeon's position.

There is, however, an argument that has been made by Frank Jackson and others that convinces me that if normative properties were natural properties then they could be picked out by non-normative natural predicates.[72] I talk about this argument in the next chapter. I said earlier that the best way for a non-naturalist to understand talk of normative properties being *sui generis* and *irreducible* was that they could not be referred to by any non-normative predicates. This means that if Jackson's argument is successful then normative properties are not *simple* in the relevant sense. Since I think that Jackson's argument is successful, I prefer not to argue that normative properties could be simple in order to refute Moore's argument so I move on instead to a different response.

§5.2 – Frege's Puzzle

I talked in the last section about the parallels between the open question argument and the paradox of analysis. Another problem regarding meaning which closely resembles the open question argument is *Frege's puzzle*.[73] Intuitively, claims like, 'Hesperus is Phosphorus', and ,'Hesperus is Hesperus', differ in meaning, even though 'Hesperus' and 'Phosphorus' are co-referring. Someone could doubt that, 'Hesperus is Phosphorus', is true (many people did), but no

---

[71] Sturgeon (2005).
[72] Jackson (1998).
[73] A similarity noticed by Church (1946), Kalderon (2004), Schroeder (2007) and Cullison (2008).

one could doubt that, 'Hesperus is Hesperus', is true. That being so, having the same reference cannot be sufficient for two names to have the same meaning, or so the argument goes.[74]

The open question argument is more or less the same argument but applied to predicates rather than names. Moore challenges us to explain how, if *goodness* is identical to the natural property *pleasurable*, it is possible to doubt that, 'Pleasure is good', is true. If *goodness* is the same as *pleasurable*, then, 'Pleasure is good', should be equivalent in meaning to, 'Pleasure is pleasure'. We know that these claims are not equivalent in meaning since we can doubt the former claim but not the latter. The same applies, Moore argues, to all identity statements containing normative and natural terms. So, *goodness* cannot be a natural property.

I suggested in the last section that in response to the open question argument, the analytic naturalist might just adopt a *partners in innocence* response, arguing that since everyone needs an account of non-obvious synonymy to answer the paradox of analysis anyway, the analytic naturalist is no worse off than anyone else. With Frege's Puzzle in mind, the naturalist might adopt a different *partners in innocence* response. They might argue that since everyone needs some account of how it is that we can doubt identity statements, the naturalist is no worse off than anyone else. We could solve the open question argument by adapting whichever solution we find to Frege's Puzzle.[75]

That being said it would, again, be more satisfying if we had something more positive to say. Frege's solution to his puzzle was to suggest that the meaning of an expression is not exhausted by its referent. This opens up the possibility that normative and natural terms might have different meanings but still have the same reference. Just as 'Hesperus' and 'Phosphorus' differ in meaning even though they both refer to Venus, so might 'causes pleasure' and 'good' differ in meaning even though they both refer to the same natural property.

---

[74] Frege (1997).

[75] This is Mark Schroeder's main response to the open question argument (2007, ch.5).

This suggestion could be used to explain why, 'Is N good?', is an open question. Where two expressions refer to the same property but are not identical in meaning, it is possible for competent language users to doubt identity statements containing those terms. The classic example used in the literature is, 'Water is H_2O'. Just as, 'Is Water H_2O?', could appear open to people who were competent with both 'Water' and 'H_2O', so, 'Is N good?', could appear open to people who were competent with both 'N' and 'good'. These forms of naturalism can explain the intuitive openness of questions like, 'Is N good?', on the grounds that on their view normative *concepts* are not reducible to (and hence have a different cognitive significance to) natural *concepts*, even though normative properties are natural properties.

This is a different response to the open question argument to those that I have considered up to this point. Previously, I suggested ways that the naturalist could maintain that normative concepts are analysable in natural terms. Now, the suggestion is that the naturalist might agree that we could not discover that, say, *goodness* is *causing pleasure* by analysis alone, but still maintain that normative properties are natural properties. It cannot be discovered by analysis alone that, 'Water is H_2O', is true but that does not make us doubt that *being water* and *being H_2O* are the same property. The naturalist can say that the same is true in the case of identity claims containing normative and natural terms.

§5.3 – Horgan and Timmons' Objection

There have been some interesting challenges to naturalist theories which argue that normative and natural terms might work like the, 'Water is H_2O', case. Terry Horgan and Mark Timmons, have argued that the open question argument works even against these forms of naturalism.[76]

On 2-dimensional semantics, an a-intension is a proposition which fixes the referent of a particular term. The proposition depends on various folk platitudes concerning the term in

---

[76] I consider some of Horgan and Timmons' other arguments in chapter 6.

question. On Frank Jackson's 2-dimensional semantics, for instance, the a-intension of the term, 'Water', is something like 'Water is the clear potable liquid that fills our oceans and lakes'.[77] On this view, knowledge of the a-intension of a term is necessary for conceptual competence with that term. A result of this is that though we can doubt that 'Water is H_2O' is true, we should not be able to doubt that 'Water' denotes the clear, potable liquid that fills our oceans and lakes. Another way to put this might be to say that, while 'X is water, but is X H_2O?' may be an open question, 'X is water but is X the clear potable liquid that fills our lakes and rivers?' is not.

At this point Horgan and Timmons point out that the same is not true in the normative case. Jackson assigns to 'good' the a-intension 'the property that would play the goodness-role in mature folk morality'. Fulfilling the goodness-role requires making true enough of the claims about 'good' that we would regard as platitudinous in certain idealised circumstances. In other words, 'good' refers to whatever natural property makes true as many of the platitudes of "mature folk morality" as possible. For example, in mature folk morality we might deem that 'killing someone is not typically a good thing' is a platitude. Whichever natural property would, if it were the referent of 'good', make true as many of these mature folk morality platitudes as possible, is the natural property that is identical to *goodness* on Jackson's account. For example, killing someone does not typically cause people pleasure, so the natural property *pleasurable* would make true the platitude 'killing someone is not typically a good thing to do'. If the supposition that 'good' denotes *pleasurable* also makes true more of the mature folk moral platitudes than the supposition that 'good' denotes some other natural property, then *pleasurable* is the natural property denoted by 'good'.

Horgan and Timmons argue that if such an account were correct then the question, 'X has the property that would play the goodness-role in mature folk morality, but is X good?' would be a closed question, just as, 'X is water but is X the clear potable liquid that fills our lakes and

---

[77] Jackson (1998)

rivers?', is. But, the normative question does not seem to be closed. It might be the case that people are normatively misguided, such that the natural property that would make true most of their mature folk morality's platitudes would not be *goodness*. It seems to be a contingent matter what natural property would fulfil the goodness-role in mature folk morality, but not a contingent matter what is good. For example, it might turn out that mature folk moralists are all utilitarians and that *pleasurable* fulfils the goodness-role. This would not settle the matter of whether good is that which causes pleasure. In other words, it would still be an open question whether utilitarianism was true. So, the 'Water is H_2O' example is not a good analogy for the case of normative and natural terms.

That may seem to be just be a problem with Jackson's particular semantic account. But Horgan and Timmons think that it generalises to any account of how the reference of 'good' gets fixed.[78] For example, on causal theories of meaning *goodness* is the natural property that causally regulates our use of the word 'good'. But, 'X is good but does X have the property that causally regulates our use of the word 'good'?' also seems to be an open question. It might be that our use of 'good' has been guided by some natural property which, in fact, many bad actions have.

§5.4 – Responses to Horgan and Timmons

There are three main responses to this argument. All these responses correspond to the three main responses to Moore's Open Question Argument so I do not dwell on them here lest I labour the point.

First, we could argue that on the correct account of the reference-fixing relation the following really will be a closed question: 'X is good but does X have the property R [the natural property that that fixes the referent of 'good'?] Or, we could at least argue that we are not in a position to know that all questions of that form are open. For all we know there might be some way of, for

---

[78] See Horgan & Timmons (2000, p.149).

instance, spelling out just what mature folk morality, such that we really would not be able to question whether the things that the mature folk moralists call 'good' are in fact good.

Second, we could argue that even though the question, 'X has the property that would play the goodness-role in mature folk morality, but is X good?' is open, that does not mean that it is not an analytic truth that *goodness* is the property that plays the goodness-role.[79] We might only know that that *goodness* is that which plays the goodness-role by doing the kinds of *philosophical analysis* that I spoke about in the last section, and philosophical analysis can be informative and non-trivial. Arriving at a correct semantic account by philosophical analysis will be a complex and controversial process, so it really would not be surprising if it was non-obvious.

Third, we could deny, *pace* Jackson, that questions like, 'X is water but is X the clear potable liquid that fills our lakes and rivers?', are closed or that that they are even answerable by analysis alone, preserving the analogy between the water case and the normative case. Stephen Laurence and Eric Margolis make a good case for this point:

> "Recall that being watery is, roughly, a matter of being clear, liquid, drinkable, falling from the skies as rain, filling the rivers, lakes, oceans, etc. But liquid can't be part of the a priori specification of its a-intension, since this would rule out water in its gaseous and solid forms. Similarly, whether water is clear depends to a large extent on the amount and type of impurities in a given sample, and since what counts as an impurity is itself subject to empirical investigation, clear isn't a better candidate than liquid. Consider also the situation of someone who grows up with no exposure to lakes or oceans - maybe a nomad in the Sahara. If Jackson's analysis is right, such a person should be able to know a priori that water fills oceans and lakes or else she lacks the concept WATER altogether. We take it, however, that neither of these possibilities is at all plausible, so one has to

---

[79] Jackson gives this response (1998).

cross off *fills the oceans and lakes* as well. In fact, we doubt that any of Jackson's purported "watery" characteristics are immune to revision."[80]

In other words, 'X is water but is X the clear potable liquid that fills our lakes and rivers?', really could be doubted by someone who was competent with the term 'water', so there is no disanalogy with the normative case.

This might seem just to be a problem with Jackson's theory of reference, but the same point can be applied to other theories of reference too. Take a Boyd-style causal theory of reference. When it comes to the normative case, Laurence, Margolis and Angus Dawson point out that we should not expect the question, 'X is good but does X have the property that causally regulates our use of the word 'good'?' to be closed.[81] Even if the causal theory of reference is correct, determining that it is so will be the result of a long, complex and contentious theoretical enquiry. The results of such an enquiry will undoubtedly not be obvious to every competent language user.

Nor need the a-intension of a term be something which we could know purely *a priori* through philosophical analysis. Though there is not space to do the matter justice here, it can be argued that semantic theories are really more like natural science hypotheses than they are the result of pure armchair reasoning.[82] Semantic theories are, on this view, empirical theories which explain how and why human beings use various sounds and symbols in the ways that they do. They are as open to revision by, and dependent on, empirical information as any other empirical theory. If that were the case, then there would be no reason to expect that every competent language user could come to recognise the correct theory of reference, even in principle, through philosophical analysis alone. If that is the case then then normative terms will not have an a-intension which all competent language users necessarily know (or are in a position to know) purely *a priori*. There

---

[80] Margolis & Laurence (2003, p.262)
[81] Laurence, Margolis & Dawson (1999).
[82] See Millikan (1987) and Dowell (2016).

would then be no disanalogy between the case of 'water' and the case of normative terms, so Horgan and Timmons' revival of the open question argument would fail.

The same responses that were available to Moore's original argument are also available to Horgan and Timmons' attempt to revive the argument. These responses are more or less as effective against Horgan and Timmons's argument as they are against Moore's, which is no surprise since the arguments are fundamentally very similar. Since, as I have argued, we have good reason to suppose that the standard responses to the open question argument stand up, we also have good reason to doubt Horgan and Timmons' variation on that argument.

§5.5 – Bedke's Challenge

Matthew Bedke offers a different and interesting challenge to the applicability of the 'Water is H_2O' example to the normative case.[83] He takes an argument from David Chalmers and Frank Jackson against certain reductive accounts in the philosophy of mind and applies it to reductions in metaethics.[84] Chalmers and Jackson are interested in 'phenomenal gap' accounts, which hold that brain states are identical to conscious states but that brain-state *concepts* are non-equivalent to consciousness *concepts*.[85] This supposed 'phenomenal gap' between thinking about brain states and thinking about conscious experience is used to rebut some of the standard objections to reductionism about consciousness, such as the conceivability of philosophical zombies. The apparent conceivability of zombies - which have brain states but no consciousness – is supposed to present a problem for reductionists. 'Phenomenal gap' accounts have it that we can conceive of beings with brain states but no consciousness because brain-state concepts and consciousness concepts have a different cognitive significance, even though they both refer to the same thing.

---

[83] Bedke (2012). Gibbard (2003, ch.5) makes a similar point.
[84] Chalmérs and Jackson (2002).
[85] See e.g. Papineau (2002), Perry (2001), Levine (2001).

The relevance of this kind of account to the open question argument is obvious.[86] If we can maintain that the difference between phenomenal and brain-states concepts can explain why we can judge that a question like, 'X has such-and-such brain states but is X having such-and-such a conscious experience?', is open, then we can use the difference between normative and natural concepts to explain why, 'Is N good?', is open.

In response to the phenomenal gap account, Chalmers and Jackson argue that if x-properties are identical to y-properties then knowledge of all the y-facts would suffice for knowledge of all the x-facts. They return to the example of water and H_2O. While the concepts <water> and <H_2O> are distinct, knowing all the H_2O-facts, plus possession of the concept <water>, would lead to us knowing all the water facts. A complete picture of all the H_2O-facts leaves out nothing relating to water. They go on to argue that, by contrast, we could know all the facts pertaining to people's brain states without knowing all the facts about their conscious experience. This suggests that brain-state concepts and consciousness concepts cannot really be picking out the same properties after all. The brain-state concepts leave out some important information that the consciousness-concepts contain. Therefore, brain states cannot be identical to conscious states.

In the same vein, Bedka argues that if we knew all the natural facts then we would not know all the normative facts, meaning that normative properties cannot be natural properties. On this picture, even if normative and natural concepts are distinct, they can only count as coreferring if there are certain semantic entailments between them. Particularly, it needs to be the case that an ideal agent could, given all the natural truths and the normative concepts, infer all of the normative truths. They should be able to do this by simply applying the normative concepts to the natural truths they have. So, if 'good' just referred to the natural property *pleasurable*, an ideal agent, given all the natural facts and given the application conditions for the concept <good>,

---

[86] Yetter-Chappell and Chappell (2012) and Laskowski (2019) both adapt the phenomenal gap account to respond to the open question argument.

should be able to determine all the facts about which things are good. But, Bedke alleges, an ideal agent clearly would not be able to do this. He concludes that we need to invoke properties beyond the purely natural ones to make normative claims come out as true.

§5.6 – Responses to Bedke

The naturalist has two main ways to respond to Bedka. First, naturalists could reject the claim that if we knew all the natural facts then we would not know all the normative facts, or to at least deny that we are in any position to know that this claim is true. If we knew all the natural facts, including all the facts about meaning, metaethics and metasemantics[87], then, for all we know, we may well come to know all the normative facts. I take it that normative naturalism has a non-zero chance of turning out to be true. That being so, there is at least a chance that once all the facts were in we would have learned that normative properties are natural properties, and so would be able to infer all the normative facts from our knowledge of natural properties. The assumption that we would not know all the normative facts upon learning all the natural facts may just be a failure of imagination. Our ability to accurately judge what we could infer once we knew every single natural fact is significantly limited at best.[88]

The other response is to challenge the claim that where x-properties are identical to y-properties, knowledge of all the x-facts would necessarily suffice for knowledge of all the y-facts. There are two main ways that we can understand fact-talk. If we understand 'facts' as being something like *states of affairs* then knowledge of x-facts will necessarily suffice for knowledge of y-facts if x-properties are y-properties. If some normative property is identical to a particular natural property then an action's having that normative property is the same state of affairs as that action's having the natural property. In that case, the normative fact that that action is, say, good, would be the same fact as the natural fact that that action, say, causes pleasure. However, if this

---

[87] Assuming, controversially, that facts about meaning, metaethics and metasemantics are natural facts.
[88] Bedke (2012, p.120) recognises the possibility of this reply but simply dismisses it.

is the case then knowledge of all the natural facts will suffice for knowledge of the normative facts, if naturalism is true. If some natural fact consists of the same state of affairs as some normative fact, then knowing that natural fact trivially suffices for knowledge the normative fact, since those states of affairs are one and the same. For example, if an action's causing happiness is the same state of affairs as its being good, and knowing a fact just requires knowing the that the states of affairs that it consists in obtains, then knowing that an action causes pleasure suffices for knowing that that action is good since these are the same facts. Of course, anti-naturalists will deny that normative facts are natural facts so they will not think that the knowledge that a particular natural state of affairs obtains will suffice for knowledge of a normative fact, but would they need an independent argument for this claim.

Bedke needs to understand talk of facts differently. He might instead understand 'facts' as referring to true interpreted sentences. We can know that one sentence is true without knowing that another sentence which refers to the same state of affairs is true if we are unaware that the terms used in those sentences refer to the same properties. Understood this way though knowledge of all the natural facts might not suffice for knowledge of all the normative facts. I might for instance be fully opinionated on which sentences, which use only natural concepts, are true of a situation without having any view about which sentences, which use normative concepts, are also true.

This need not mean that normative properties cannot be natural properties, however. Take the example of indexical concepts. In John Perry's classic example, a man pushes a shopping trolley around a supermarket trying to determine who has spilled a line of sugar on the floor.[89] Unbeknownst to him, a bag of sugar in his own shopping cart has a tear in it and he is in fact following a trail left by himself. The man seems to know that the claim 'The shopper with the torn bag of sugar is making a mess' is true, but he does not believe the claim 'I am making a

---

[89] Perry (1979)

mess', even though he is the shopper with the torn bag of sugar. The moral of the story is that we can have knowledge of all the non-indexical facts without having knowledge of all the indexical facts.[90] However, we do not take that to mean that the person referred to by his use of 'I' is somehow ontologically distinct from the person with the leaking bag of sugar.[91] Similarly, it could be argued that we could know all the natural facts (in the *true-sentence* sense of *fact*) without knowing all the normative facts, even if normative properties were natural properties, if natural concepts are distinct from normative concepts. At least, I do not see what reason (beyond the fact that it follows from certain, controversial, semantic theories)[92] that Bedke gives us to accept that this could not be the case.

So, if we understand fact-talk the first way then to claim that knowledge of all the natural facts does not suffice for knowledge of all the normative facts begs the question. While, if we understand fact-talk the second way then we have no reason to expect that knowledge of all the natural facts should suffice for knowledge of all the normative facts.

It might be argued that indexical concepts are a special case and that we cannot infer anything about the relation between natural and normative concepts from them.[93] We could then point to other potential examples but these will all be a bit more controversial. Many philosophers are of the opinion that chemical and biological objects are nothing (ontologically speaking) over-and-above their microphysical constituents but that there are nevertheless no 'bridge laws' between

---

[90] It could possibly be argued that the man in the supermarket could in fact figure out that it was he who spilled the sugar from all the non-indexical facts. I do not think that the same could be argued in relation Perry's other example of Rudolf Lingens in the library (Perry (1977), cf. Lewis (1979a)) however; an example which illustrates the same point.

[91] James Lenman suggested to me in conversation that a philosopher like Geoffrey Medell might argue that reflection on these kinds of cases supports a kind of substance dualism. The idea would be that 'I' refers to a non-physical mind while non-indexical terms like 'the shopper' refer to a physical body. This would undermine the analogy I am trying to make since it would show that indexical and non-indexical terms are not co-referring. I cannot see how this could be made into a plausible position, however. 'I' in claims like, 'I have two legs', could not refer to a non-physical mind, and a non-indexical like 'the shopper' could not refer to a physical body that is distinct from a mind in claims like 'the shopper believes that 2+2=4'.

[92] Neil Mehta (2013) offers a semantic theory on which there might be no semantic links between some coreferring concepts. He argues, convincingly to my mind, that his theory is more parsimonious and explanatorily powerful than the one that Bedke relies on.

[93] Chalmers (2004) presents some arguments against reductionism in a review of Perry (2001) that are in the same ballpark as Bedka's but for which analogies to indexical concepts arguably do not help.

micro-physical concepts and chemical and biological concepts such that one could get all the biological facts from the micro-physical facts.[94] Since these examples are controversial I do not think that they can be used to decisively settle the matter. However, as long as it remains unsettled whether we could always infer all the facts about some domain of properties from all the facts about the domain of properties where the first domain reduces to the second domain, the anti-naturalist cannot be said to have yet refuted naturalism.

§6 – Conclusion

The three main kinds of responses that naturalists have made to the open question argument seem to me to hold up. If we suppose that there are analyses of philosophically interesting terms that could not be doubted then we have no reason to think that 'good' and other normative terms are an exception. If analysis is not always obvious then there is no reason to expect that, 'Is N good?', must be a closed question. Moreover, even if good cannot be analysed, that does not mean that *goodness* could not be a natural property. I now move on to look at some ways that people have reinterpreted Moore's arguments so that they do not rely on any controversial assumptions about meaning or content.

---

[94] Block and Stalnaker (1999) make this point too.

Chapter 2: The Open Question Argument Part 2

In the last chapter, I argued that G.E. Moore's open question argument does not refute naturalism. Moore's argument, at least as I interpreted it, relies on some dubious claims about meaning which the naturalist can reject. Andrew Altman and Peter A. Sutton have both argued, though in different ways, that we can interpret Moore in such a way that his argument does not rest on any claims about meaning.[95] They argue that, so understood, the open question argument is not weak to the responses that have been standardly be made to it. I argue that Altman and Sutton's versions of the open question argument are in fact weak to the standard responses (or slightly modified versions of them) and that their arguments also fail for other reasons. I focus mostly on Altman's argument because I think that there is more to say about it and because I think that what there is to say about Altman's arguments largely apply, *mutatis mutandis*, to Sutton's argument.

§1

According to what Altman calls, "The Standard Interpretation" of the open question argument, Moore argues that *goodness* cannot be a natural property because 'good' has a different cognitive significance to all natural terms. This is the just the interpretation I looked at in the last chapter. Altman lays out the standard objections to Moore's argument on the Standard Interpretation, which I also covered in the last chapter. By way of a reminder, here they are again:

> *Question-begging* - The open question argument is said to be question-begging in the sense that it assumes that no natural term 'N' has the same cognitive significance as 'good' to support the contention that every possible question of the form, 'Is N good?' will be *open*. This assumption is just what is in contention between (analytic) naturalists and their

---

[95] Altman (2004), Sutton (2014).

opponents (assuming that Moore's other premises are correct), so the argument begs the question.

Hasty generalisation - Moore's argument generalises from our intuitions about a very limited numbers of naturalist theories to every possible naturalist theory. Moore only considers the views that *goodness* is identical to *pleasurable*, *desired*, *approved of* and *that which we desire to desire*, but there are many more views available. Moore is not justified in generalising about all cases from the limited number of cases that he considers.

The Paradox of Analysis - Moore's argument seems to require that there are no non-obvious analyses. This is unfortunate since there seem to be many putative examples of non-obvious analyses. We should not infer from the fact that everyone would accept, 'pleasure is pleasure', but not, 'pleasure is good', that 'good' cannot be analysed as 'pleasure'.

Synthetic Naturalism[96]- Moore's argument seems to assume that if two expressions have different cognitive significance then they must refer to distinct entities but this is not necessarily the case. It is often claimed that the predicates "Is water" and "Is composed of H_2O molecules" both refer to one and the same property, *being water*, even though these predicates do not have the same meaning; people knew perfectly well the meaning of the predicate "is water" prior to the discovery in chemistry that water is composed of H_2O molecules. The same might be said of 'good' and natural terms that refer to the same property.

Altman is happy to accept that the standard objections refute Moore's argument on the Standard Interpretation. He motivates his own interpretation of Moore's argument largely based on the assurance that it is not subject to these objections. Because of that and because I already

---

[96] This is something of a misnomer since some analytic naturalists, such as Frank Jackson, can use a version of this objection.

discussed these objections at length in the last chapter, I will, in what follows, assume that if these objections apply to Altman's argument then that argument is unsound.

§2 – Altman's Interpretation of Moore

According to the Standard Interpretation, Moore was concerned to show that *goodness* could not be identical to any natural property because for any natural property, we can always doubt whether objects with that natural property are also good, whereas we cannot doubt that objects with that natural property have that natural property. This shows, Moore argues, that *goodness* cannot be identical to any natural property. Altman finds a different reading.

The passage from *Principia Ethica* that Altman most focuses on appears is §13. In this passage, Moore is trying to refute two alternatives to his view that *goodness* is a simple property. He considers whether it might be held that 'good' is meaningless and whether 'good' might denote a complex (analysable) natural property, in the same way that 'horse' might denote the complex property *is a hoofed quadruped of the genus Equus*.[97] Moore's reply to the idea that 'good' denotes a complex property begins as follows:

> "The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition may be offered, *it may always, be asked, with significance, of the complex so defined, whether it is itself good.*"[98]

Altman calls this the "key sentence" for his interpretation.[99] At first blush, Moore seems not to be talking about whether it makes sense to doubt whether some *object* with a particular natural property must also have the property *goodness*, rather, he seems to be talking about whether it makes sense to say of a *natural property* that it has the property *goodness*. On Altman's interpretation, the crucial point about the property *goodness* is that it cannot be have itself as one

---

[97] Moore (1903, p.60)
[98] ibid. p.67. My italics
[99] Altman (2004, p.397)

of its properties. Following Altman, I will say that when one property has another property, the first property *exemplifies* the second property or that the second property is *predicated* of or *instantiated by* the first. Altman thinks that *goodness* cannot exemplify itself (or, in other words, it cannot be predicated of itself) but that natural properties can exemplify *goodness*. This shows that *goodness* is not identical to any natural property, argues Altman.

We might understand the question, 'Is good good?', to be asking if *goodness* is self-identical, but we might also understand it as asking whether the property *goodness* exemplifies itself. Understood in the second way the question, 'Is good good?', is, according to Altman, "literally meaningless".[100] Asking 'Is good good?', is like asking, 'Is largeness large?'. The properties *goodness* and *largeness* just are not the kind of things that can themselves be good or large, respectively. Altman thinks that there are very few properties which can sensibly be said to exemplify themselves; properties like *being a property* or *not being a number* are some of the rare cases. Altman writes, "there is, as a rule, something philosophical awry with self-predication".[101]

Meanwhile, it seems to makes perfect sense to ask of any natural property whether it is good (here understood as asking whether it has the property *goodness*, not whether that property is identical to *goodness*). 'Is pleasure good?' is a perfectly intelligible question, for example. Here the naturalist has a problem, for if *goodness* can be intelligibly predicated of any natural property, but *goodness* cannot be intelligibly predicate of itself, then *goodness* cannot be identical to a natural property. This is a case of Leibniz's law, since natural properties have the property of *intelligibly having goodness predicated of them*, whereas *goodness* does not have the property of *intelligibly having goodness predicated of it*, *goodness* cannot identical to any natural property.

Altman argues that both the Standard Interpretation of Moore and his interpretation are supported by the text of *Principia Ethica*. However, Altman also thinks that on his interpretation

---

[100] ibid. p.404
[101] ibid., p.405

Moore's argument is much more philosophically defensible. On his interpretation, Moore's argument is not vulnerable to the standard responses, so it is a more charitable interpretation of Moore than the Standard Interpretation. I think that on examination we can see that Altman's interpretation of Moore is not really supported by the text, but since my main concern is whether naturalism can be refuted, I leave these exegetical issues until the appendix. In what follows, I argue that Altman's interpretation of Moore's argument is not a problem for naturalists and that it is vulnerable to the standard criticisms of the Open Question Argument.

§2.1

On Altman's interpretation of Moore's Argument the key point is that it is an open question whether a given natural property exemplifies *goodness*, but not an open question whether *goodness* exemplifies *goodness*. As such, he thinks that on his interpretation Moore would not vulnerable to *The Paradox of Analysis*, since he is simply not concerned with whether normative predicates have the same cognitive significance as any natural terms. Altman's concern is *metaphysical* (or so he argues). He is interested in the metaphysical question of which properties can exemplify *goodness* and not semantic questions concerning the meaning of 'good'. Because of this, Altman does not, unlike Moore on the Standard Interpretation, require any dubious claims about meaning to be true for the argument to work.

Altman also argues that Moore would not be vulnerable to *Synthetic Naturalism* on his interpretation. He does need the assumption that if a question of the form, 'If X X?', does not make sense, but, 'Is X Y?', does, then 'X' and 'Y' cannot be coreferential. But this assumption does not rule out that terms like 'Water' and 'Being composed of H_2O molecules' may be coreferential yet distinct in meaning.

Finally, Altman does not think that *Hasty Generalisation* or *Question-Begging* is a problem on his view either. He says that *Hasty Generalisation* is a problem for Moore because he only considered a

few of the popular naturalist theories of his day and not the more sophisticated theories that have since arisen. He then goes on to consider Richard Boyd's view that good is a homeostatic property cluster involving the satisfaction of various human wants and needs.[102] He says that since it clearly makes sense to ask, 'Is Boyd's homeostatic property cluster good?', his view rules out Boyd's view. The same would apply, he assures us, to other naturalistic theories too. So, his argument does not require a hasty generalisation; the generalisation is well founded, since it applies to the more sophisticated naturalist theories. Altman's argument is also not guilty of begging the question since his claim that 'Is N good?' always makes sense is well grounded, or so he thinks.

I will argue that Altman's argument is weak to versions of all the standard responses and that it has other problems as well. To make it easier to refer to individual premises in what follows, here, almost verbatim, is how Altman sets out his argument:

> (1) Good does not admit of intelligible self-predication.
>
> (2) Good can be intelligibly predicated of [any natural property] P.
>
> (3) Properties cannot be identical with one another if one admits of intelligible self-predication and the other does not.
>
> (4) Therefore, goodness is not identical to any natural property.[103]

I will postpone discussion of premise (1) until §5. In the next section I give reasons to think that premise (2) is false and then in §4 I give a reasons to think that (3) is false.

§3 – Objections to Premise (2)

§3.1 – Is Goodness attributable to properties?

---

[102] Boyd (1988)
[103] Altman (2004, p.400)

Altman's evidence for the second premise of his argument is just the intuition that questions of the form, 'Is pleasure good?', seem to make perfect sense to us. But we should not assume that questions like, 'Is pleasure good?', are necessarily asking of a property whether it is good, as opposed to asking, say, whether a set of objects that have a property like *pleasurable* also have the property *goodness*. Philosophers have different views about which kinds of entities are potential bearers of value and it is by no means uncontroversial that properties are among those entities. If *goodness* cannot be a property of natural properties then there is no asymmetry to explain between *goodness* and natural properties; neither can exemplify goodness.

The example Altman discusses is the question, "Is being yellow good?" He says that the answer to that question would be something like, "In general, yellow is not good or bad, but in a particular painting or color scheme for a house, it may well be good or bad".[104] It is not clear to me whether, 'Is being yellow good?', is really asking of the property *being yellow*, whether it has the property *goodness*, as opposed to asking of yellow objects whether they also have the property *goodness*. It seems just as reasonable to say that it is yellow *objects*, or a *state-of-affairs* in which something is yellow, or the *fact* that something is yellow, which are really the potential bearers of *goodness*. Yellow objects might be good or bad in virtue of being yellow, but it would not follow that their colour properties are, *strictly speaking*, themselves good or bad. If X is F entails that X is G, it does not mean that X's property *being F*, itself has the property *being G*. I think that we can go further and suggest that the property *yellowness* is a bad candidate for being a bearer of value. If it was *yellowness* itself which was the potential bearer of *goodness* then it would be puzzling why it has the property *goodness* in some cases and not others; after all, the property *yellowness* is the same thing everywhere it is instantiated.

What about a question like, 'Is being nationalistic good?' The philosopher who thinks that, *strictly speaking*, properties are not bearers of value, is going to be able to give plausible enough

---

[104] ibid., p.398

interpretations of this question. On one very natural reading, the question is shorthand for asking whether actions which demonstrate an individual's being nationalistic are good actions. For example, if John advocates for the elimination of foreign aid on the grounds that the government should only look after its own citizens I might well ask, 'Is being nationalistic good?', to question whether John's advocacy is a good thing. On another suggestion, the question is asking whether a state of affairs in which people have the property *being nationalistic* will also be a good state of affairs. Suppose I think that nationalism is good. I need not then think that, *strictly speaking*, the natural property *being nationalistic* has the property *goodness*. I might just think that all those *states-of-affairs* which consist in a person's being nationalistic also have the property *goodness*. On a third account, the question asks whether the *obtaining* of the state of affairs that consists in people's *being nationalistic* is good. In other words, it would be asking whether the *fact* that people are nationalistic is good. The linguistic evidence does not seem to offer a compelling reason to accept that properties themselves can, *strictly speaking*, themselves be good.

It might be that objects are the only bearers of goodness, or actions, or states-of-affairs, or facts, or propositions, or worlds, etc. If (natural) properties are not possible bearers of goodness then the question of whether a natural property has the property *goodness* makes no more sense than asking whether *goodness* instantiates itself. Whether or not properties are bearers of goodness is too complex a topic to cover in detail here but we can sketch some reasons to resist the conclusion that properties do instantiate goodness. As N. M. Lemos points out, in the case of other normative properties it seems implausible that they can be exemplified by properties.[105] *Rightness*, for instance, can plausibly be attributed to actions but it seems strange to say that it is a property of the properties of actions. Suppose that giving to charity is right because it promotes happiness and alleviates suffering. It sounds odd to say that the property *promotes happiness and*

---

[105] Lemos (2005, p.183). c.f. Lemos (1991).

alleviates suffering is the bearer of *rightness* in this case, and not the act of giving to charity (or perhaps the state-of-affairs that consists in my giving to charity, or the fact that that state of affairs obtains). Plausibly, things are only right if they are voluntarily done, but it is actions, and not properties, which are voluntarily done by agents. Since some normative properties like 'right' do not seem to be exemplified by other properties, we need some reason to think that things work differently in the case of *goodness*.[106]

Since this worry involves controversial issues in the metaphysics of normativity, I will put it to one side and assume in what follows that properties can be *good*. I will move on to see if Altman's argument stands up granted that assumption.

§3.2 – Hasty Generalisation and Begging the Question

I think that *Hasty Generalisation* and *Begging the Question* can be applied to Altman's argument. The objection originally went like this: if we grant Moore that in cases of successful analyses it will always be impossible to doubt that the analysans and the analysandum denote the same thing, we need some reason to think that this will not be the case for every possible analysis containing normative and natural terms. Moore seemed to assume that this would be the case based on his examination of very few candidate naturalistic analyses of normative terms, so Moore is guilty of making a hasty generalisation. For all Moore has told us, there could be some analyses N of, 'good', on which we could not sensibly doubt that, 'N is good', is true. Since what is at stake between the naturalist and Moore (if Moore's other premises are correct) is whether or not, 'N is good', is not obviously true for every value of N, Moore is also guilty of begging the question against his opponent.

The same objections can be made, *mutatis mutandis*, to Altman's argument. Altman seems to assume, from his consideration of a few questions of the form, 'Is N good?', that all such

---

[106] Moreover, if Altman's argument only works against a Naturalist account of *goodness* and not a Naturalist account of *rightness* then its scope and power would be diminished.

questions will make sense. The naturalist could argue that on the correct analysis, N, of 'good', it really would not make sense to ask, 'Is N good?' In those conditions once we had arrived at the correct analysis of 'good' it would be clear to us that, 'Is N good?', was asking of N/good whether it instantiates itself. Supposing that Altman is right that it is meaningless to ask of a property whether it instantiates itself, it then really would be meaningless to ask, 'Is N good?', for the correct analysis 'N' of 'good'. Altman's claim that this would not the case for any value of N seems just as vulnerable to criticism as Moore's assumption that the answers to all claims of the form, 'N is good', would be open to doubt. Altman considers whether it makes sense to ask if *pleasurable* has the property *goodness*, but he does not consider many other natural properties. Therefore, if *Hasty Generalisation* is a good objection against Moore then it seems to be a good objection against Altman too.

Suppose that 'brother' can be analysed as 'male sibling'. If that were the case then, plausibly, *being a brother* and *being a male sibling* would be the same property.[107] In that case, the question, 'Does John's property *being a brother* have the property *being a male sibling*?', would be understood by competent language users as asking whether a property exemplified itself. Altman would have to say that that question does not make any sense since he thinks that asking if a property exemplifies itself never makes sense. The naturalist could ask for a reason to suppose that something similar is not true in the normative case. If there is an analysis, N, of 'good' on which it is clear to competent language users that 'N' and 'good' refer to the same property then according to Altman the question, 'Is N good?', should not make sense either. Altman has not given us a reason to doubt that there is an analysis of 'good', so he has not given us reason to think that, 'Is N good?', will always make sense.

---

[107] I am ignoring many of the issues I raised regarding analysis in the last chapter because they are more relevant to the *Unobvious Analysis* objection that I apply to Altman's account later.

Following on, Altman's argument could also be said to beg the question against the naturalist since (if we accept Altman's other assumptions) exactly what is at stake between the naturalist and the anti-naturalist will be whether it makes sense to ask of every single natural property whether it is good. Since, Altman assumes that it will always make sense to ask of any natural property whether it is good, Altman begs the question as much as Moore did. Granted, Altman does not beg the question against the naturalist in exactly the same way that Moore was originally accused of doing, but a charge of begging the question is nevertheless just as applicable.

Altman recognises the possibility of this response but he writes, "It is insufficient to charge Moore with hasty generalization unless one specifies the leading candidates that Moore overlooked."[108] Altman thinks that his argument is safe from *Hasty Generalisation* because it applies to (what he considers to be) the leading naturalist theories when he was writing, such as Richard Boyd's account.

However, this does not seem to be what the naturalist is getting at when they make objections along the lines of *Hasty Generalisation*. I think that they are really just challenging the non-naturalist to give us a principled reason to suppose that no possible naturalistic analysis could make the relevant questions closed. Moore's argument says that since all prior attempts to analyse good have failed, all future attempts will; a fairly brash inference. I do not see why Altman thinks that it would be insufficient for the naturalist to charge Moore with making a hasty generalization without also providing a counter-example proof analysis of their own. Since it is the case at most only one analysis could be the correct one, and since analyses can often extremely complicated, it does seem reasonable to question how much we can infer from the fact that a handful of attempts so far have failed. And, it seems that Altman's argument is just as weak to this objection.

---

[108] Altman (2004, pp.402-3)

With all that said, I think that there are better objections than *Hasty Generalisation* and *Question-Begging* that one can make to Altman's argument so I move on to those.

§4 – Objections to Premise (3)

§4.1 – Relational Views of 'Good'

According to Moore's view, all uses of 'good' refer to some monadic property. In *Good and Evil*, Peter Geach offered a powerful argument against this view.[109] According to Moore to say that, 'X is a good K', is to say that X is a K and that X is good. On Moore's account 'good' is a *predicative* adjective. An adjective, A, is predicative if we can infer from the fact that X is A and the fact that X is identical to Y, that Y is A for any X and Y.[110] 'Red' is a clear example of a predicative adjective. If a given car is red then it must also be a red vehicle, a red mode of transport, a red artefact, etc. An adjective is *attributive* if it is not predicative. 'Big' is an attributive adjective; we cannot infer from the fact that Jerry is a big mouse and from the fact that Jerry is an animal that Jerry is a big animal. Geach argued that, 'good', in, 'good K', works more like 'big' than it does like 'red'; if John is a good lawyer and John is a dancer it does not follow than John is a good dancer.

These considerations suggest that, 'good', in phrases like, 'good K', does not denote some monadic property. An alternative suggestion is that 'good' refers to a relational property.[111] According to some semantic theories, words like 'tall', 'fast', 'old', etc. have an argument place for a particular *standard*. When I say 'Michael Jordan is tall' I might mean that he is tall *by the standards of adult humans* or something along those lines. The predicate 'is tall' in that sentence refers to the relation of meeting the contextually specified standards for being a tall adult human.[112] According to relational views of 'good' works the same way. To say of a toaster that it

---

[109] Geach (1956).
[110] This definition is taken from Thomson (2008). It is in the same spirit as Geach's account but it is a bit clearer.
[111] Ziff (1960), Katz (1966), Phillips (1998), Thomson (2001), Copp (2001), Shanklin (2011), Finlay (2014).
[112] Semantic minimalists challenge this view of predicates like 'is tall' but theirs is very much a minority view.

is good would say that the toaster meets certain standards, e.g., in an ordinary context, standards requiring that the object toast bread efficiently, safely, evenly etc.

We could sensibly ask of a property whether it was good on this picture, because having a particular quality or property can be recommended or demanded by standards. Suppose that John harbours a desire to become a jockey and suppose I point out that he is quite short. You, unfamiliar with horseracing, ask, 'Is being short good?'. By default in this context, your question will be understood as asking whether being short for an adult male is the kind of quality that is advantageous for winning horse races, or to put it another way, whether being short for an adult male is among the standards by which we judge potential male jockeys. Take another example. Suppose I ask, 'Is being frugal good?' I think that absent any other context this would be understood as asking whether being frugal is recommended (or perhaps demanded) by some particular practical or moral standards.

Whatever problems it may have, this kind of account is well positioned to handle Altman's argument. It will be no surprise on this account if the question, 'Is good good?', outside of a particular context, does not seem to make any sense. The question will not make sense because it will be unclear what relational property the two uses of 'good' are supposed to be denoting. This means that it will be semantically indeterminate which proposition the question is concerning. Since the predicate 'is good' can denote many different relational properties given different contexts and the different objects it can be applied to, there will be no standard 'default reading' of the question, 'Is good good?'. Judith Jarvis Thomson, discussing something like this point, gives the example: "Was St. Francis better than chocolate?"[113] Because it is unclear what is meant by 'better than', (it is unclear in just what respect St Francis is to be compared to chocolate), the

---

[113] Thomson (2008, p.59)

question will not seem meaningful to us. So, if, 'Is good good?', does not make sense, the naturalist has available to them a good explanation of this.

At the same time, I think that we can come up with some examples where 'Is good good?' or an equivalent question could make sense. These will be cases where it is intelligible that the relational property picked out by the first use of 'good' has the relational property picked out by the second use of 'good'. For example, consider this exchange:

> A: "Mary is a really good person. She gives 50% of her income to charity."
>
> B: "Yes, but is being good good? After all, she can't afford a car or a mortgage"

In this case, "Is being good good?", is understood as asking whether meeting certain moral standards meets certain other prudential standards.

Or, consider these examples:

> A: "Mary is really good. She kills her targets without hesitation and her crimes are never traced back to her."
>
> B: "Is her being good good though? Her actions cause so much heartbreak to the families of her victims."

> A: "The guerrillas' plan was very good. By targeting the capital on a Saturday they ensured that they caused maximum panic.
>
> B: "But is it good that their plan was good? Many civilian lives were lost."

In these cases, it is asked whether meeting some standard associated with effectively killing or with causing panic meets a moral standard. I admit that these examples do not sound perfectly natural (we would not usually relativise two instances of the same incomplete predicate to different standards in the same sentence because it invites confusion.) Nevertheless, the question uttered by B in these examples certainly do not seem meaningless.

Now Altman might object that this suggestion misses his point. When he says that "Is good good?" does not make sense, he is understanding "good" to refer to some simple monadic property *good simpliciter* or *plain good*. Altman could argue that it makes no sense to say that goodness *simpliciter* could itself be good *simpliciter*, but it does make sense to say that other properties are good *simpliciter*, showing that goodness *simpliciter* is not identical to any natural property. However, on these naturalistic views there is no such property as *goodness simpliciter* just as there is no such thing as *being tall simpliciter*. Some people like Peter Geach argue that talk of "good simpliciter" is just based on a confusion.[114] Others like Stephen Finlay and Robert Shanklin offers vindicatory accounts of the meaning of the expression 'good simpliciter'[115] on which the expression is legitimate but on which it does not refer to some monadic property. In any case, Altman would need an independent argument to show that questions like, 'Is pleasure good?', are asking whether pleasure has a monadic property and not a relational property, if he was to put pressure on these relational kinds of naturalism.

§4.2 – The Paradox of Analysis and Synthetic Naturalism

Many naturalists think that uses of 'good', or at least certain uses of 'good', denote a particular monadic natural property. Even if Altman's view is not effective against relational naturalistic views, if it were only effective against these non-relational views then that would be an impressive result. I now want to argue that it is not effective against these views either.

Altman, responding to the potential objection that *goodness* does have itself as a property says, "there would be a strong presumption against the premise in light of the fact that the general rule seems to be that properties do not have themselves as properties."[116] Altman thinks that this

---

[114] Geach (ibid.) c.f. Aristotle (1096a 24-29)
[115] See Finlay (2014) and Shanklin (2011)
[116] Altman (2004, p.406)

'general rule' explains why the question, 'Is good good?', does not make any sense; properties do not exemplify themselves so it cannot make sense to ask of *goodness* if it has the property *goodness*.

Suppose that that is all correct. In that case, the naturalist could simply say that whatever natural property is identical to *goodness* could not have *goodness* as one of its properties since that would contravene Altman's non-self-exemplification rule. They can say there must be at least one natural property which necessarily does not have *goodness* as one of its properties. That is not quite enough to show that Altman's argument is unsound. We would need an explanation of why the question 'Is N good?', seems to be intelligible while, 'Is good good?', does not. The fact that this question is intelligible, but 'Is good good?' is not, is Altman's basis for saying that 'good' and 'N' cannot refer to the same property.

Now here is the important point: the naturalist can just say that the reason that that question, 'Is N good?', is intelligible is that it that it is not obvious that, 'N', and, 'good', denote the same property. Because it is not obvious that, 'N', and, 'good', denote the same property, it is not obvious that, 'Is N good?', is asking of a property whether it exemplifies itself, or a different property. Because of that, the question, 'Is N good?', appears to be intelligible. Since it is an open question whether 'N' and 'good' are co-referential, it is an open question whether or not, 'Is N good?', is asking whether a particular property exemplifies itself, or whether it asking if one property exemplifies some other distinct property. The question, 'Is good good?', by contrast, obviously asks of a particular property whether it has itself as a property since it features two uses of the same word.[117]

This strategy then amounts to denying premise (3) of Altman's argument. If, 'Is good good?', is not intelligible but, 'Is N good?', is, that need not be because 'good' and 'N' do not refer to the same property. It might just be because 'Is good good' is obviously asking if a property

---

[117] At least on the assumption that 'good' is not context-dependent in the way assumed by the relational views looked at in the last section

exemplifies itself, while it is unobvious that 'Is N good?' is asking of a property that it exemplifies itself.

This strategy works for both analytic naturalism and synthetic naturalism. Analytic naturalists will say that 'N' and 'good' are synonyms but not obviously so, so it is not obvious that question, 'Is N good?', is asking of a property whether it has itself as a property. Synthetic naturalists can say that 'N' and 'good' are not synonyms but that they nevertheless refer to the same property. Because they are not synonymous the question, 'Is N good?' is intelligible to us as we might think that it is asking of one property whether it has a distinct property *goodness*. Therefore, it turns out that the naturalist can offer very similar responses to Altman's version of Moore's argument as they can to the standard version of Moore's argument.

The suggestion is that, 'Is good good?', and, 'Is N good?', are in a sense asking the same thing; they both ask if good has itself as a property (though they ask it under different guises). Altman might try to resist this on the grounds that, 'Is good good?', cannot be accepted whereas, 'Is N good?', might well be. However, this would just be to return to the standard interpretation of Moore's argument. It would be to argue that 'good' and 'N' cannot refer to the same property because they have a different *cognitive significance*. As such, it would be guilty of all the faults of Moore's argument on the standard interpretation; faults which Altman believed meant the argument was 'pronounced dead'.

At this point Altman might backtrack and claim that the metaphysical obstacle to *goodness* exemplifying itself is not a problem with self-predication but with regards to some other issue. Perhaps he might say that *sui generis* non-natural properties are not in the kind of metaphysical category to exemplify *goodness* whereas natural properties are. But it is hard to see why the naturalist should be persuaded by this. At this point the objection just looks like a variant of the *Just Too Different Intuition* that philosophers like David Enoch and Michael Huemer push; that

intuitively natural properties are *just too different* to normative properties to be identical to them. I cover this kind of objection in the next chapter.

§5 - Is it Meaningless to Ask if *Goodness* Instantiates Itself?

So far, I have been taking it for granted that Altman was right to say that it was meaningless to suggest that good instantiates itself, but it is far from obvious that he is correct. I gave a few examples in §4.1 of contexts in which, 'Is good good?', seems to make sense but in those cases both uses of 'good' were not denoting the property *goodness simpliciter* or *plain goodness*.[118] However, even if we restrict our discussion to talk of *goodness simpliciter*, it seems intuitive, to me at least, that, 'Is good good?', is not meaningless. In fact, I think it quite clearly has a meaning. 'Is good good?' means roughly 'Does the property *goodness* have the property *goodness?*'.

According to perhaps the default semantic account of interrogative sentences, the meaning of a question is a function from a world to the set of propositions which are possible answers to the question in that world.[119] Understood this way 'Is good good?' can be assigned a meaning; it is a function from a world to the propositions *Goodness is a property of goodness* and *It is not the case that goodness is a property of goodness*. Perhaps that account of interrogatives sentences is not correct and on the correct account 'Is good good?' cannot be assigned a meaning, but Altman does not engage with this issue, nor does he at any point suggest what criterion of meaningfulness he is using, so we do not have to grant him premise (1). Meaningless questions do not usually have sensible answers but Altman seems to think that the question can be sensibly answered; he thinks that the answer is 'No, since goodness cannot exemplify itself'. Altman does not really say anything to defend premise (1).

Instead, he hedges his bets, as it were. He argues that even if (1) is false, naturalism can still be refuted. Since Altman does not defend (1) and since there are some reasons to think that it is

---

[118] For the time being I ignore the potential differences between these properties.
[119] Hamblin (1973)

false, the naturalist could justifiably reject this premise, rendering Altman's original argument unsound. I will now look at Altman's Reformulated Argument, which does not rely on (1) being true.

§5.1 - Altman's Reformulated Argument

If it were the case that, 'Is good good?', is in fact meaningful this would not help the naturalist, argues Altman, because he could just at that point reformulate his argument. If it does make sense to ask of *goodness* whether it is good then, by metaphysical necessity, the answer will always be negative, whereas for any given natural property, *qua* natural property, there is nothing metaphysically impossible about its being good. That being so, Altman argues, he could just put forward a variant of his argument:

> (5) *Goodness* is not identical to any natural property in virtue of the fact that *goodness* cannot, as a matter of metaphysical necessity, have itself as a property.
>
> (6) There is no metaphysical restriction against a natural property, *qua* natural property, having the property *goodness*.
>
> (7) Therefore, by Leibniz's Law, *goodness* cannot be the same as any natural property since *goodness* has the property *cannot by metaphysical necessity be good*, whereas natural properties do not have that property.

This is Altman's Reformulated Argument.

§5.2

There are two ways to tackle this argument; we could deny that, 'Is good good?', necessarily has a negative answer or we could argue that, 'Is N good?', necessarily does have a negative answer for some value of N. I will look at these both responses in order.

If we accept that, 'Is good good?', is meaningful but deny that the answer to it is necessarily negative then our others options are to say that the answer is necessarily positive, or that it is sometimes positive. It might seem surprising that the view that good does not admit of self-predication would be attributed to Moore. Many people see Moore's view as a kind of Platonism about good.[120] The idea of *goodness* as a kind of simple, *sui generis*, non-natural property strikes many people as being strongly reminiscent of Plato's idea of the Form of the Good. And Plato seemed to think that The Form of the Good was itself good, indeed it was good to the highest degree. Why should we suppose that Moore was not inclined to say the same thing about *goodness*?[121]

Altman admits, "Moore provides no illumination as to why he might accept the premise [that good cannot be good]." But he goes on to speculate, "It is entirely possible that he is concerned to halt Plato's famous third-man argument. Moore might have thought that a higher-order good would be required to make sense of the question whether good is good. And a still higher-order good would be needed to make sense of the question whether second-order good is good." This is one way to motivate the suggestion that good does not admit of self-predication but it does not seem conclusive. There are solutions to the third-man argument which do not involve dropping the premise that the Form of F is F, so avoiding the conclusion of the Third Man argument does not require that we agree with Altman.[122]

Altman also speculates that, "[Moore] may simply have thought that the self-predication premise was obviously and fundamentally true, i.e., that it was an "intuition" in his sense of the term." Intuitively, properties cannot exemplify themselves, for example, it sounds very odd to say that the property of being intelligent is itself intelligent; properties do not seem to be the kind of

---

[120] e.g. Baldwin (1993)
[121] Sutton (2014) argues that Moore think that it is tautologous that the fact that X is good is good.
[122] See Rickless (2015)

things that can be intelligent. Altman says that given how intuitively unappealing the idea that goodness exemplifies itself is, the burden of proof is on the proponent of this claim.

Altman admits that there are exceptions to the rule that properties do not exemplify themselves, e.g. *being a property*, *being self-identical* etc. That being so, the intuition that properties cannot, by metaphysical necessity, exemplify themselves, has no probative value, since Altman admits that it is false (if we even have that intuition in the first place). Altman probably intends for the relevant intuition to be weaker. He probably means for the intuition to be that properties cannot, by metaphysical necessity, *usually* exemplify themselves or something like that. However, this weaker intuition is compatible with the claim that good is good, since good might be one of the unusual cases of a self-exemplifying property.

Altman might instead think it is just intuitive that *goodness* in particular cannot exemplify itself. However, it is not clear how strong this intuition is. For example, it does not seem unintuitive to say that Martin Luther King Jr.'s *goodness* was itself a good thing. It reasonable to say that it was *good* (*simpliciter*) that MLK was good (*simpliciter*), much better (*simpliciter)* than if he were bad (*simpliciter)*. Examining this issue further would probably require getting more bogged down in a lot of issues in normative metaphysics so I move on to what I think is a better response.

§5.3

The preceding considerations tell against Altman's 'reformulated' argument. They all suggest reasons to suppose that *goodness* might be able to exemplify itself, undermining the distinction that Altman wants to draw between *goodness* and a natural property. There is, however, a more damning objection we could make. We could undermine the distinction that Altman wants to draw in a different way, by arguing that for some natural property N, the answer to the question, 'Is N good?', is necessarily negative if Altman's other assumptions are correct. This is a variation on an objection I considered in the last section. Altman argues that if, 'Is good good?', makes

sense then the answer is necessarily 'No', since *goodness*, by metaphysical necessity cannot have itself as a property. This is a problem for the naturalist, argues Altman, because there is no metaphysical restriction on natural properties, *qua natural properties*, being good. Altman argues that since there is a metaphysical restriction that applies to *goodness* that does not apply to natural properties (*qua* natural properties), *goodness* cannot be a natural property.

However, if, as the naturalist thinks, *goodness* is identical to a natural property N, then the answer to the question 'Is N good?' must also necessarily be 'No'. If N is identical to goodness and if *goodness* cannot be good, then any property which is identical to *goodness* can obviously not be good either. Altman might be right to say that for any natural property there is no metaphysical restriction, *qua natural property*, to its being good, but that is neither here nor there since that claim is compatible with there being a natural property which, *qua being identical to goodness*, cannot itself be good. To reiterate, if there is no metaphysical restriction, *qua natural property*, to a natural property being good, that does not show that there is no metaphysical restriction, *tout court*, to natural properties being good. It is not N's being a natural property that means that it cannot exemplify goodness, rather, it is N's being identical to goodness which means that it cannot itself exemplify goodness (if we grant Altman that goodness cannot exemplify itself). Of course, Altman does not think that any such natural property exists, but he needs an independent argument for that claim or he just begs the question against the naturalist. So, Altman's reformulated argument is invalid.

Altman might argue that the answer to, 'Is N good?', cannot be necessarily negative for every value of N since we can always doubt what the answer is, but this would obviously just be to return to the Standard Interpretation of Moore's argument which Altman rejects. If goodness cannot self-exemplify because of a *metaphysical* restriction, as Altman suggests, then the fact that people can doubt whether, 'Is N good?', has a negative answer should not matter. If N cannot, by metaphysical necessity, be good (because N is identical to goodness), then the correct answer

to, 'Is N good?', is necessarily 'No', regardless of whether or not that is the answer people would in fact give.

So all in all, there are many possible ways we might put pressure on Altman's reformulated. argument. It is not clear that *goodness* cannot be predicated of itself, nor that it can be predicated of any natural property. This concludes all I want to say about Altman's arguments, I move on now to briefly look at a very similar argument from Peter Sutton.

§6 – Sutton's Argument

Peter Sutton also claims to have found a 'new' open question argument which centres around issues of self-predication, but the details of his argument differs from Altman's argument.[123] Altman could, if what I have said up to now has been right, adopt this argument as a last ditch attempt to refute naturalism.

Whereas Altman would say that a claim like, 'The fact that X is good is good' ,is either nonsensical or necessarily false, Sutton thinks that it is necessarily true.[124] He does not give much support for this claim. He might be persuaded by examples like the Martin Luther King example that I considered in the last section or perhaps he is committed to some kind of Platonism from which he thinks the claim follows. In any case, it does not seem outrageous to think that a thing's *goodness* is itself *good*, at least in some cases. I am happy to grant Sutton's premise for the sake of argument.

The claim, 'The fact that X is good is good', contrasts with a claim like, 'The fact that X causes pleasure causes pleasure', which, if true at all, are only contingently true, argues Sutton. Things that increase net pleasure might not increase net pleasure. If we are prudish, or anti-hedonistic,

---

[123] Sutton (2014).
[124] Sutton considers whether 'The fact that X is good is good' is nonsensical, but he says that it would not matter since he could just revert to using Altman's argument if it were. Since I have shown that Altman's argument does not work I think that it would matter.

or sadistic then perhaps the fact that certain things cause pleasure might cause net displeasure. This goes for any natural term we could substitute for 'causes pleasure'; we can always imagine a world in which the fact that something has a particular natural property is not good. So, goodness cannot be the same as any natural property since, even in non-intensional contexts, we cannot substitute a natural term denoting that property for 'good' *salva veritate*. Naturalists are committed to saying that the truth conditions of a claim like, 'The fact that X is good is good', is the same as a claim like, 'The fact that X causes pleasure causes pleasure', so if the former claim is necessarily true and the latter claim is not, then naturalists have a problem. Like Altman, Sutton argues that his own version of Moore's argument is superior to the traditional ones because it does not rely on any controversial assumptions about meaning or cognitive significance.

I think it can be shown that Sutton's argument does not work. First, I need to show that regardless of which metaethical theory turns out to be true it will be the case that normative properties are necessarily coextensive with some (broadly) natural property. Naturalists and non-naturalists disagree in that non-naturalists think that *goodness* is something ontologically over-and-above natural properties whereas the naturalist does not, but both naturalists and non-naturalists agree that moral goodness is necessarily *coextensive* with some (broadly) natural property.[125] If a certain form of utilitarianism is true, for instance, then 'good' will denote a property which is necessarily coextensive with 'causes pleasure', whether naturalism or non-naturalism are true.

It might be thought that the possibility of particularism shows that *goodness* need not be necessarily coextensive with any natural property. Particularists think that whether or not something is good is dependent on many different features of the particular context that that thing is in. They reject that there are any general normative principles like, 'Good actions are

---

[125] It might be thought that theological voluntarists would disagree since properties like *that which God approves of* are not natural properties but we could understand 'natural' very broadly to encompass supernatural properties. They may sound absurd but it is a common practice in metaethics dating back to Moore.

those that cause pleasure'. The fact that something causes pleasure can be good in some contexts but bad in others (e.g. the pleasure a sadist derives from torture). The same applies, particularists argue, to all normative considerations; sometimes they count in favour of some action and sometimes against, depending on very particular features of the context the action takes place in.

However, particularists accept that there is at least one normative principle.[126] This just follows from the strong supervenience of the normative on the natural; a principle, which enjoys almost universal assent.[127] The strong supervenience of the normative on the natural is a principle which says that two worlds cannot be different in any normative respect without being different in some (broadly) natural respect. This seems quite intuitive. Assuming the assassination of JFK was wrong in this world, it does not seem possible to imagine a world where every single natural fact was exactly the same but assassinating JFK was not wrong.

Now if we take some possible good (*simpliciter*) thing we could give a complete description of the natural features of that thing, including all the natural features of the world in which it takes place, and call that w_1. Because the normative supervenes on the natural, we know that there could not be a world which was identical to the world described by w_1 but in which the relevant thing was not good. Next, we take another *good* thing and describe fully the natural features of its world and call that w_2. We could repeat this process for every possible good thing. We then put all these descriptions into a massive disjunction; e.g. $w_1 \, v \, w_2 \, v \, w_3...v \, w_n$. Call this disjunction 'W'. 'W' does not feature any normative terms but it will be necessarily equivalent to the predicate 'good'. For any good thing it will be trivially true that it satisfies the long disjunction description e.g. $w_1 \, v \, w_2 \, v \, w_3...v \, w_n$ since a complete description of that action will necessarily be one of the disjuncts. Particularists must concede that the normative principle, 'Things which are W are *good*' is true. This is not a problem for particularists since, if W is massively disjunctive, the principle

---

[126] See e.g. Dancy (2004)

[127] Nicholas Sturgeon raises some interesting objections to the supervenience thesis (2003, 2009) but I simply do not have space to answer them here.

could not be comprehended by agents and so would be useless for normative decision-making. The motivation for being particularist would then remain since there would not be general normative principles we could use in making normative decisions. The important point is that whatever view we accept, assuming the strong supervenience of the normative on the (broadly) natural, there will be some non-normative expression which is necessarily coextensive with 'good'.

We could then call the natural *property* that is denoted by 'W', *N*. The normative property *goodness* will be necessarily coextensive with the property N, since any action which is good must also exemplify N.[128] Whichever first-order ethical theory turns out to be true there will be some expression that only uses natural terms which denotes a property which is necessarily coextensive with 'good'. The expression need not be an incredibly long disjunction, however. If a certain form of utilitarianism is true, for instance, then 'good' will denote a property which is necessarily coextensive with *'pleasurable'*. Naturalist utilitarians think that *goodness* just is the property *pleasurable* so they think that *goodness* is, trivially, necessarily coextensive with *causes pleasure*. Non-naturalist utilitarians meanwhile think that *goodness* and *pleasurable* are distinct properties, but nevertheless when something exemplifies one property it must also exemplify the other too. In either case, *goodness* will be necessarily coextensive with *pleasurable*.

Now we have all we need to refute Sutton's argument. Suppose for the moment that utilitarianism is true. Assume for contradiction that, 'The fact that X is good is good', is necessarily true and, 'The fact that X causes pleasure causes pleasure', is contingently true, if true at all. These claims are just the premises of Sutton's argument.

Since we have assumed utilitarianism, if, 'The fact that X is good', is true then, 'The fact that X causes pleasure', must also be true. If 'good' and 'causes pleasure' refer to necessarily coextensive

---

[128] Jackson (1998), Gibbard (2003), Dancy (2004), Ridge and McKeever (2005).

properties then any world in which it is true that X is good is also a world in which it is true that X is pleasurable. Since Sutton thinks that good is predicative the same must also go for higher-order uses of 'good'. So, if it is necessarily true that, 'The fact that X is good is good', then, 'The fact that X is good causes pleasure', must be necessarily true also. If the fact that X is good has the property *goodness* in every possible world, and if *goodness* is necessarily coextensive with *pleasurable*, then the fact that X is good must have the property *pleasurable* in every possible world. If that is the case then 'The fact that X causes pleasure causes pleasure' must be necessarily true also, again because we have assumed that *goodness* and *causes pleasure* are necessarily coextensive. However, we know that Sutton thinks that any claim like 'The fact that X causes pleasure causes pleasure' is, at most, contingently true. Since that claim is, at most, contingently true the claim 'The fact that X is good is good' must be at most contingently true also. However, this contradicts the assumption we started with that 'The fact that X is good is good' is necessarily true. So, whether naturalism or non-naturalism is true, on the assumption that utilitarianism is true, it cannot not the case that 'The fact that X is good is good' is necessarily true and 'The fact that X is good causes pleasure' is, at most, contingently true. We can run the same argument, *mutatis mutandis*, for whichever first-ethical theory turns out to be correct. Therefore, Sutton's argument fails.

Assuming the normative supervenes on the natural then we know that there is some natural term which refers to a property which is necessarily coextensive with the property referred to by 'good'. That being so, it must be the case that for some natural term we could substitute it for 'good' without changing the truth value of any claims in which it appears (if we exclude, for simplicity's sake, intensional contexts). Since every object which is good is also N, any claim of the form, 'X is good', will be true iff, 'X is N'. So, Sutton's argument does not work.

Sutton could perhaps argue against the supervenience of the normative on the natural. However, if he were to do that the naturalist could just adopt a position like Nicholas Sturgeon's non-

reductive naturalism. Since on that view it need not be the case that there is any natural term which picks out the natural property that (according to naturalists) is identical to *goodness*, the fact that no natural term can be substituted for 'good' *salva veritate* would not be a problem; in fact, it would be exactly what we would expect. So, whether or not the principle of supervenience holds, naturalists can answer Sutton's challenge.

§7 - Conclusion

In this chapter, I have shown that three arguments against naturalism fail. Altman's first argument was that since, 'Is good good?', is meaningless and, 'Is N good?', is not it cannot be the case that 'N' and 'good' denote the same property. Altman, I argued, fails to establish either that 'Is good good?' is meaningless, or that 'Is N good?' is not meaningless for all values of 'N', so both premises fail.

Altman's second argument was that if, 'Is good good?', is meaningful then necessarily it has a negative answer, whereas, 'Is N good?', does not necessarily have a negative answer, from which it follows that 'N' and 'good' cannot denote the same property. I, again, challenged both premises and the validity of this argument.

Sutton's argument, meanwhile, was that since, 'Is good good?', necessarily has a positive answer, whereas questions like, 'Is N N?', do not, 'N' and 'good' cannot be co-referring. Both premises of Sutton's argument cannot be at the same time correct. Since *goodness,* on almost everyone's account, is at least necessarily coextensive with some (broadly) natural property, it must be the case that if, 'Is good good?', is necessarily true, then, 'Is N N?', is necessarily true also, for some value of N.

Appendix – Has Altman interpreted Moore correctly?

Altman writes in the introduction of his paper, "my fundamental aim is philosophical rather than exegetical". [129] At the same time, throughout his exposition Altman takes the time to cite textual support for his interpretation. If he were correct then it would undermine my reading of Moore in the last chapter, so the matter is worth looking at briefly, but if the reader is not interested in these exegetical matters then they are invited to skip to the next chapter.

The first thing to acknowledge is that Altman does concede that there are passages in *Principia Ethica* which seem to contradict his interpretation and support the standard interpretation. [130] But, he argues, when Moore first formulates his argument he does so in a way that does not chime well with the standard interpretation. [131] If there are some passages that support his interpretation and others that support the standard interpretation, and if Moore's argument is stronger on Altman's interpretation than it is on the standard interpretation, then considerations of charity will push us to prefer Altman's interpretation. I have already argued that Altman's version of the argument is no threat to naturalism so his interpretation is not any more charitable than the standard one. But we can also see that Moore's text does not support Altman's interpretation either. In fact, the passage that Altman cites specifically to support his interpretation in fact gives very little support to that interpretation.

I mentioned earlier that the passage from *Principia Ethica* that Altman most focuses appears in §13 of the first chapter. Here is the "key sentence" for Altman's interpretation again:

> "The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition may be offered, it may always, be asked, with significance, of the complex so defined, whether it is itself good." [132]

---

[129] Altman (2004, p.396)
[130] e.g. Moore (1992, p.95, p.176)
[131] Altman (2004, p.397) referencing Moore (1992, pp.66-69)
[132] Moore (1992, p.67)

Moore seems to be concerned with whether it makes sense to say of a *natural property* that it has the property *goodness*. Altman argues that this is a problem for the standard interpretation because on that view the question, 'Is pleasure good?', is asking of objects which have a particular natural property (in this case the property *pleasant*) whether they also have the normative property *goodness*. On Altman's interpretation the question, 'Is pleasure good?', asks whether the natural *property* that naturalists identify with *goodness* is itself good. The "key sentence" arguably makes more sense if we interpret Moore as Altman does than if we interpret Moore according to the Standard Interpretation. That is the first piece of textual evidence Altman gives for his view.

Additionally, in this passage Moore says that it can always be asked "with significance" whether pleasure is good. According to the Standard Interpretation "with significance" means something like *cognitive significance*. The thought is that 'good' and any natural term N have different cognitive significance since 'Is N good?' can be doubted while 'Is N N?' cannot be doubted. Altman interprets Moore differently here. By, "with significance", Altman takes Moore to mean, "intelligible", taking his cue from the fact that Moore seems to use the term, "intelligible", in place of, "with significance", later in the same passage.[133]

Altman takes "intelligible" to mean just 'possesses meaning'. He argues that this is a problem for the Standard Interpretation. If "significance" is just intelligibility then the standard interpretation cannot be correct since trivial identity questions like, 'Is pleasure pleasure?', are perfectly intelligible. The identity question 'Is pleasure [identical to] pleasure?' may be trivial but it still has meaning; it is syntactically well-formed, all its constituent words are meaningful, etc. So, Moore cannot be primarily concerned with the fact that identity questions like 'Is pleasure pleasure?' and 'Is pleasure good' have different significance, since "significance" is just meaningfulness, and both these questions are meaningful. Instead, he must be concerned with questions like 'Is pleasure good?' and 'Is good good?' understood not as identity questions but questions

---

[133] Moore (1993, p.67)

concerning whether one property exemplifies another property. Altman thinks that understood this way 'Is pleasure good?' is meaningful and 'Is good good' is not.

To recapitulate then, Altman gives two reasons to think that this passage favours his interpretation over the standard one. First, Moore seems to be talking about whether a natural property is good, and not whether an object with that natural property is good. Second, Moore says that closed questions are unintelligible, but if closed questions were just asking of one thing if it was identical to itself then they would not be unintelligible.

I think it would be puzzling if Moore meant what Altman says he does given what Moore goes on to say in the same passage. Shortly after the "key sentence", Moore writes:

> "It may indeed be true that what we desire to desire is always good; perhaps, even the converse may be true: but it is very doubtful whether this is the case, and the mere fact that we understand very well what is meant by doubting it, shews clearly that we have two different notion before our mind.""[134]

Altman writes, "The open question, in its initial formulation then, is a question about the good of natural properties, and not a question about the good of objects possessing those properties".[135] But, in this sentence at least, it seems very much that Moore is asking whether the object that possesses the natural property *being the thing we desire to desire*, also possesses the property *goodness*. Moore here is saying that because we can doubt that those things which we know to be desired to desired are good it shows that being a thing we desire to desire is not the same as being good. Additionally, Moore here seems more concerned with the *cognitive significance* of believing of some object that it is good, rather than its meaningfulness. This lines up with the Standard Interpretation and conflicts with Altman's.

---

[134] ibid.
[135] Altman (2004, p.397)

Now Moore does write in the same passage:

> "If we carry the investigation further, and ask ourselves "Is it good to desire to desire A?" it is apparent on a little reflection, that this question is itself as intelligible as the original question, "Is A good?" – that **we are, in fact, now asking for exactly the same information about the desire to desire A, for which we formerly asked with regard to A itself**. But it is also apparent that the meaning of this second question cannot be correctly analysed into "Is the desire to desire A one of the things which we desire to desire?": we have not before our minds anything so complicated as the question "Do we desire to desire to desire to desire A?"[136]

This passage might seem at first glance more congenial to Altman's interpretation. Altman takes Moore here to be pointing out that the question, "Is it good to desire to desire A?", is intelligible, because that tells against the view that goodness is identical to being that which we desire to desire. It tells against that view because good cannot predicated of good but can be predicated of being that which we desire to desire (though Moore never says this explicitly).

But notice here that Moore's question is not actually about whether the property of *being that which we desire to desire* is good, it is about whether the desire to desire A is good (see bolded text). In other words, Moore is talking about whether a mental state is good, not whether a property is good. Moore finds it very implausible that to ask whether that mental state is good is equivalent to asking whether that mental state is what we desire to desire since the second question is much "complicated" than the first (and so has different cognitive significance). Again, *pace* Altman's interpretation, Moore seems to be concerned with the question of the good of things (in this case the mental state the desire to desire A) with a particular natural property (in this case *being that which we desire to desire*), and not whether the natural property itself is good. In line with the

---

[136] Moore (1992, p.67)

standard interpretation Moore here is questioning whether goodness is a natural property based on the differences in cognitive significance between questions containing the predicate 'good' and questions containing the predicate 'being that which we desire to desire'.

Altman also tries to find supports for his interpretation with what Moore says a bit further down the same passage. Altman says of this passage,

> "Moore reinforces the idea that his argument concerns the intelligibility of asking about the good, not of the objects which possess a given natural property, but of the property itself. **He writes that, regardless of which property a naturalist theory picks out as identical to good, "Everyone does in fact understand the question, 'Is this good'".** Accordingly, his argument appears to involve the thesis that, for any natural property (or set of them), the question of whether the property is good has meaning: the fact that "everyone" would understand a question establishes that it has meaning".[137]

To see why this is a misreading I need to quote the passage from Moore more fully than Altman does. Moore writes:

> "It is very easy to conclude that what seems to be a universal ethical principle is in fact an identical proposition; that, if, for example, whatever is called "good" seems to be pleasant, the proposition "Pleasure is the good" does not assert a connection between two different notions, but involves only one, that of pleasure, which is easily recognised as a distinct entity. But whoever will attentively consider with himself what is actually before his mind when he asks the question "Is pleasure (or whatever it may be) after all good" can easily satisfy himself that he is not merely wondering whether pleasure is pleasant. And if he will try this experiment with each suggested definition in succession, he may become expert enough to recognise that in every case he has before his mind a

---

[137] Altman (2004, pp.397-8)

unique object, with regard to the connection of which with any other object, a distinct

question may be asked. Everyone does in fact understand the question "Is this good?"

When he thinks of it, his state of mind is different from what it would be, were he asked,

"Is this pleasant, or desired, or approved?"[138]

When Moore talks about "pleasure" here he is not talking about the property that a naturalist

might identify with *goodness*, rather, he is talking about the set of objects that have that property,

i.e. objects that are pleasant. This can be seen by the fact that he says that according to the

naturalist "Pleasure is the good" concerns only one entity. Moore always uses the phrase "the

good" to mean the set of good things, not the property *goodness* which he denotes simply with the

adjective "good".[139] If according to Moore "Pleasure" is identical to "the good" on the naturalist

picture, then "Pleasure" must denote a set of objects with a natural property, and not, *pace*

Altman, the property itself. So, when Moore uses as an example the question "Is pleasure (or

whatever it may be) after all good?", he is not trying to demonstrate to us that goodness can be

predicated of some natural property. And, when he then suggests "experimenting" by asking "Is

this good?" for each naturalist suggestion and seeing if we understand the question, "this" does

not refer to a natural property, but rather, the set of objects with that natural property. Hence,

Altman's interpretation is a misreading of Moore's text.

Additionally, the fact that Moore is talking about the fact that everyone "understands" the

question, "Is this good?", does not lend support to Altman's interpretation. Altman wants us to

believe that the talk of "understanding" favours his interpretation over the Standard

Interpretation because on his picture Moore is interested in the fact that questions like 'Is good

good?' do not make sense whereas on the Standard Interpretation Moore is concerned with the

---

[138] Moore, (1993, pp.68-9)
[139] See e.g. ibid. pp.60-1: "I do not mean to say that *the* good, that which is good, is thus indefinable… It is just because I think there will be less risk of error in our search for a definition of 'the good' that I am now insisting that *good* is indefinable. I must try to explain the difference between the two. I suppose it may be granted that 'good' is an adjective. Well 'the good', 'that which is good', must therefore be a substantive to which the adjective 'good' will apply: it must be the whole of that to which the adjective will apply, and the adjective must *always* truly apply to it."

fact that questions like 'Is pleasure pleasure?' are tautological. The reason he thinks his interpretation is to be preferred is that tautological questions are perfectly understandable even if they are pointless, so the fact that closed questions are tautologous and open questions are not, cannot be Moore's primary concern. But the fact that Moore talks about "understanding" here is not very significant. What Altman neglects to mention is that in this part of the text Moore has moved on to refuting the suggestion that "good" is meaningless. What he is doing is showing that since we understand the question, "Is this good?", the predicate "good" clearly has a meaning. Then to back up his other point, Moore is making clear that we understand the meaning to be distinct from the meaning of terms like "pleasant", "desired" and "approved".

Let us return now to the "key sentence" for Altman's interpretation:

> "The hypothesis that disagreement about the meaning of good is disagreement with regard to the correct analysis of a whole, may be most plainly seen to be incorrect by consideration of the fact that, whatever definition may be offered, it may always, be asked, with significance, of the complex so defined, whether it is itself good."

Though it is admittedly a bit unclear on first reading, it now seems that "the complex" Moore refers to is in fact the set of objects which possess the complex natural property that the naturalists identify with *goodness*, and not the property itself. Moore is not asking whether the complex property naturalists identify with *goodness,* has the property *goodness*, rather, he is pointing out to us that we can ask of the set of *objects* with that complex natural property whether it also has the property *goodness*; suggesting that those properties cannot be the same. If it seems strange to understand "the complex" here to refer to a set of objects with a complex property, and not the property itself, recall that Moore elsewhere uses the definite article to distinction between objects with a property and the property itself. Moore uses "the good" to denote objects which fall under the extension of the adjective "good", and the adjective "good" to refer to the property *goodness*. I suggest he is doing the same with "the complex" in this passage.

What of the fact that Moore seems to use "with significance" and "intelligible" interchangeably? Altman uses this, plus the fact that trivially true statements like "pleasure is pleasure" are intelligible, as evidence against the standard interpretation. Altman says that he takes "intelligible" in "a straightforward way" to mean "possesses meaning". Since he thinks that 'Is good good?' does not possess meaning, whereas 'Is pleasure good?' does, Altman can explain why Moore seems to use "intelligible" where he had previously used "with significance". If the standard interpretation meanwhile required that we understand "intelligible" to have some other, very non-standard, meaning then that might count against that interpretation.

However, there is a perfectly ordinary sense of "intelligible" that the standard interpretation can use. If someone, in an ordinary context and completely out of the blue, were to ask me, "Am I currently dead?", I would be completely puzzled by their question even though it has meaning (in the sense that it is grammatically well-formed and the words in it are all meaningful).[140] Because I assume that they know that they are alive, because I expect people not to ask questions to which they already know the answer and because there is not any obvious alternative question that they might be implicating (in Grice's sense of the word), I will be totally stumped. It would be perfectly felicitous in the context for me to say, "I'm sorry but your question is unintelligible to me. I do not understand what you're asking". In other words, I might say that a question is unintelligible to mean that it is unintelligible why the question is being asked. Similarly, the closed question, 'Is pleasure pleasure?', can be said to lack an intelligibility that 'Is pleasure good?' has. There is no reason to ask me if pleasure is identical to pleasure, since the answer is obvious so I might well say that the question is unintelligible. So, the standard interpretation has no problem on this count.

---

[140] By 'ordinary context' I exclude, for example, contexts in which we are playing a murder mystery game.

I conclude then that even the passage Altman brings up to provide textual support to his interpretation of Moore's argument does not in fact provide support to his interpretation.

Chapter 3: The Normativity Argument

For most of the time since the publication of *Principia Ethica* the open question argument has been seen as the most significant objection to naturalism but in light of now well-known problems with the argument anti-naturalists have increasingly argued that naturalism can be shown to fail in other ways. Some philosophers think that even if naturalists can say that normative terms are distinct in meaning from natural terms, it is just obvious that normative *properties* could not be natural *properties*. The idea that normative properties might be identical to certain natural properties is seen by some as obviously mistaken, or even absurd.

Part of the debate might come down to a misunderstanding of what naturalism is. It is sometimes said that naturalists identify normative properties with *non-normative* properties.[141] If that were the case then it would be easy to see why naturalists are accused of missing something crucial out of their account of normative properties. Were *goodness*, for instance, identical to some non-normative property N then goodness would not be normative since N is not normative and goodness is identical to N. However, the objection is too quick. The naturalist does not identify normative properties with *non-normative* properties; rather, they identify them with natural properties. Those natural properties, they are committed to saying, are normative. The anti-naturalist might then insist that this cannot work since natural properties could not be normative, but they need to supply independent arguments for that claim, otherwise they just beg the question against the naturalist.

One also comes across the idea that naturalism, *qua* a form of reductionism, entails that normativity does not really exist or that it is at least, in some important sense, *less real* than the non-naturalist believes it to be. We find this in Graham Oddie's *Value, Reality and Desire* where he writes, "That which is reducible is less real than that to which it reduces"[142]. These complaints

---

[141] e.g. Miller (2003), Cline (2015).
[142] Oddie (2005, ch.1)

seem to me to miss the mark. The naturalist thinks that normativity is just as real as anything else that exists.[143] They believe that things really are good, bad, obligatory, etc. Certainly, they have a different account to other philosophers of what something's been good, bad, obligatory, etc. consists in, but that account is not committed to saying that normative properties are 'less real' than other properties. Following the rejection of vitalism and the adoption of the idea that *being alive* reduces to a number of purely physical states and processes, we have not started saying that *being alive* is somehow 'less real' than the physical states and processes that being alive reduces to or consists in. Similarly, we should not say that anyone who thinks that the normative reduces to the natural thinks that normativity is somehow less than fully real.

Following G.E. Moore, philosophers have also invoked Bishop Butler's dictum - "Everything is what it is, and not another thing" - to suggest that reductions are unworkable. The thought is that when we offer a reduction of a normative property A to natural property B, we deny that A is A and say that it is B instead. We can find this kind of idea in Thomas Nagel, who writes: "If values are objective, they must be so in their own right, and not through reducibility to some other kind of objective fact. They have to be objective values, not objective anything else".[144] That a thing is not itself is of course absurd. But this objection is unfair to the naturalist; if goodness is happiness-maximisation then goodness is still goodness and happiness-maximisation is still happiness-maximisation.[145] And, to say that goodness reduces to happiness-maximisation is not to say that goodness is not really a normative property, rather, it is to say that happiness-maximisation is a normative property.

---

[143] One might argue, along Platonist lines, that since naturalists think that normativity is natural they must think that it is in some sense less real than the kinds of perfect *sui generis* non-natural qualities that non-naturalists take normative properties to be. But the idea that the natural world is less real than some other reality is, I think, not one that has been at all established and it is not even clear to me what it means, so I think that we can avoid this possibility. Moore offers a nice criticism of this kind of view (1903, p.166-7).

[144] Nagel (1986, p.138). For discussion see Schroeder (2005, p.8)

[145] Frankena (1939, p.472), Prior (1949), Baldwin (1993), Sturgeon (2003), Schroeder (2005).

However, even once we correct these misunderstandings, naturalism still strikes many philosophers as implausible. According to these philosophers, whatever their commitments *viz* the meaning of normative terms or the content of normative concepts, naturalist accounts cannot succeed because they all make a fatal fundamental error in identifying normative properties with natural properties. By committing themselves to this metaphysical thesis naturalist accounts give up any hope of capturing the sense in which normative properties are normative. This is the *Normativity Objection*.

When an action has some natural property like *minimises suffering,* that fact obviously could not be the same as that action's *being right.* If rightness were just the same as some natural property then rightness would not have the particular normative quality that it in fact does have, or so these philosophers believe. This kind of normativity argument is pressed against the naturalist in a few different ways. In what follows I look at a few different cases for the normativity argument, leaning heavily on Derek Parfit's discussion in the second and third volumes of *On What Matters,* since Parfit goes into more detail than non-naturalists tend to.

§1 - The Just Too Different Intuition

Sometimes non-naturalists argue that we simply know that normative properties are not natural on reflection. In identifying the normative with the natural the naturalist fails to capture the sense in which the normative is of its own special, *sui generis* kind. Normative properties, it is alleged, are intuitively distinct in kind from natural properties; to suggest otherwise is to commit a category error. Similar kinds of intuitions drive the case against reductions in other areas of philosophers, such as reductions in philosophy of mind of phenomenal states to physical states in the brain.

In part six of *On What Matters* Derek Parfit uses this intuition to form what he says is a "decisive" argument against naturalism. Parfit makes his case quickly and simply. Some things are very obviously not of the same kind, Parfit gives the following examples,

> "Rivers could not be sonnets, experiences could not be stones, and justice could not be – as some Pythagoreans were said to have believed – the number 4."[146]

The same, though "much less obviously"[147], goes for natural facts and normative facts.[148] Given the meaning of normative terms it could not turn out that they refer to natural properties, just as given the meaning of 'heat' it could not have turned out that heat was a particular shade of blue. Normative facts, if they exist, are in their own separate, distinctive category like mathematical facts. To argue that normative or mathematical facts are the same as some fact of another kind is to make a basic category error.

Parfit uses as his target naturalist theory a kind of reasons internalism like that defended by Bernard Williams; roughly, the position that the normative property *is what you ought to do* is the same as the natural property *is what would fulfil your present fully informed* desires.[149] Parfit thinks that it is obvious that Williams-style internalism cannot be correct. To illustrate this point Parfit turns to an example named *Burning Hotel* the details of which are as follows:

> "You are in some Burning Hotel, and you can save your life only by jumping into some canal. I am outside your hotel, which I know to be on fire, and I can see you at some window above the canal".[150]

The claim, 'You ought to jump', in this case could not state the same fact as, 'Jumping would do most to fulfil your present fully informed desires, or is what, if you deliberated in certain naturalistically describable ways, you would choose to do', given the difference in meaning

---

[146] Parfit (2011b, p.324)

[147] ibid., p.325

[148] We need not assume that there are irreducibly normative facts for this objection to work (ibid., p.327). Instead, we could just defend the claim that natural facts could not be normative. Error theorists could then accept this objection. So too, could some expressivists (e.g. Ayer (1936), Gibbard (1990) but not Blackburn (1998) or Gibbard (2003)).

[149] Williams (1981, ch.8).

[150] Parfit, (2011b, p.283)

89

between these claims. The fact that you ought to jump just could not possibly be the same as a causal or psychological fact. It might be thought that this objection misses the mark against synthetic naturalists who do not think that normative and natural claims have the same meaning, but Parfit does not require that these claims have the same meaning on the naturalist's account. Parfit's argument just requires that we can accurately tell from the respective meanings of normative and natural claims that they refer (if they are true) to facts in different basic categories.

Some might argue that this cannot be right because there is no sharp distinction between the normative facts and natural facts as evidenced by the fact that truths containing 'thick concepts' are partly normative and partly natural. For example, to claim that some act was generous might be thought to imply both that it was *proto tanto* good and that it involved the transfer of resources from one person to another person or persons. Parfit argues that this response does not work because we can distinguish between the purely normative and purely natural parts of such concepts (as I just did with the concept *generous*) and see that these parts are evidently not of the same kind.[151]

§2 - Responses

§2.1

David Enoch, a non-naturalist like Parfit, has a nice term for the intuition driving the thought that the normative just cannot be natural, he calls it the "Just Too Different Intuition".[152] But while Enoch finds this intuition plausible he does not think that simple appeals to it constitute any kind of fatal objection to naturalism.[153] He notes that many people simply do not share this intuition, so the naturalist and the non-naturalist find themselves at an impasse from which he thinks no fully satisfying general argument against naturalism can be generated. He writes,

---

[151] ibid. p.326
[152] Enoch (2011). Other philosophers who report having these kinds of intuitions include Donagan (1981), Johnston (1989), Dancy (2005, p. 132), Scanlon (2000, pp. 57–8) and Huemer (2005)
[153] Enoch (2011), ch.5

"[A]ny argumentative move here will likely start either with accepting or with rejecting the just-too-different intuition, thus in a way begging the question against the other party to the reducibility debate."[154]

Enoch, in my view, judges the dialectical situation correctly. The normativity argument is not one that you feel will convince many who were not already strongly predisposed against naturalism. To illustrate, see how Michael Huemer, arguing along very much the same lines as Parfit, presents one argument against naturalism:

(1) Normative properties are radically different from natural properties.
(2) If two things are radically different, then one is not reducible to the other.
(3) So, normative properties are not reducible to natural properties.[155]

The first premise of this argument is just a flat denial of naturalism and the only people that are going to find it convincing are those that were already inclined to reject naturalism anyway. The same kind of issue seems to affect Parfit's normativity argument. The argument does not seem to amount to much more than simply saying that naturalism seems incorrect.[156]

§2.2

I think we should also worry about the examples used in Parfit's presentation of the argument. First, he points out that things like rivers and sonnets, or experiences and stones, are clearly of a different kind. There is an important disanalogy between these kinds of examples and the case of the normative and the natural, however. Facts about rivers do not supervene on facts about sonnets (or *vice versa*) and facts about experiences do not supervene on facts about stones (or *vice versa*) but it is widely agreed that normative facts do supervene on natural facts; an action that is right in one world could not be wrong in a world where all the natural facts were the same.[157] Plausibly, part of the reason that it seems obvious to us that stone-properties and experience-

---

[154] Enoch (2014), p.108
[155] Huemer (2005), p.94. Huemer actually talks about 'value properties' rather than 'normative properties'. I changed it just so that the argument was consistent with the rest of my discussion. I do not think that it is a philosophically significant change
[156] Schroeder (2009), Fleming (2015).
[157] Railton (2016) also points this out.

properties could not be the same is that their extensions are so different. We can easily imagine a world in which there are many stones but no conscious experiences, and *vice versa*. The same is not true of the normative and the natural; given the supervenience of the normative on the natural there must be some broadly natural property with the same extension as a normative property. So, the first kinds of examples that Parfit uses as analogies are potentially misleading.

Second, Parfit then illustrates how his argument applies to the case of the normative and the natural by returning to talk about Williams-style internalism and the *Burning Hotel* example. He argues that we can clearly see that in this example the fact that it jumping the course of action that would do most to fulfil my present fully informed desires is different to the fact that it is what I ought to do. If we suppose for the sake of argument that that is true, it does not tell us that every other form of naturalism cannot work. Again, part of what is doing the work here might just be that internalism offers an extensionally inadequate theory of what we ought to do. Faced with the example I might think, 'Yes, Parfit is right, *what I ought to do* could not be the same as *what does most to fulfil my present fully informed desires*, since it would still be the case that I ought to jump even if jumping did not do most to fulfil my present fully informed desires.' But this would just show that internalist theories face a kind of extensional inadequacy problem, it would not show that naturalists in general are guilty of a category error. Non-naturalists often argue in the following way; they take one very simple and implausible naturalistic reduction of a normative property, point out that it is intuitively unappealing, and then claim without much argument that this result must generalise to show that no naturalistic reduction is possible. That may not be exactly what Parfit is doing with the normativity argument, but his use of the *Burning Hotel* example to support it does, I think, trade on the same kind of intuitions.

§2.3

Even if we give credence to the just too different intuition it hardly seems, as Parfit claims, that it would be decisive. Naturalism may simply be a view which sounds a bit implausible to us initially

but which makes up it for it by having significant theoretical advantages, e.g. ontological parsimony, a straightforward explanation of the supervenience of the normative on the natural, etc.

Huemer does not agree. He argues that we can simply see that all forms of naturalism are false and that beyond that no further argument is needed. Someone who cannot see that, say, goodness is not a different kind of thing from any natural property just does not understand what goodness is or what natural properties are, and they need not be argued with further. [158]

Against this, one can point to the fact that many people clearly have no such intuition to cast doubt on the idea that the falsity of naturalism is quite so obvious and that naturalists have no understanding of normative concepts. Additionally, we can point to people who have strong intuitions that normative properties are natural properties. In my experience these people tend to be utilitarians who think that it is just obvious that rightness is maximising happiness (or something else along similar utilitarian lines like desire-satisfaction) and that people who have views to the contrary are lead astray by self-interest, religious belief, residual attachments to outdated ideas like honour, or a kind of squeamishness around making sacrifices for the greater good.[159] I do not think that this view is plausible, but I do not know any independent reason for thinking that these intuitions are unreliable while the just too different intuition is not.

Additionally, many people who may not have the intuition that normative properties are natural, *do* have intuitions that seem to entail it. For instance, many people have the intuition that the only facts there are are broadly natural facts, and they have the intuition that there are some normative facts, e.g. that it is a fact that torturing people for fun is wrong. Together these intuitions would seem to entail that normative facts must be natural facts.[160]

---

[158] Huemer (2005, ch.4)
[159] e.g. E. and M. Clark (1941), Peter Singer (1993), Harris (2010).
[160] It might be thought that naturalism is not entailed by these intuitions because quasi-realism is an option. There is not space to discuss the matter properly here but quasi-realists have in recent years increasingly turned away from

And of course, many people have the intuition that non-naturalism is fantastical or even downright ridiculous. Lots of people react to non-naturalism with a kind of disbelief that anyone might suppose that normative properties are *sui generis* and non-natural, and that we come to know about them through some mysterious faculty of 'intuition' (Huemer) or through 'non-causal responses' (Parfit). The point is not that these intuitions suggest that naturalism is right, rather, it is that we would need some reason beyond the bare fact that some people have the just too different intuition to doubt that naturalism is right. There may be some story about why we should take the just too different intuition seriously but not any conflicting intuitions, but I do not think that non-naturalists have yet given us it. With all these different, conflicting intuitions in the air one is tempted to think that the fact that some people have a particular given intuition on the matter is, by itself, really neither here nor there.

§2.4

Even if we accept that people feel that the normative is intuitively different to the natural there is scope for claiming that that intuition is consistent with some forms of naturalism, for it is not clear whether the intuition, to the extent we have it at all, really concerns normative *properties* as opposed to normative *concepts*. Suppose it is said that the just too different intuition is supposed to be a pre-theoretical feeling that many of the folk are disposed to have. It seems plausible to wonder in that case whether folk intuitions are really fine-grained enough to distinguish between issues about normative *facts* and *properties* on the one hand and normative *concepts* or the meaning of normative *terms* on the other. If asked whether goodness seems to be a different kind of thing to any natural property someone might ask themselves, 'Well, I can think that something is good without thinking that it has any particular natural quality, so being good must be a different kind of thing to natural qualities.' I suspect that in this kind of way the kinds of intuitions underlying

---

claiming that all facts are natural facts precisely because it is fits awkwardly with their commitments to both the irreducibility of normative concepts on the one hand and a kind of deflationism about fact-talk on the other. (e.g. Toppinen (2014), Ridge (2014), Price (2013), Bex-Priestley (2018))

the open question argument might get mistaken for the just too different intuition. I think that Parfit encourages this with the way he discusses *The Burning Hotel*. He writes:

> "You know that, unless you jump, you will soon be overcome by smoke. You might then believe, and tell yourself, that you have *decisive reasons* to jump, that you *should, ought to,* and *must* jump, and that if you don't jump you would be making a *terrible mistake*. If these normative beliefs were true, these truths could not possibly be the same as, or consist in, some merely natural fact, such as the causal and psychological facts stated by (C)."

Parfit starts by getting us to think about normative thought processes, and then he moves straightaway to claiming that the referents of these thoughts could not possibly be natural facts. It is easy to feel that a kind of sleight-of-hand is going on here. It may be intuitive that *thinking* that I ought to jump could not be the same as *thinking* that jumping does most to fulfil my present desires, but is it so intuitive that the *property* referred to by one thought could not be the same as the *property* referred to by the other? Compare this example:

> You know that there is water in the glass. You might then believe, and tell yourself, that it is drinkable, that light passes through it, that it is the same material that fills our lakes and rivers. If these beliefs were true, these truths could not possibly be the same as, or consist in, some mere *chemical* fact such as the facts stated by claims about hydrogen and oxygen atoms.

The first sentence leads us down the garden path, prompting us to see a difference between folk concepts and chemical concepts, only then for the second sentence to push us to the intuition that water facts and chemical facts are not the same. If this is the kind of reasoning process going on in the normative case however then the intuitions are no threat to synthetic naturalists who agree that normative concepts are very different from natural concepts. It is not even a worry for analytic naturalists who have developed multiple strategies for accounting for the intuitive difference in cognitive significance between normative and natural concepts, as seen in the last two chapters.

§2.5

Alternatively, it might be that the just too different intuition is not supposed to be some pretheoretical hunch had by the folk, but is instead meant to be a kind of compelling conviction that assails one once one has reflected on and digested issues in the metaethical literature. It would be uncharitable to suppose that philosophers with this intuition were not properly distinguishing *properties* and *concepts* that pick out those properties.

But if this is the case then, again, it is striking that there are so many naturalists/reductionists around. Additionally it is reasonable to suspect that the just too different intuition might be a product of 'theory contamination'; that is, an intuition that arises out of someone's theoretical commitment to non-naturalism as opposed to an antecedent datum supporting it. If the just too different intuition just amounts to the thought had by non-naturalists that non-naturalism seems right and naturalism seems wrong, then it does not seem very troubling.

Huemer does speculate that the reason that people are opposed to non-naturalism is because, "it does not fit with the spirit of the modern age" and that because of "biases", including irrational "scientistic", "politically correct" or "cynical" attachments, people have been lead to think that opposing metaethical views might be workable.[161] If that were the case then we might think that we can disregard intuitions that normative properties might be natural properties on the basis that philosophers who have these intuitions did not come to their views in rational or reliable ways. But, first, I do not think that Huemer says enough to ground this piece of speculation so I think that we can safely pass over it. Second, we can point to examples of philosophers who did not seem to have the just too different intuition but who either predate the 'modern age' (e.g. Jeremy Bentham) or who do not seem to be scientistic, cynical or unduly concerned with being politically correct (e.g. Peter Geach).

§2.6

---

[161] Huemer (2005, pp.240-247)

Finally, even if we suppose for the sake of argument that the just too different intuition is widespread and that it really does concern *properties* and not just *concepts*, why should we think that it has any probative force? Perhaps time and motion do not intuitively seem to be relative and perhaps water does not intuitively seem to be H_2O but this does not lead us to reject any theories which suggest otherwise.

Parfit and Huemer both argue that there is an important difference between the just too different intuition and these scientific cases.[162] It did not seem obvious to competent users of the pre-scientific concept <water> that water could *not* turn out to be H_2O. By contrast, argue Parfit and Huemer, it does seem obvious that given the content of the concept <good> the property *good* could *not* turn out to be some natural property. Parfit thinks that it follows from the very meaning of 'good' that *good* could not be a natural property. To suppose otherwise is just to change the subject and talk about something other than *good*. The same is not true of thinking that water is not a molecule or that time is not relative.

But are things really this clear cut? Mark Schroeder gives the following example:

> It does seem pretheoretically obvious that splitting a drop of water must always yield smaller drops of water. But the thesis that water is H_2O entails that this is false. It entails that water is made out of smallest bits. But water does not seem to be made out of smallest bits; it seems to be continuous.[163]

The same goes in the case of time. Here is an example from Eric Margolis and Stephen Laurence:

> [N]othing seems more inviolable than the belief that if a is older than b today, then a will be older than b next year too; but Relativity Theory implies that this isn't so. [164][165]

---

[162] Parfit (2011b). Huemer (2005, ch.4).
[163] Schroeder (2009, p.202)
[164] Laurence and Margolis (2003), p.271.
[165] Another example might include the idea that heat is intuitively a kind of substance (Railton's example, 2017). Putnam's *Robot Cats* (1962a) example may also be used.

It may have seemed obvious to people that certain claims which entail that water is not a molecule or which entail that time is not relative must be true, but these people are mistaken. Other examples could include the idea that heat is intuitively a kind of substance[166] or Putnam's *Robot Cats* example.[167] At first blush it might seem obvious to some that it follows from the concept <cat> that cats are animals, but if prompted to imagine a world in which it turns out that what we believe to be cats are in fact robots many (though not all) have the intuition that we would have learned that cats are robots and *not* that there are no cats, suggesting that it is not a conceptual truth that cats are animals after all.

I am not suggesting that it was part of the content of the pre-scientific concept <water> that water is not a molecule; we are not inclined to say that people living prior to circa 1750 were not using the same concept as us when they spoke about 'water'. Rather, I am suggesting that what may seem obvious to us regarding some object need not be a very good guide to the content of the concepts that picks that object out. This opens up the possibility that even if it seems intuitively obvious that the property *good* could not be a natural property (which I am not convinced of), that does not necessarily tell us that it follows from the meaning of the term 'good' or the content of the concept <good> that *good* could not be a natural property.[168] Laurence and Margolis point out,

> "There's always the possibility that an apparent a priori analysis is really just a reflection of a lack of theoretical imagination." [169]

I think that the same could be said about the normative case. Even if it is true that the just too different intuition is widespread and stable, we need not think that it represents a *conceptual truth* or forms part of the meaning of our normative terms, as Parfit and Huemer seem to think. It may just be like the false intuition that time is absolute.

---

[166] Railton (2016)
[167] Putnam (1962a)
[168] Putnam (1962b)
[169] Laurence and Margolis (2003), p.271.

All in all then, the normativity argument does not seem too convincing. It is not obvious that people, by and large, have the intuition that normative properties cannot be natural properties and even if they do it is not obvious that that gives us much reason to reject naturalism. It is hard to know whether Parfit intends to put much stress on this argument. While he does say that he thinks it is "decisive", he also says that it is "much less obvious" that the normative and the natural are distinct than that, for instance, rivers and sonnets are.[170] He also goes on to say that despite his many years of attacking naturalism, some of the central claims of synthetic naturalism can seem plausible to him, which presumably would not be the case if naturalists were guilty of an obvious and egregious category error.[171] In any case, the normativity argument as he first presents it does not seem to put much pressure on naturalism.

§3 - The Normativity Argument Revived

Sometimes when philosophers present something like the normativity argument they go further than simply saying that naturalism is obviously false on reflection, they also try to identify a feature that normativity has that the natural lacks in order to prove, by Leibniz's law, that one cannot be identified with the other.

Often they have trouble articulating clearly just what this feature is. Richard Joyce, for example, argues that naturalism cannot make sense of the "practical oomph" that morality in particular has.[172] But Joyce never gives us a clear sense of what "practical oomph" is and even seems to suggest that it is not possible to explain this feature of normativity in an illuminating way.[173] At times, Parfit also talks as if this special feature of normativity is something totally ineffable.[174]

---

[170] (2011b, p.325)
[171] ibid., p.349
[172] Joyce (2006, ch.2, ch.6. 2011 p.)
[173] Though he tells us (2011, p.525) that it is *not* a categorical reason-giving force, which would seemingly be the most obvious candidate.
[174] e.g. Parfit (2016, pp.73-4.) cf. Wittgenstein (1965)

Without being especially clear on what the *explanada* are, it feels unfair to burden the naturalist (and everyone else who is not a non-naturalist) with having to explain it. Citing some mysterious, ineffable quality that normativity has, and arguing that no naturalist could hope to capture it is no real advance on simply stating that naturalism seems wrong. If the naturalist has to deny that normative properties are normative in some sense that cannot be expressed or clearly stated then it does not seem costly to do so.

At other times though, anti-naturalists are a bit clearer about what normativity requires. Parfit argues that naturalists can capture *some* of the senses in which a claim can be normative, but not the sense in which they can involve *reasons to act*. Here is how he spells out this idea:

> On the *reason-involving* conception, normativity involves reasons or apparent reasons. When I call some claim
>
> > "*normative* in the *reason-implying sense*, I mean roughly that this claim asserts or implies that we or others do or might have some reason or apparent reason." [175]

Other non-naturalists like Jean Hampton, T.M. Scanlon, Joseph Raz and Jonathan Dancy also seem to have something like this in mind when they say that naturalists miss out on the normativity of the normative.[176] [177]

§4 – Response

There might be scope for the naturalist to bite the bullet. At least when it comes to *moral* claims, the naturalist could argue that we do not *necessarily* have any reason to do that which we morally ought to do. Given that most people usually have desires or goals which are furthered if others are treated decently, and if their welfare is promoted, etc. it so happens that we usually have a

---

[175] Parfit (2011b, p.267)

[176] Hampton (1998), Scanlon (2000), Raz (1999) and Dancy (2006)

[177] Another way to understand what is meant by a property's being normative might be that believing that some action has that property necessarily entails some link to believer's motivation. I consider the issue of whether naturalists can and should say that normative property are normative in this sense in chapter 5.

reason to do what is right. We can use this fact to try to explain why it seems to us that behaving morally is rational. But nothing on this picture guarantees that it is always rational to behave well. Philippa Foot puts this point in a memorable way: "We are not conscripts in the army of virtue, but *volunteers*".[178]

Faced with examples of strange hypothetical agents who by stipulation have no desires or end furthered by behaving morally naturalists could either insist that it is actually more intuitive that they have no reason to behave morally, or they could argue that such a person is not possible.[179] But, among other problems, this kind of response does not translate very well to handling normative concepts like <all things considered ought>. The idea that I have no reason to do what I morally ought to do might be coherent, whereas the idea that I have no reason to do that which all-things-considered I ought to do borders on unintelligibility. It seems reasonable to expect that an account of normative concepts should explain why at least some of them seem to entail claims about reasons.

A better response for the naturalist is to agree that there is a reason-implying sense of normativity and to capture it in naturalist terms. The naturalist can agree with the non-naturalist that if I ought *all-things-considered* to do something then I have a reason to do it. Take the Williams-style reasons internalist theory that Parfit often using as his target. According to this view, I have a reason to do something if it would fulfil one of my present fully-informed desires, and I ought *all-things-considered* to do something if it would *do most* to fulfil my present fully-informed desires. On this picture it is easy to see why the concept <all things considered ought> is reason-implying; if something would do most to fulfil one of my present fully-informed desires then that implies that I have a reason to do it, since having a reason to do something, on this

---

[178] Foot (1979, p.170.)

[179] Sharon Street, Bernard Williams, David Sobel and others argue take the first kind of response. Mark Schroeder, Michael Smith and Kantian constructivists use the latter kind of response. These philosophers would not however deny that we have reason to do that which *all things considered* we ought to do.

picture, consists in its fulfilling one of my present fully-informed desires.[180] In other words, if something does most to fulfil my present fully-informed desires then *a fortiori* it must fulfil one of my present fully-informed desires. So, on this naturalist account, if something is what I ought to do all things considered that (almost trivially) entails that I have a reason to do it.

Other naturalists can adopt the same pattern, they can say that if something has the natural property picked out by <all things considered ought to> then that implies that it must also have the natural property picked out by <has a reason to>. The naturalist can capture why other normative concepts are *reason-implying* along similar lines. Suppose that the non-naturalist says that an action's being *good* implies that there is a reason to do it. The Williams-style internalist could say, for example, that an action is good iff it would fulfil someone's present fully-informed desire. This would straightaway imply that if an action were good for a person to perform then they also have a reason to do it. And, they can as easily explain why the normative concept of <has a reason to> is itself normative in the *reason-involving* sense, since every claim trivially implies itself.

If the kind of reasons fundamentalism/buck-passing account of normative concepts favoured by so many non-naturalists is correct then the way ahead for naturalists is clear. They can offer a naturalistic reduction of the concept <has a reason to> and then show how other normative concepts reduce to that concept, i.e. show that when an action has the natural property picked out by a normative concept it must also have whatever property is picked out by the concept <has a reason to>.[181] Mark Schroeder suggests that all naturalist reductionisms should take this form.[182] But if, instead, something like the concept <good> is fundamental, then naturalists

---

[180] A potential caveat concerns cases where no possible action would fulfil my present fully-informed desires. In that case, trivially, any action would do most to fulfil my present fully informed desires even though no action would fulfil any of my desires. A think that the nternalist could easily finesse their account to handle this but since I am not interested in defending internalism specifically I do not pursue the best way to do this.

[181] Schroeder (2005, 2007), Smith (1994), Copp (1995)

[182] Schroeder (2007, Ch.5).

could still show that normative concepts are reason-implying, they would just have to then also show how the concept <has a reason to> reduces to <good>.

Of course, this does not show that Williams-style internalism or any other naturalist reduction actually succeeds in explaining why normative concepts are reason-implying. To do that they would need to put forward a specific reduction of the concept <has a reason to> which captures the things we want to say about reasons, which is in other ways theoretically plausible, and which has the right links to the other normative concepts. But nothing the non-naturalist says suggests that this is in principle impossible. So, I think that naturalists are perfectly safe from this version of the normativity argument.

There is another slightly different way that Parfit argues that naturalists cannot explain why normative properties are normative. Sometimes when Parfit talks about the sense in which concepts can be normative he does not say that they must involve reasons *per se*, but that they must involve *irreducibly normative reasons*.[183] It is *this* that shows that naturalism is unworkable, he thinks. Clearly, naturalists cannot say that normative concepts are normative in *that* sense, but that does not matter because the complaint straightforwardly begs the question against the naturalist.[184] Whether there are irreducibly normative reasons, or whether reasons admit of naturalistic reduction, is exactly what is at issue between the non-naturalist and the naturalist, so this way of framing the normativity objection is no advancement on simply asserting that normative properties do not seem to be natural. I move on now to look at Parfit's final formulation of his normativity argument.

§5 - The Normativity Argument Revived Again

§5.1

---

[183] e.g. Parfit (2016, p.70)
[184] Frankena makes this point about a similar argument (1939, pp. 473-4)

In the third volume of *On What Matters*, Parfit concedes that the naturalist has a plausible response to his normativity argument as he originally set it out. This time he uses in his examples hedonic act utilitarianism (HAU), the view that says that acts are right if and only if, or just when, these acts minimize the total sum of suffering minus happiness. If HAU is correct then an action's having the natural property *minimizes suffering* makes it the case that that action has the normative property *is right*. If both naturalism and HAU were correct then *minimizes suffering* would also be identical to *is right*.

Previously, Parfit had just flatly denied that the natural property that *makes* acts right could also be the same as the property *rightness*. This, he thought, counted strongly against naturalism which is committed to saying that a property like *minimizes suffering* makes that act have the normative property *is right* and is identical to *is right*. Discussing how a hedonic act utilitarian might respond to this objection, Parfit writes:

> "These Naturalists might reply:
>
>> Your claims are mere assertions. You agree that
>>
>> (J) having molecular kinetic energy both non-causally makes an object hot and is the same as being hot.
>>
>> We believe that, if HAU were true, we could similarly claim that
>>
>> (K) being an act that minimizes suffering both non-causally makes an act right and is the same as being right.
>>
>> Why do you deny that (K) could be true? Where is the disanalogy?
>>
>> This reply is, I concede, plausible. Given the similarity between (J) and (K), it may seem dogmatic to declare that being an act that minimizes suffering couldn't be the same as being right."[185]

---

[185] Parfit (2016, p.79)

As we can see in this passage, Parfit came to agree that his original normativity argument was question-begging, as I have been arguing.

He does however go on to suggest a new way to put his normativity argument. He does this by identifying a feature that the normative has that the natural does not, but this time he does not make the equally question-begging assumption that normative claims must involve *irreducibly* normative reasons. This argument is quite different from the other formulations, so much so that I am not sure that it should really be called a normativity argument, but since Parfit considers it to be so I will address it here.

Parfit's new argument focuses on the relation, *non-causal making*. That a liquid is made of H_2O molecules *makes* it the case that the liquid is water, but a liquid's being made of H_2O molecules does not *cause* that liquid to be water; causes cannot the same as their effects. Similarly, that a figure is closed and has three sides makes it the case that that figure is a triangle, but it does not *cause* the figure to be a triangle.[186] Parfit agrees that natural properties can *non-causally* make actions have particular normative properties. What he wants to argue is this could not be true if those natural properties were also identical to those normative properties.

§5.2

Parfit's first move is to argue that the claims (J) and (K) in the quoted passage above are not analogous. He writes:

> (J) and (K) are not, I believe, relevantly similar. The wording of these claims suggests that what corresponds to *being hot* is *being right*. But that is not so. Heat is *the property that can have certain effects*—such as causing certain sensations, melting solids, etc. What corresponds to heat is not *the property of being right*, but *the property that makes acts right*. What is relevantly similar to (J) is not (K) but

---

[186] Schroeder's example (2005), (2007, ch.4)

(L) being an act that minimizes suffering both non-causally *makes* an act have the property that makes acts right, and is the *same* as having the property that makes acts right.

I should have claimed that, unlike (K), (L) might be true. (L) is like the true claim that

(J) having molecular kinetic energy both non-causally makes an object hot and is the same as being hot.

Parfit's point here is subtle and potentially confusing. If you tell someone that something is 'hot' then you tell them that that thing has other, *distinct* properties like *boils liquids*, *melts solids*, etc., whereas telling someone that 'something has a certain molecular kinetic energy' does not tell them that it has those other properties. Parfit thinks that that is just part of the respective meanings of the expressions 'heat' and 'molecular kinetic energy'. (J) is informative because it tells us that when something has a particular molecular kinetic energy that makes it have the, other, different properties, e.g. *boils liquids*, that we associate with the term 'heat'.

Synthetic naturalists want to say that 'being hot' is analogous to 'is right' and that 'having a certain molecular kinetic energy' is analogous to 'minimises suffering' (or whatever natural expression that denotes the natural property that is identical to *being right*) but Parfit thinks that this is mistaken. 'Being hot' is more analogous to 'makes acts right', while 'being right' is more analogous to 'boils liquids'. This is because an object's having the property *boils liquids* is a result of it having the property *has a certain heat* (which is the same property as *having a certain molecular kinetic energy*). Since an action's *being right* is (assuming utilitarianism) a result of its having the natural property *minimises suffering*, *being right* is more analogous to *boils liquids* (the effect) than it is to *being hot* (the non-causal maker). Therefore, (J) is more like (L) than it is like (K).

With that cleared up, Parfit moves on to argue that a natural property could not both be the property that make acts right and be identical to the normative property *being right* just as the property *being hot* could not both non-causally make it the case that liquids boil and be identical to the property *boils liquids*.

Something cannot be the same as one of its properties, Parfit claims. The Sun cannot be the same as its brightness, for example. This goes for properties as well as for objects, e.g. a square cannot be the same as its property of *being square*, which in turn cannot be the same as its higher-order property, *being a shape property*. The claim that no property can be the same as one of its higher-order properties entails that if a property has a particular effect, then that property cannot be the same as its higher order property *being the property that has such-and-such an effect*. Additionally, where a property has a particular effect, that property could not be identical to the effect it has.

Parfit uses the following example:

> "(R) Just as the Sun couldn't be the same as the Sun's property of being bright, the Sun's brightness couldn't be the same as any of this property's higher-order properties, such as the property of being the property that makes the Moon shine. Nor could the Sun's brightness be the same the as the property of being made to shine, or of shining."

Even though the *making* relation mentioned in (L) is not, unlike in (R), a causal relation the same goes for (L), claims Parfit. The naturalist must say that there is some natural property N which makes act right but which is also identical to the normative property *being right*. However, N cannot be the same as its higher-order property *being the property that makes acts right*. N could also not be the same as the properties *being made to be right* or the property *being right*. Therefore, the naturalist's identification of N with *being right* must be incorrect.

Parfit then says that the naturalist might reply that the scientific analogy concerning heat shows that the following is true:

(U) the relation of *non-causal making* always implies *being the same as*.

But (U) is not true. *Non-causal making* is an asymmetrical relation, having molecular kinetic energy makes something hot, but not vice versa. Whereas, *being the same as* is a symmetrical relation. Additionally, *non-causal making* often holds between different properties, being killed by a meteorite makes you unlucky but the property *being killed by a meteorite* is not the same as *being unlucky*.

So, Parfit concludes, where an action's having a natural property non-causally makes it have a normative property, those properties cannot be the same.[187]

§6 - Response

The first thing to say is that the naturalist has no need to defend (U). No naturalist argues that because having a particular molecular kinetic energy both makes things have a particular heat and is the same as having a particular heat, so *being right* must be the same as whatever natural property non-causally makes it the case that acts are right. Rather, naturalists argue that examples like the case of heat show us that it merely *could* be the case that some natural property is the same as a normative property and that reference to that natural property can figure in explanations of why certain actions have that normative property.[188] In other words, the naturalist uses the heat example to show that there *might* be synthetic identity claims relating natural and normative properties. They do not use the example to try to show that normative properties *must* be natural properties. So while Parfit is correct to say that (U) is false, it is of no cost to the naturalist because they would not, and need not, defend (U).

The second thing to say is that if Parfit's argument were sound it would be a potential problem for his positive metaethical view too. Parfit is a reasons fundamentalist. He takes reasons to be in a sense fundamental to other normative properties. On this view, normative properties like *goodness* or *wrongness* are nothing over and above certain *reason* properties. For example, to call something good (in the reason implying sense) is, roughly, to say that certain facts about that thing's nature that would give us strong reasons to respond to it in a positive way, such as wanting it or using it.[189] I take it that Parfit would accept that if a thing's nature gives us strong reasons to respond to it in a positive way that *non-causally* makes it good. We can run Parfit's third normativity objection against his own view. If that objection works then *goodness* must be

[187] McNaughton and Rawling (2003) make a similar argument.
[188] See e.g. Railton (2016)
[189] Parfit (2011a, p.38). cf. Scanlon (2000, p. 97. 2014, p.2)

something distinct from *gives us strong reasons by its nature to respond to in a positive way*. Similar moves could be made, *mutatis mutandis*, against any view which reduces normative concepts to other normative concepts. So, like the open question argument, Parfit's argument overgeneralises; a good reason to suspect that it does not work.

With those preliminaries out of the way, I will move on to address his argument directly. Parfit's argument is that, *contra* naturalism, it could not be the case that where an action has a property that makes that action have another property, those properties are identical. That has some intuitive appeal. If one thing makes another, it sounds very odd to suggest that they could then be one and the same. But talk of 'making' in this case might be misleading because it suggests some kind of causal, creator-creation relationship. 'Making' in the normative cases, however, is supposed to be *non-causal* and so could be denote some relation compatible with naturalism. Naturalists can say that just as *being a closed a figure with three sides* both non-causally makes something be a triangle and is identical to the property *is a triangle*, so for some natural property, that natural property non-causally makes something have a normative property and is the same as that normative property.

 It will be useful to have labels for some properties in what follows:

> n – Maximises happiness.
> m – Being Right
> o – Being the property that makes acts right.

According to the naturalist suggestion on the table n is the natural property that makes acts right, o is one of n's higher-order property and m is the same property as n. Parfit argues that, if HAU is true, then n cannot be the same as o because o is one of n's properties and nothing is the same as one of its properties. Parfit is wrong to say that nothing can be the same as one of its properties, as we saw in chapter 2. The property *is a property* plausibly has itself as one of its properties. So does the property *is not a number*. Parfit gives us no reason for thinking that

normative properties are not also examples of properties that have themselves as a property. That issue does not really matter however since the naturalist does not think that n and o are the same property anyway; rather, they think that n and m are the same. I do not see why this part of Parfit's argument is even relevant. Parfit is trying to refute the suggestion that having the natural property that makes acts right is the same as *being right*, but the property that makes acts right on the naturalist account is n, not o.

Parfit does however also argue the property that makes acts right (n) cannot be the same as being right (m). The analogy he uses is of the property that makes the moon shine (the Sun's brightness) which he says cannot be the same as one of the Moon's properties *being made to shine.* Similarly, if n makes acts right, then n cannot be the same as *being right* (m).

But we can accept that if one of the Sun's properties makes the Moon have a property then those properties are not the same while arguing that it works differently in other cases. The Sun's brightness *makes* the moon shine by *causing* it to shine. We know that the Sun's brightness cannot be the same as the Moon's *shining* because we know that causes cannot be the same as their effects. But the *making* relation that holds between natural and normative properties is supposed to be *non-causal*, so we have no reason to think that the what goes for Parfit's example will generalise to non-causal cases.[190] Though Parfit assures us that what is true of examples of *causal making* will also be true of examples of *non-causal making,* he does nothing to ground that claim. His Sun example is a red herring.

Parfit does however touch on potential example of non-causal making too. It seems right to say both that heat is molecular kinetic energy and that having a certain molecular kinetic energy makes something hot. This seems like an example of non-causal making that holds between identical properties; this is why it is a good example for naturalists. However, as we have seen,

---

[190] Parfit agrees that on anyone's account the connection between something's natural property and the normative properties that it is made to have by having those natural properties, cannot be a causal connection. Parfit (2011a, p.299)

Parfit thinks that this example is not a good analogy for the relation between normative and natural properties. Saying that something is hot is really saying that it causes certain effects, e.g. boiling liquids, melting solids. So, when we say that having a certain molecular kinetic energy makes an object hot, what we are really saying is that it causes the object to have certain effect like *boiling liquids*. A more apt analogy for the relation between normative and natural properties then would be the relation between molecular kinetic energy and its effects, e.g. boiling liquids, melting solids etc. But just as a property like *boils liquids* could not be the same as the property that causes it - *having a certain molecular kinetic energy* - so *being right* could not be the same as the property that makes acts right. So, once we examine the analogy more closely, Parfit thinks, we see that it actually undermines naturalism.

This argument is not very convincing, however. There are two possibilities: either the relation between having a certain molecular kinetic energy and having certain effects like boiling liquids is a causal relation or it is not. If it is a causal relation then, just as with Parfit's Sun example, the example is disanalogous to what the naturalist wants to say about the relation between normative and natural properties. In other words, if we want to say that having a certain molecular kinetic energy *causes* liquids to boil, then the fact that the property *having a certain molecular kinetic energy* is distinct from the property *causes liquids to boil*, is of no importance, since the naturalist does not say that the relation between having a particular natural property and having a normative property is a causal one.

Alternatively, if the relation is *non-causal* then it is not clear that Parfit is right to say that the property *having a certain molecular kinetic energy* and the property *having such-and-such an effect* are necessarily distinct properties. In order to see this let me give a more specific example. Imagine some water is heated to 100 degrees centigrade at sea level so that it boils in a particular way (e.g. it evaporates at such-and-such a rate, it bubbles with a certain amount of force, etc.). Let us refer the property of having the amount of molecular kinetic energy that that water has, E. Having E,

we might say, *non-causally* makes it the case that that water (at sea level) boils in the particular way that it does.[191] It is not that having that amount of energy *causes* the water to boil in the way that it does, we might suggest. Boiling in the way that it does is plausibly just part of what it is for water (at sea level) to have that particular amount of molecular kinetic energy.

But if the relation between E and the water's boiling in a particular way is a non-causal relation, then it is not obvious that E and *is boiling in such-and-such a way* are necessarily distinct properties. I think that, *pace* Parfit, the reverse is actually more intuitive; it seems more plausible than not that they are the same property. Having a particular amount of molecular kinetic energy *just is* what it is for water at sea level to have the property of boiling in the particular way that it does in this scenario. The water's boiling in the way it does and its having a certain energy do not intuitively seem to be distinct qualities of the water, they are one and the same quality. If that is right then there can be cases of non-causal making between identical properties. This is just what the naturalist wants to say about the relation between normative and natural properties, which is that a certain natural property is identical to *being right* and that it makes actions right.

But even if Parfit is right about the water case, we can point to other cases. It is legitimate to say that the fact that some shape is a closed figure with three sides *makes* it a triangle and it also seems fine to say that the property of *is a closed figure with three sides* is the same property as *is a triangle*. Some naturalists want to say that something similar happens in the case of n and m.[192] They think that we can offer *constitutive* explanations of an act's being right, wrong, good, bad, etc. in terms of that act's natural properties. Nothing that Parfit says seem to show that that is not possible. So, even if Parfit were right about the water case then, it would not be costly for the naturalist since there are some other examples of non-causal making between identical

---

[191] I say 'boiling in the particular way that it does' because, of course, water boils when it has greater amounts of molecular kinetic energy too. Water does not however, boil in exactly the same way, (e.g. does not evaporate at the same rate, does not bubble with the same force), when it has *more* kinetic energy than water at 100 degrees does.
[192] See especially Schroeder (2007, ch.4) and Finlay (2014, ch.1). Their views are heavily influenced by Jeffrey King's (1998) work on analysis which I discussed in chapter 1.

properties that they can appeal to. Parfit has to show that there are no cases of non-causal making between identical properties, not just that there are some cases of non-causal making between distinct properties. So far he has not done that.

Finally, what of the fact that *non-causal making* is an asymmetrical relation, whereas *being the same as* is not? If normative properties are just natural properties then why does the fact that an act has a natural property make it the case that that act has a normative property, but not vice versa?

This is admittedly a puzzling issue at first blush, but there are a few different suggestions in the literature for how to understand constitutive reductive explanations. On most proposals, the asymmetry of *non-causal making* claims is explained by the fact that 'P makes Q' conveys different information to 'Q makes P'. For example, 'Having a certain molecular kinetic energy makes it have particular heat' offers us an informative scientific explanation of heat in terms of molecular kinetic energy, whereas, 'Having a certain heat makes it have a certain molecular kinetic energy', in a sense, gets things backwards as heat is not fundamental to our understanding of molecular kinetic energy.

There are then different suggestions for understanding why some non-causal making explanations are informative whereas others are not. I will run through some of these quickly without evaluating how plausible they are or whether they can truly explain the asymmetry of non-causal making explanations.

One suggestion is that the claim, 'P makes Q', informs us that one linguistic expression which we did not previously know the extension of, 'Q', picks out the same property as another linguistic expression that we did know the extension of, 'P'.[193] For example, discovering that minimising suffering makes something right informs us that the extension of the term 'right' are all those actions that minimise suffering.

---

[193] Jackson (1998, p.123)

According to another suggestion, constitutive explanations inform us that we can come to know that something has a particular property in the same kind of way that we come to know that it has another property.[194] For example, if we discover that minimising suffering makes something right, we come to know that we can learn that something is right through broadly empirical means, namely, by measuring how much suffering it causes.

A third suggestion has it that non-causal making claims help us to understand the structure of the property in question, e.g. learning that being a unmarried adult man makes you a bachelor is to learn that the property *being a bachelor* is composed of the properties *is unmarried, is an adult* and *is a man*.[195] Complex properties are composed of simpler ones but not vice versa, so the asymmetry of non-causal making is accounted for on this suggestion.

A fourth suggestion is the same as the third suggestion except that, rather than these constitutive explanations telling us about the structure of a *property,* they tell us about the structure of a *concept*.[196]

We do not need to determine which, if any, of these suggestions is correct. Everyone needs some account of non-causal making to explain why it is legitimate to give constitutive explanations like, 'Being a closed figure with three sides makes it a triangle'. There are several potential accounts in the literature already. So, the naturalist need not be worried by the fact that according to their account an action's having a certain natural property makes it have normative another, even though those properties are the same.

Conclusion

---

[194] Shafer-Landau (2003). He does not consider himself to be a naturalist, but he does argue that normative properties are *constituted* by natural properties. This makes him a naturalist as I have been using the term since on such views normative properties are not, ontologically speaking, *over and above* natural properties in a relevant way.
[195] Schroeder (2007, ch.4)
[196] Finlay (2014).

I have reviewed three ways that we might understand the normativity argument against naturalism and I have found them all wanting. That naturalism might seem incorrect to some philosophers is not a very pressing objection against naturalism. Nor is it very difficult for naturalists to explain why normative properties are normative in the reason-implying sense, or why natural properties can both non-causally make things have particular normative properties while being identical to those properties. I now move on to consider Parfit's other principal argument against naturalism.

Chapter 4: The Triviality Argument

Parfit thinks that the normativity argument decisively rules out naturalism but, acknowledging that many people will be unconvinced, he also offers other arguments. The one he believes to be most significant is the triviality argument, which he devotes the most discussion to and names a chapter after.[197] This argument has generated much discussion, but most of the replies have misunderstood the point that Parfit is trying to make. However, once we understand the point that Parfit is trying to make we can see that one of the premises of the triviality argument just begs the question against the naturalist. Parfit claims that he does not beg the question because he has offered decisive auxiliary arguments for the premises of his argument. It is these auxiliary arguments and not the triviality argument, which is of the most philosophical interest. I identify the main auxiliary arguments and show that they do not work. I also, along the way, dampen the hopes of Parfit and Peter Railton, that naturalists and non-naturalists might come to agree with one another.

§1 – The Triviality Argument

The triviality argument is designed to show that normative claims, understood along naturalist lines, could not be both *substantive* and *positive*. In order to be *substantive* a normative claim must be *significant*, in the sense that we could disagree with it or it could tell us something we do not already know. To be *positive* a normative claim must state or imply that when some act has a natural property, like *maximises happiness*, it also has some other, *different* normative property.

Throughout, Parfit uses the example of the utilitarian claim

> (A) When some act would maximise happiness, this act is what we ought to do.

---

[197] Parfit (2011b, ch.26)

On the kind of non-naturalist account that Parfit favours (A) tells us that acts which have the natural property *maximises happiness*, also have the non-natural, non-ontological, irreducibly normative property *being what we ought to do*. It is easy for philosophers who hold this account to explain why (A) is substantive and positive. (A) is substantive since we might not have already known that acts which maximise happiness also have the different normative property of being what we ought to do. Additionally, someone could sensibly disagree with (A); they might believe instead that only actions with some other natural property are the actions we ought to do or they might believe that there is no such property as being what we ought to do. Non-naturalists can straightforwardly explain why (A) is positive too since *being what we ought to do* is, on their account, a different normative property from the natural property *maximises happiness*.

Naturalists meanwhile cannot explain why claims like (A) are both substantive and positive. Naturalist utilitarians are committed to something like the following claim:

> (C) When some act would maximise happiness, this property of this act is the same as the property of being what we ought to do.

But (C) is incompatible with the fact that (A) is substantive and positive, so naturalist utilitarianism must be false, or so Parfit argues. Accepting (C) does not require accepting there is any analytic link between the meaning of the expressions 'maximises happiness' and 'is what we ought to do', or between the content of the *concepts* <maximises happiness> and <is what we ought to do>. (C) only requires that normative terms and concepts pick out the same property as some natural terms and concepts. As such, Parfit's argument applies equally to both analytic and synthetic forms of utilitarian naturalism.

Parfit presents his argument in the following form:

> (1) (A) is a substantive normative claim, which might state a positive substantive normative fact.

> (2) If, impossibly, (C) were true, (A) could not state such a fact. (A) could not be used to imply that when some act would maximise happiness, this act would have the different property of being what we ought to do, since (C) claims that there is no such different property. Understood along Naturalist lines (A) would

be only another way of stating the trivial fact that, when some act would maximise happiness, this act would maximise happiness.

Therefore,

This form of naturalism is not true.[198]

Mark Schroeder points out that in this form Parfit's argument is invalid.[199] (1) is true as long as (A) states or implies that when an act would maximise happiness it would have *some* different normative property. So, (1) could be true if (A) implies that acts which maximise happiness have the normative property *good* or the normative property *what we have most reason to do*. But, (2), if true, only rules out that (A) implies that when some act maximises happiness it could not have the normative property of *being what we ought to*. This premise is silent about whether (A) might imply that happiness-maximising acts has normative properties other than the property *being what we ought to do*. As a result, (1) and (2) could be true and the conclusion false. To fix this Parfit could simply tweak the second premise so that it said something like this:

> (2') If, impossibly, (C) were true, (A) could not state such a fact. (A) could not be used to imply that when some act would maximise happiness, this act would have *a different normative property* since (C) claims that there is no such different property.

With that tweak in place, Parfit's argument is valid, but it only tells against one particular form of naturalist utilitarianism if it is sound. Parfit thinks, however, that his argument generalises. For any claim of the form, 'If an act is M then it is N', where 'M' is a natural predicate and 'N' is a normative predicate, that claim can only be positive and substantive if it says of an act that has a natural property that is also has a different normative property. However, according to the naturalist, some claim of the form, 'If an act is M then it is N', merely states the normatively trivial fact that actions with a particular natural property have that selfsame property. If

---

[198] Parfit (2011b), pp.343-44. I alter premise (2) slightly so that Parfit's argument applies to analytic naturalists as well as synthetic naturalists.

[199] Schroeder (2016)

normative claims of the form, 'If an act is M then it is N', are positive and substantive then for any 'M' and 'N' those terms cannot denote the same property. Therefore, no naturalist account could work. It does not matter whether or not on the naturalist account in question 'M' and 'N' have the same *meaning*, just that they are co-referring. Because his argument has this feature, Parfit thinks that it is not vulnerable to the same objections that the open question argument faces.

§2 - Response 1 – Railton and Dowell & Sobel

Parfit's triviality argument has generated a lot of discussion but most responses to it have missed the mark. Parfit imposes on the naturalist two requirements: that they show that normative claims like (A) are substantive and that they show that they are positive. Most naturalist responses have focused on the first of these requirements, but, as we shall see, this is a mistake because Parfit is willing to grant that the naturalist can explain why normative claims are substantive, just not how they could be *both substantive and positive*.

(A) is *substantive* as long as it is potentially informative and could be disagreed with. The second premise of Parfit's arguments states that if naturalism were true then normative claims would just state the normatively trivial fact that actions with a natural property have that selfsame property. Most philosophers who have responded to Parfit have thought that he is here claiming that if normative claims like (A) merely stated the fact that actions with a natural property have that property then they could not inform us of something we did not already know, nor could they be sensibly disagreed with. [200] On this interpretation, Parfit's argument is more or less just a variation of Moore's open question argument, which I already covered in Chapters 1 and 2. It is natural to think that Parfit is here trying to argue along the same lines as Moore. Moore argued that if naturalism were true then a normative claim like 'Pleasure is good' would be equivalent in

---

[200] Copp (2017), Jackson (2016) and Lenman & Lutz (2018) also interpret Parfit's point in this way.

meaning to a trivial claim like 'Pleasure is pleasure'; that looks similar to the point that Parfit is trying to get at when he says that naturalists cannot explain why normative claims are non-trivial. If we understand Parfit's argument this way then the challenge for the naturalist is to explain the difference in cognitive significance between accepting normative claims and accepting trivial natural claims like 'Pleasure is pleasure'.

To meet this challenge the naturalist can just take one of the many extent responses to Moore's argument and apply it, *mutatis mutandis*, to Parfit's argument. Different naturalists try to do this in different ways. One example comes in Peter Railton's paper *Two Sides of the Metaethical Mountain*.[201] Railton takes Parfit to be challenging the naturalist to say what information a claim like

> (C) when some act would maximise happiness, this property of this act is the same as the property of being what we ought to do.

provides over and above a trivial claim like

> (E) being an act that would maximise happiness is the same as being an act that would maximise happiness

Railton's response is to argue that (C), unlike (E), informs us that the property *maximises happiness* fulfils the *job description* associated with the normative concept <being what we ought to do>. He makes his case by using the analogy to the concepts <heat> and <molecular kinetic energy>, which are co-referring but associated with different information. The job description associated with the concept <heat> is something like *that which melts solids, evaporates liquids, etc*. If we are told that when something has the property of having a certain molecular kinetic energy then it necessarily has a certain heat, we are informed that things with a natural property also have the property of fulfilling a certain job description associated with the concept <heat>. This is information that could not be provided simply by telling saying to someone, "Something with a certain molecular kinetic energy has a certain molecular kinetic energy", since the concept <molecular kinetic energy> is not associated with the same job description as the concept

---

[201] Railton (2016).

<heat>. The same goes for (C) and (E). If we learned that *maximising happiness* was the same as *being what we ought to do* we would have learned that *maximising happiness* is the property that fits the job description of <being what we ought to do>. Railton is circumspect when it comes to detailing what the job descriptions associated with normative concepts are, but he does suggest that normative concepts are associated with other normative concepts and that they are related to intention and action in ways that paradigmatically natural concepts are not.

Whether or not Railton shows that normative claims, understood along naturalist lines, can be substantive, it does not fully address the argument that Parfit intends to make. Parfit accepts that claims like (E) could give us substantive information as Railton claims, but he argues that this information could not also be *positive*. To be positive, in Parfit's sense, a normative claim must tell us that when some act has a natural property it has another *different* normative property. The claim (C) does *not* imply that when some act has a particular natural property, it therefore has a different normative property. The only information it gives us is *negative*; it tells us that when some act has a natural property, it does not have some other different property of being right, since naturalists do not believe there is any property of rightness distinct from that natural property.[202]

The same pattern arises in Janice Dowell and David Sobel's debate with Parfit. Dowell and Sobel understand Parfit's argument along the same lines that Railton does. They write:

> "Parfit's Triviality Objection purports to show that [naturalists] are unable to do so much as state informative identities between the normative and the natural".[203]

They model their response to this challenge on Robert Stalnaker's approach to explaining why identity claims involving co-referring proper names can be informative. It is not necessary to go far into the details, but the idea is that claims like (C) update the 'conversational common

---

[202] Parfit (2016, pp.94-6)
[203] Dowell and Sobel (2017, p.153)

ground' – roughly, the set of presuppositions and background beliefs underlying a conversation - in ways that (E) does not.[204] Particularly, (C) rules out that we are speaking a language in which the expression 'maximises utility' denotes a property other than the property that is denoted by the expression 'what we ought to do'.

In his response, Parfit does not take issue with Dowell and Sobel's contention that claims like (C), understood along naturalist lines, could be informative. His point of contention is that this information could not be *positive*. As a result, he quickly dismisses Dowell and Sobel's argument as being based on a misunderstanding of his argument.[205] Merely showing that normative claims are substantive is not an adequate response to Parfit's challenge since it only fulfils one of the two requirements that Parfit imposes.

§3 – Response 2 - Railton Again

Railton's first reply to Parfit, which we have just seen, is to challenge the idea that normative claims, understood along naturalist lines, cannot be informative. Railton also, however, offers a more irenic response.[206] He explores whether a synthetic naturalist could accept that there are non-natural properties if those properties are also non-ontological, as Parfit believes normative properties to be.[207] On this response, the naturalist, in effect, concedes that Parfit's Triviality argument shows that claims like (C) could not be true, but then shows that the denial of claims like (C) is consistent with naturalism.

Railton takes Parfit to be saying that distinct concepts must be tied to distinct properties. He discusses the concepts <heat> and <molecular kinetic energy>. Railton takes Parfit to be saying that because those concepts are distinct they must refer to distinct properties. A claim like

---

[204] Stalnaker (1981, 2014).
[205] Parfit (2017, pp.230-232).
[206] Railton (2016)
[207] Parfit (2016, pp.53-4)

(G) heat is the same as molecular kinetic energy

tells us that the property *molecular kinetic energy* has a different second-order property *heat*. Because the concept <heat> supplies different information to the concept <molecular kinetic energy>, the properties *heat* and *molecular kinetic energy* are different properties. Railton comments that this is an unusual way to understand talk of 'properties' - we do not usually think of *properties* as conveying information, rather, we think that only *properties under descriptions* convey information.[208] Still, if the naturalist goes along with Parfit's understanding of 'properties' they can agree with much of what he says, or so suggests Railton.

Normative concepts are distinct from natural concepts on both Railton and Parfit's views. Since normative concepts carry different information to paradigmatically natural concepts, Railton thinks he can now agree that normative concepts pick out some kind of non-ontological, non-natural property, *in Parfit's sense of the word 'property'*. Since these properties are non-ontological this does not conflict with his commitment to methodological naturalism – the principle that, in inquiry, we should see how far we can get on the assumption that all there are, are the kinds of entities that figure in the sciences. While non-natural irreducibly normative properties do not figure in the sciences, if they are non-ontological then naturalists need have no objection to including them. And since normative concepts like <wrong> do not play fundamental explanatory roles in the sciences it makes sense for the naturalist to say that *properties* like *wrong* are non-natural (as long as we are using 'property' in the sense that Railton takes Parfit to be doing).

Since on Parfit's view (as Railton understands it) to say that two properties are distinct is to say that the concepts they are associated with carry different information, we can accept that, *in Parfit's sense of 'property'*, things have non-natural properties without reifying the former in any way that would trouble a methodological naturalist. To say that, 'Killing has the non-natural property

---

[208] Railton (2016, p.53)

wrong', would be to say that the concept <wrong> applies to killing and that that concept <wrong> is not equivalent in meaning to any natural concept. The synthetic naturalist could agree with all this.

When it comes to talk of irreducibly normative facts, Railton also sees reason to think that he and Parfit can agree. Parfit distinguishes between *referential* and *informational* senses of facts.[209] On the former sense the claims, 'Heat is heat', and, 'Heat is molecular kinetic energy', state the same fact, on the latter sense one claim states a trivial fact and the other states an interesting fact. Parfit thinks that when we talk of 'normative facts', it is the informational sense of fact that is of interest to us. Railton argues that the synthetic naturalist can agree that 'Killing is wrong' and 'Killing does not maximise happiness' (or whatever) state different facts *in the informational sense of 'fact'*, since they agree that those sentences carry different information. Since the concept <wrong> is one that does not fundamentally figure in the natural sciences, Railton is happy to say that, 'Killing is wrong', expresses a kind of *non-natural* fact, distinct from the fact expressed by the claim, 'Killing does not maximise happiness'.

Additionally, on Parfit's view, only natural facts have *normative importance*.[210] The claim,

> (J) Your wine is poisoned

states a natural fact which gives you a reason not to drink the wine. The normative claim,

> (K) the fact stated by (J) gives you a reason not to drink the wine,

states a normative fact. But, that fact is not *normatively important*; it does not provide you with any further reason not to drink the wine over and above the fact stated by (J), according to Parfit. Since, on Parfit's view, non-natural normative facts do not have normative importance, Railton thinks that that is a reason to think that the naturalist can grant that talk of non-natural

---

[209] Parfit (2011b), ch.26, §94.
[210] Parfit (2011b), pp.279-80

normative facts is legitimate. Since non-natural normative facts do not need to figure in any, "fundamental metaphysical explanation of what really matters", these facts can figure into a broadly methodologically naturalist program.[211] So, Railton suggests, all in all there is very little that Parfit says about normative matters that the naturalist could not say also.

Parfit is enthusiastic about Railton's proposal.[212] He writes that he and Railton now agree. Parfit thinks that if the naturalist is happy to allow that there are non-ontological non-natural normative properties and use those to explain the normativity of normative claims then synthetic naturalism is okay.

I do not think that Parfit and Railton really agree. The reason is that Railton misunderstands Parfit's view of properties, something which Parfit actually points out.[213] Railton has it that on Parfit's view where two terms carry different information they refer to different properties. Since on Railton's view normative and natural terms carry different information, he thinks he can allow that in Parfit's sense of the term 'property', normative and natural terms refer to different properties. Just as the properties *being a certain heat* and *having a certain molecular kinetic energy* might be said to be different because the expressions 'being a certain heat' and 'having a certain molecular kinetic energy' convey different information, so might the synthetic naturalist say that the properties *maximising happiness* and *being what we ought to do* are different properties. At the same time, the naturalist could maintain that *in a causal-explanatory sense* normative properties are natural properties, i.e. that normative properties are nothing over and above natural properties ontologically speaking.

However, Parfit denies that this is his view of properties, though admittedly it is hard to identify exactly what he thinks properties are. On his view, *heat* and *molecular kinetic energy* are the same property even though the terms 'heat' and 'molecular kinetic energy' convey different

---

[211] Railton (2016, p.53).
[212] Parfit (2016, ch.42).
[213] Parfit (2016, pp.93-4)

information. In fact, earlier in *On What Matters*, in response to one of Frank Jackson's arguments, Parfit writes that whenever two concepts necessarily have the same extension then *if those concepts refer to natural properties*, like <heat> and <molecular kinetic energy>, then plausibly they must both refer to the very same property.[214]

Instead, Parfit argues that the claim

(G) 'Having a certain heat is the same as having a certain molecular kinetic energy'

is more informative than

(O) 'Having a certain heat is the same as having a certain heat'

even though they state the same fact in the referential sense, because the former also tells us that the property *having a certain molecular kinetic energy* is related to certain other properties associated with the term 'heat' like *causing certain sensations*, *being melted* and *being turned into a gas*. (G) and (O) both state the same fact in the referential sense, but because they do so in different ways (G) conveys information that (O) does not.

Since Parfit does not think, *pace* Railton, that two properties must be distinct if they (or terms that pick them out) carry different information it does not seem as clear that the naturalist can agree with Parfit after all. When Parfit says that to be substantive and positive claims like (A) must be relating one property to a different normative property he does not mean merely that claims like (A) should convey some information over and above the trivial claim that a natural property has a natural property, otherwise he would say that a claim like (G) was relating two different properties as well. When Parfit says that normative properties cannot be natural properties he seems to be using 'properties' in the same kind of sense that Railton would. Railton has to say that normative properties belong to the same ontological category as natural properties, even if normative terms are not equivalent to natural terms. Parfit says that normative

---

[214] Parfit (2011b, p.296).

properties cannot possibly belong to the same ontological category as natural properties (because normative properties belong to no ontological category on his view). There is no chance of their views being reconciled.

§4

Parfit does admit that both Railton's misunderstanding and Dowell and Sobel's, may be partly his fault because of how he stated his view.[215] I think that Parfit is right about this. In his responses to naturalists, Parfit is very quick to accept that naturalists can show that normative claims are at least substantive. Given that, one wonders why he made it one of the two explananda. Since he thinks it is relatively easy for naturalists to explain why normative claims are substantive but impossible to explain why they are *positive*, there seems no reason to challenge naturalists to explain how normative claims could be both substantive *and* positive, rather than just challenging them to show that normative claims like (A) can be positive. It is also strange that Parfit calls his argument the 'triviality argument', rather than something like the 'positivity argument', since the positivity requirement seems to be doing the work. These features of Parfit's argument invite misunderstanding.

I suspect however that the main reason that philosophers have not seemed to address the argument Parfit is intending to make is down to considerations of charity, for understood correctly Parfit's argument is straightforwardly question-begging. In assuming that a normative claim like, 'Maximising happiness is right', must be positive, Parfit begs the question against the naturalist. The naturalist thinks that some natural property N is identical to *rightness*. Parfit's assumption that all claims of the form like, 'Actions which are N are right', must be positive, requires that 'N' and 'right' denote distinct properties. It is not clear why Parfit should be entitled

---

[215] Parfit (2016, p.95)

to assume that normative claims are all positive since whether 'N' and 'right' denote the same natural property or not is exactly the point of contention between him and the naturalist.

And that, I think, is all that really needs to said to refute Parfit's triviality argument. The first premise of Parfit's argument is tantamount to declaring that naturalism is false and so is question-begging in the most obvious way. It is easy to sympathise with Mark Schroeder (who I think does gauge Parfit's intentions correctly) when he says, "I have to confess that Parfit's triviality objection is one of the most puzzling arguments I have ever encountered in philosophy."[216]

In his reply to Schroeder, Parfit claims that he does not beg the question since he provides arguments for his first premise, and since Schroeder does not address these auxiliary arguments the triviality argument is still to be defeated.[217] Since the first premise of Parfit's argument is tantamount to a denial of naturalism, any auxiliary arguments for that premise are going to be more interesting arguments against naturalism than the triviality argument itself. It is strange that the triviality argument has been billed as a new, devastating objection to naturalism given that the philosophical interest really lies in the justification for its first premise.

In any case, when we come to look at *On what Matters*, it turns out to be difficult to identify any argument for the first premise of the triviality argument which applies generally to all forms of naturalism and which is not obviously question-begging. Seemingly, the normativity argument, or similar flat appeals to the intuition that naturalism seems wrong, is supposed to be doing some of the work, but as I argued in the last chapter that argument is not convincing. In what follows I try to draw out from the text some of the other arguments that Parfit might use to establish the first premise of his triviality argument and show that they do not provide a very strong case against naturalism.

---

[216] Schroeder (2016, p.223)
[217] Parfit (2016, ch.44).

§5

§5.1 – Hard and Soft Naturalism

To deny the first premise of the triviality argument is, Parfit argues, to accept what he calls 'hard naturalism'.[218] *Hard naturalists* think that, at least in principle, normative facts could be stated in purely natural terms. Parfit gives Richard Brandt and Frank Jackson as examples of hard naturalists. Hard naturalists will say that normative claims like (A) are not positive. Since they think that a normative term like 'is right' can be restated as 'maximises utility', it is coherent for them to say that (A) does not say of actions with one property that they have another *different* normative property, rather, (A) says that actions with one natural property have that natural property. The problem with hard naturalism is that it is just very implausible that we could do without normative terms and concepts. Normative facts just could not be restated in purely natural terms without loss, or so thinks Parfit. I will say more about Parfit's case against hard naturalism shortly.

According to what Parfit calls 'soft naturalism' we must use normative terms and concepts, or that we at least have very good reason to, even though normative facts are natural facts.[219] Parfit attributes this kind of position to philosophers like Stephen Darwall and Peter Railton. Parfit thinks that soft naturalism is, *prima facie*, a much more plausible position than hard naturalism and he even goes as far to say that the central claims of soft naturalism sometimes strike him as true.[220] All the same, he argues that soft naturalism ultimately founders.

Soft naturalists cannot deny that normative claims are positive, argues Parfit. Soft naturalists have to say that normative claims like (A) are positive, for if (A) just said of actions that had a

---

[218] Parfit (2011b, p.344)
[219] On Parfit's formulation naturalists also claim that all facts are natural facts. I prefer my way of phrasing it because it allows that people who believe in theological or supernatural facts, say, could still count as ethical naturalists. Some of the major neo-Aristotelians like G.E.M. Anscombe, Peter Geach and Alastair McIntyre might fall into this category and I take it it would be preferable if their views counted as naturalist.
[220] ibid. p.349

natural property that it had that natural property, we would not need or have good reason to have normative terms and concepts. If (A) just states the trivial fact that happiness-maximising actions are happiness-maximising, then we could state the same fact with the purely natural claim (E):

> (E) being an act that would maximise happiness is the same as being an act that would maximise happiness

So, soft naturalists must say that normative claims like (A) state some fact other than the one stated by (E). The problem is that there is no obvious candidate for which fact this would be. In other words, soft naturalists must argue that normative claims like (A) tell us that acts which have some natural property have some other different normative property, but this leads them into what Parfit calls, *The Lost Property Problem*; the problem of being unable to say which normative property a claim like (A) attributes to an action.[221] The most obvious candidate for the normative property that (A) attributes to happiness-maximising actions would be the property *being what we ought to do*. But, if this is the case then (A) would not state a positive, normative fact given the naturalist's commitment to (C), which says that *being what we ought to do* is not a different property to *maximising happiness*. There are no other obvious candidates for the normative property that a claim like (A) could attribute other than *being what we ought to do*. That being so it seems likely that if (C) were true then (A) must state the normatively trivial fact that acts with a natural property have that natural property but provide no positive normative information in addition to this.

Additionally, even if soft naturalists do identify a different normative property that is attributed to happiness-maximising actions by a claim like (A) (e.g. the normative property *being what I have most reason to to*) they will have to give a naturalistic account of that property, i.e. they will need to say which natural property it is identical to. To then explain why claims which attribute *that* property to actions are positive they would have to find yet another normative property which

---

[221] ibid. p.345

those claims attribute to those actions. And then they would have to do the same thing for this other property, and so on for ever.

Parfit goes further and argues that soft naturalism is not even a coherent position. If, on the one hand, soft naturalists were right that we needed normative terms and concepts then it would be false to say that normative facts were also natural, for if they were they could be stated using purely naturalistic vocabulary. If, on the other hand, soft naturalists say that normative facts could be stated in natural terms then soft naturalism simply fails in its own terms and collapses into hard naturalism. This is the *Soft Naturalist's Dilemma*.[222]

So, if naturalists argue that we do need normative terms and concepts then their position ceases to be coherent, but if they argue that we do not need normative terms and concepts then their position cannot explain why normative claims like (A) are *positive*. In what follows I argue first that soft naturalism is coherent and that soft naturalists need not say that normative claims are positive. After that, I argue that Parfit has failed to show that hard naturalism is implausible.

§5.2 - Parfit's Case against Soft Naturalism

Parfit argues that soft naturalists cannot claim that we need, or have good reason to have, irreducibly normative terms and concepts if normative properties are natural properties. By 'irreducibly normative terms and concepts' he must mean normative terms and concepts which are distinct in meaning from any natural terms and concepts which refer to the same properties, just as 'H_2O' and 'water' are distinct in meaning but co-referring. If normative properties are natural properties then we could just as well pick out normative properties with natural terms as with normative terms. So, we do not need, or have good reason to have, those normative terms. The natural terms would do by themselves. This means that soft naturalism collapses into hard naturalism.

---

[222] ibid. p.366.

Parfit's argument involves two questionable assumptions. The first is the assumption that if normative properties were natural properties then we must be able refer to those properties using some non-normative, natural terms. Nicholas Sturgeon denies this assumption.[223] Sturgeon sees no reason to suppose that the natural sciences must have a term for every natural property. Just for the purposes of illustration, suppose a kind of particularism is correct, meaning that whether an action is, say, good depends on many particular features of a situation. On this view, what makes an action *pro tanto* good in one situation might make it *pro tanto* bad in other situations. For example, that going to the cinema causes pleasure might be *pro tanto* good but that sadistic cruelty causes pleasure might be *pro tanto* bad. If some such view were correct then there might be no simple natural predicate, like 'pleasurable', that we could use refer to all good actions. Nevertheless, we could think that an action's *goodness* in any given situation were identical to those natural properties that (constitutively) make it good. Parfit says nothing that refutes Sturgeon's view.[224]

We might think, as Frank Jackson does, that there must be a way to refer to normative properties using purely natural terminology.[225] I covered this argument in chapter 2, but here a brief sketch of it by way of reminder.[226] For any possible right action we can give a complete description, w_n of that action's natural properties and the properties of the world in which it takes place. Because the normative supervenes on the natural we know that actions which are naturalistically identical to the action described by w_n will also be right. If we put all these descriptions into a huge disjunction of the form $w_1, w_{2...} w_n$ then we will have a description, containing only natural terms, which can be applied to any right action. This huge disjunction then is a purely natural

---

[223] Sturgeon (1988. 2005, p.109).
[224] Parfit addresses Sturgeon directly but misunderstands his position. Parfit (2011b, p.365, p.368) quotes Sturgeon as saying that if some form of naturalism is true then we would 'be able to say, in entirely non-moral terms, exactly which natural properties moral terms refer to', and 'moral explanations would be in principle dispensable'. But in the section Parfit is quoting from Sturgeon (1988, pp.242-243) is only talking about what would the case if *reductive* naturalism were true; a form of naturalism that he rejects. In fact in the very same passage Parfit quotes from, Sturgeon makes it clear that the position Parfit attributes to him is one he in fact rejects.
[225] Though Jackson would prefer to say 'descriptive terminology' for reasons I do not need to go into.
[226] See chapter 2 §6.

expressions with which we can denote right actions. This might seem to refute what Sturgeon says. If goodness is a natural property then even if there is no common natural term for it in our language, we can construct one in the above way. This might seem to show that Parfit is right to say that if normative properties are natural properties then they can be referred to using purely non-normative, natural terminology.

However, if normative properties could only be picked out by extremely long disjunctions, or even infinitely long disjunctions, then we would have excellent reason to have normative terms in addition to natural terms. Very long disjunctive expressions are unwieldy and difficult to follow. Infinitely long disjunctions are impossible to state or fully understand. So if Jackson is right that we can in principle state all possible normative properties in natural terms, that would not show that Parfit was right that soft naturalism is incoherent. The soft naturalist can say that we have excellent reason to retain normative terms and concepts because we could not refer, in any practicable way, to normative properties without them.

Following on from that, the second questionable assumption Parfit makes is that we cannot need, or have good reasons to have, two terms or concepts which are distinct in meaning if they are co-referring. This seems like an odd assumption, even if Sturgeon is wrong and normative terms like 'good' are coreferring with very simple natural terms like 'causes pleasure'. No one believes that we do not need, or do not have good reason to have, the term 'water' since we also have the co-referring term 'being composed of H_2O molecules'. There are many possible reasons we might have to have two such terms. I will outline a few of these.[227]

First, different coreferring terms might be associated with different conceptual or inferential roles.[228] Take the case of 'water' and 'composed of H_2O molecules'. We can know that one term

---

[227] See also Copp (2012, 2017) who does something similar in response to Parfit.
[228] It need not be the case that an inferential/conceptual role *semantics* is correct for what I say to work. Terms/concepts may be associated with inferential/conceptual roles without the *meaning* of these terms and concepts being determined by those inferential/conceptual roles. Railton suggests something along these lines for normative concepts (2016).

applies to some liquid without knowing that the other does, even though these expressions are co-referring. Suppose for illustrative purposes that competence with the term 'water' requires an ability to infer that the liquids to which the term applies are potable and that competence with the expression 'is composed of H_2O molecules' requires an ability to infer that the liquids to which the term applies contain hydrogen atoms. If someone knew that 'is composed of H_2O molecules' applies to the liquid in the glass we can infer straightaway that there are some hydrogen atoms in the glass but we cannot necessarily infer that the liquid is potable. At the same time, someone else could be competent with the term 'water' and know that it applies to the liquid in the glass without thereby being able to infer that there are hydrogen atoms in the glass, but they would be able to infer that it is potable. This gives us a reason to have both 'is composed of H_2O molecules' and 'water' in our vocabulary. By applying the latter term we imply something about the liquid in the glass which would not necessarily imply if we used the term 'water' instead and vice versa.

Similarly, normative and natural terms might be associated with different inferential or conceptual roles. A particularly clear example of this comes in Allan Gibbard's *Thinking How to Live*.[229] Gibbard argues that normative properties are natural properties but that normative *concepts* are 'plan-laden', meaning that they are used to rule out particular courses of action. On this picture, we can infer from someone's using a normative term something about their intentions or motivations which we would not be able to infer if they used instead a coreferring natural term.

A naturalist need not accept the kind of semantics that Gibbard carves out for normative terms in order to justify the usefulness of normative terms however. There are other options available. Supposing that *rightness* is identical to m*aximising happiness*, it might be the case that we still need the concept <rightness> because it has inferential links to other concepts like <ought> and

---

[229] Gibbard (2003)

<wrong> which the concept <maximising happiness> does not.[230] In other words, by telling you 'X is right', you can infer that X is not wrong and that you ought to do X, but you cannot necessarily infer those things by my saying 'X maximises happiness'.

The second reason we might want to have distinct, co-referring expressions is that people can be unaware that co-referring terms are co-referring.[231] It is useful to have the term 'heat' in addition to the term 'molecular kinetic energy' so that we can communicate heat/molecular kinetic energy related information to people who know what 'heat' refers to but don't know what 'molecular kinetic energy' refers to. Similarly, if, as soft naturalists believe, we are not sure exactly which natural terms picks out normative properties, it is helpful to retain normative terms which do.

Third, people can disagree about whether co-referring terms are co-referring. Suppose that I think that water is a primitive element whereas you know that it is composed of H_2O molecules. Since 'water' and 'H_2O' are coreferring you could state, in a referential sense, your view that water is H_2O even if our language did not have the term 'water'; you could just say 'H_2O is H_2O'. However, this would make it impossible for us to disagree. I will not object to your claim 'H_2O is H_2O'. The term 'water', it is sometimes said, is associated with the *job description*, which is something like 'that clear potable liquid that fills our lakes and rivers'. Having a term with that job description allows us to ground our disagreement since we disagree about the fundamental chemical composition of the liquid in our lakes and rivers. Similarly, some naturalists argue, normative terms are associated with a different job description to the natural terms that refer to the same properties.

Fourth, non-equivalent but coreferring terms can also be associated with different implicature, presuppositions or other categories of information that are the result of the pragmatics of our language. According to one view of 'thick' normative terms, a term like 'generous' has the same

---

[230] Railton (2016).
[231] Sturgeon (1988).

semantic content as a purely descriptive term like 'Disposed to give large amounts to others', but it also has pragmatically expresses the speaker's approval of that character trait. Some naturalists have argued that something similar happens with even 'thin' normative terms like 'right'. These terms refer to the same property as some natural term but also express the speaker's *desires*, *pro-attitudes, approval, etc.* Having terms that work in this way serves an important social function since we can influence the each other's behaviour and attitudes by expressing our own attitudes in this way. The soft naturalist can argue that this gives us good reason to have normative terms, whose meaning is not reducible to natural terms, even though normative terms refer to natural properties.

I think we could probably add more reasons but those four will be sufficient. Given how plausible it is that we sometimes need, or have good reason to have, non-equivalent but coreferring terms I am inclined to think that Parfit must intend his argument against the soft naturalist to work a bit differently. I suspect he means to say that if, as the soft naturalist supposes, we need, or have good reason to have, irreducible normative terms and concepts *in order to represent all the things that there are*, then it cannot be the case that normative properties are natural properties. If we could just as well pick out normative properties with natural terms then, *with regards to the goal of picking out all the properties that there are*, we could use just natural terms instead.

However, that, *with regards to the goal of picking out all the properties that there are*, we need only have natural terms is something I think that the naturalist can and should accept. If we have natural terms to pick out all natural properties, and if normative properties are natural properties, then we must have natural terms to pick out normative properties. If *rightness* is *maximising utility* then we can refer to *rightness* either with the terms 'right' or 'maximises utility'. We might still have reasons for keeping these two separate terms, as I just argued, but we do not need both terms *in order to represent all the things that there are*. Having just one or the other term would do with regards

to the goal of referring to the properties that there are. As far as I can tell, no soft naturalist has any reason to deny this, so their position is not incoherent. Nor does Parfit give us any reason for thinking that it is costly for the naturalist to say that we could refer to normative properties with natural terms.

§5.3 – Parfit's Case against Hard Naturalism

Suppose now that Parfit is right that naturalists must say that, at least in principle, we could do without normative terms and concepts. Would that show that naturalism is inadequate? In my view, Parfit does not do much to establish this. In his response to Schroeder, he speaks as if he had provided a damning refutation that applies to all hard naturalist theories, but when we come to look at the text his attack on hard naturalism is almost entirely focused on attacking Richard Brandt's account.[232] Particularly, Parfit attacks Brandt's idea that we could understand calling someone's desire 'rational' to just mean that that person would still have that desire after they have finished undergoing cognitive psychotherapy – a process that requires carefully deliberating on full information.[233] Parfit argues that this is not our ordinary sense of 'rational' since, in the ordinary sense, it would not be rational for someone suffering from anorexia to starve themselves to death, even if they would desire to do so after cognitive psychotherapy.[234] Brandt is aware of this but he believes that his definition of 'rational' is preferable to the ordinary one because the ordinary definition makes no sense unless we are willing to allow that there are non-natural facts; a possibility that Brandt rules out.[235] Parfit believes that Brandt's definition is not superior to the ordinary one because, firstly, he thinks that it is plausible that there are non-natural facts and, second, if we understood normative concepts in Brandt's sense then we would

---

[232] Parfit (2011b, part 6).
[233] Brandt (1979).
[234] Parfit (2011b, §98).
[235] Or at least, this is what Parfit takes Brandt to be arguing.

have no way of asserting that in cases like that of the anorexia sufferer the rational desire is not the one that would survive cognitive psychotherapy.

Whatever the merits of Parfit's case against Brandt it is difficult to see why it should amount to a general attack against hard naturalism. If Parfit is right then Brandt's definition of 'rational' is not extensionally adequate; it does not mark as 'rational' the things we are intuitively inclined to call rational. However, hard naturalists could assign to 'rational' a different extension to the one that Brandt assigns to 'rational', so Parfit's case against Brandt will not generalise.

In fact, we know that there is at least one possible hard naturalist position that Parfit cannot think extensionally inadequate. Parfit thinks that there is some natural property N which (non-causally) makes desires rational. A natural result of this, is that that N will be necessarily coextensive with the normative property *rational*. Parfit has to concede that a hard naturalist who said that N just was identical to 'rational' would at the very least not be vulnerable to the objection he makes to Brandt. In other words, Parfit will not be able to find a case where some action is rational but it does not have the property N that the hard naturalist identifies with the normative property *rational*. So, his criticism of Brandt cannot generalise.

Beyond his argument against Brandt, Parfit does not offer any argument explicitly designed to show that hard naturalism fails. In lieu of such an argument, we are forced to scramble around to try to identify another auxiliary argument for the first premise of Parfit's triviality argument in part 6 of *On What Matters*. I run through three of these now.

§5.4 - Naturalists must be Subjectivists?

Parfit suggests at one point that naturalists will inevitably be subjectivists.[236] If that is the case then it might be right to say that naturalism will be extensionally inadequate, since extensional

---

[236] This was much more explicit in the 2009 draft of *On What Matters* (pp.562-563), but see Parfit (2011b, p.339) where he suggests that defeating subjectivism is "enough" to defeat naturalism.

inadequacy is the main objection made against subjectivist accounts.[237] If that objection is a good one and if naturalists must be subjectivists then Parfit will have a good objection against naturalists.

However, it is not clear why Parfit thinks that naturalists must be subjectivists. Part of it seems to be that he thinks that subjectivism offers what appears, at first blush, to be a promising explanation of normativity in terms of motivation. On subjectivist accounts, what is rational is tied, in one way or another, to our desires. Subjectivists might try to use this fact to explain why normative facts, on their account, are normative. Of course, Parfit thinks that this is not in fact an adequate explanation of normativity, but he seems to believe that this will be the motivation for naturalists to be subjectivists.

However, as we saw in the last chapter, there is no reason that naturalists need to explain normativity in terms of motivation. If to be normative is to be *reason-involving*, as Parfit thinks, then naturalists just need to give a natural reduction of the concept <reasons>, and show that other normative concepts are related to the concept of <reasons>. There is no reason why such a reduction must take the form of a subjectivist theory, so there is no reason that the naturalist must have an extensionally inadequate theory.

§5.5 - The Nose Blowing Argument

At one point, Parfit argues that we can see by analogy that a claim like

> (A) When some act would maximise happiness, this act is what we ought to do.

could not state the same fact as

> (J) Maximising happiness is the same property as being what we ought to do.

The analogy is with the claim

---

[237] See Parfit's own case against subjectivism (2011a, ch.3).

(V) Blowing your nose is what you ought to do.

This claim could not plausibly state the same fact as:

(W) The property of blowing your nose is the same as the property of being what you ought to do.[238]

This does not strike me as a very strong argument. I grant that (V) is not naturally interpreted as stating the same fact as (W). The 'is' in (V) is presumably the 'is' of predication rather than the 'is' of identity, since no one could realistically believe that something is what you ought to do iff it involves nose-blowing. Parfit is right that (V) is most plausibly interpreted as saying that actions with one property have some other distinct property, but the naturalist can agree with this. They do not have to say that every normative claim of the form 'X is what you ought to do' is equivalent to a claim like (W). Consider these claims

(G) Making your mother proud is what you ought to do.

(G') When some act makes your mother proud, this act is what you ought to do.

A utilitarian naturalist could accept (G) and (G') without committing themselves to saying that the property *makes your mother proud* is identical to the property *being what you ought to do*. They might just think that acts which involve making your mother proud also maximise happiness, and so also think that acts which have the property *making your mother proud* also have the property *maximises happiness/is what we ought to do*.

§5.6 - Advice Argument

Another point Parfit makes is that if normative claims just expressed natural facts then they could not be used to give advice.[239] This, I think, is a better argument. If a normative claim like (A) just states the same fact as a natural claim like

(E) being an act that would maximise happiness is the same as being an act that would maximise happiness

---

[238] Parfit (2011b, p.350)
[239] Parfit (2011b, p.278-282)

then we might doubt that (A) could be used to give advice. (E) does not seem to be a piece of advice, it seems to be silent on normative matters. So, if (A) states the same fact as (E) then (A) cannot be advice either. But, intuitively (A) is advice, so the naturalist account of normative claims like (A) must be inadequate.

To make this argument work we need to know what, according to Parfit, makes a claim a piece of advice. This is difficult to do very precisely. At one point in the text Parfit seem to characterise advice just as that which necessarily involves telling us what we *should do*, how we *ought to act*, what we have *most reason* to do, etc. But, if we take this approach then the naturalist can explain why (E) is not advice but (A) is very easily. (A) is advice, on the naturalist's account, because (A) contains the normative predicate "ought to" whereas (E) contains no normative predicate. If advice just involves the application of normative predicates then naturalists have no difficulty with explaining why (A) is advice, they just have a different take on the nature of the referent of those normative predicates to the non-naturalist.

At other times, Parfit suggests that in order to be advice a claim must tell us, or must imply, that something has an *irreducibly* normative property. To be advice a claim must tell us what we ought to do or have reason to do, where telling us what we ought to do is understood as referring to an *irreducibly* normative fact. But why should we accept this? Parfit thinks that it is just obvious, but as explored previously, the fact that some philosophers think that non-naturalism is obviously true does not tell against naturalism. Here we are just back to the question-begging assertions that normative predicates simply could not refer to natural properties.

Another option might be to understand advice in terms of its *cognitive significance.* The thought might be that normative claims, understood along naturalist lines, could not constitute advice since we might advice someone that something was, say, what they had most reason to do, without also believing that it, say, maximised happiness, and we might believe that an action maximises happiness without also thinking that it is an advisable thing to do. But the advice

argument, understood in this way, comes out as a variation of Moore's open question argument and so it could be dealt with using the strategies I defended in chapters 1 and 2.

We might say that in order to be a piece of advice a claim must necessarily exerts a kind of pressure on the audience to behave in a certain way. In other words, advice necessarily has a kind of *prescriptive* or *recommendatory* illocutionary force. We might also claim, relatedly, that if I advise someone to do something then necessarily I will *desire* that they do that thing. These claims about advice might seem difficult to capture this on a naturalistic picture. According to naturalists, normative claims just attribute to actions particular natural properties. Seemingly, someone can tell someone that something has a natural property without *prescribing* or *recommending* any particular course of action. Also, someone could accept that some action has a particular natural property without thereby desiring anything in particular. This is more or less a version of the motivational internalist's argument against naturalism.

Motivational internalism is a problem for all cognitivists, including non-naturalists like Parfit. Since Parfit is not a motivational internalist I think we can safely assume that Parfit does not think that in order to be advice an expression must be a *prescriptive* speech act, nor that it must have a necessary link to the speaker's desires. Parfit believes that people can make normative claims without necessarily desiring to influence the actions of others, meaning that on his view a prescriptive illocutionary force does not necessarily accompany sincere normative utterances. If the naturalist cannot explain why normative claims are advice, *in this sense*, then neither can Parfit, meaning that his advice argument would not have any dialectical purchase. I have more to say about motivational internalism in the next chapter. Since it is just as much a problem for Parfit as for the naturalist I put it aside for now.

Even if we put motivational internalism to one side, we might think that the naturalist needs to say something about why normative utterances are *at least often understood* as attempts to influence the behaviour of others, even though they merely tell us that certain actions have certain natural

properties. It is worth mentioning that naturalists have several options for how to do this. Philosophers such as Allan Gibbard and David Copp argue that normative properties are natural properties but both, though in quite different ways, argue that crucial to the meaning of normative claims is the expression of a kind of desire that the audience act a particular way.[240] The expression of this attitude exerts a kind of conversational pressure on the audience to obey.

Other philosophers also think that normative terms express desire-like attitudes but they argue that this is because of the *pragmatics* of normative language rather than down to the meaning of normative terms.[241] I will just give the briefest sketch. We give people information for some particular purpose. If I tell someone that smoking shortens their life expectancy I am not understood as merely informing them of some medical fact, I am understood as advising them not to smoke. The naturalist can offer explanations of this fact. Given that people are assumed to speak for some purpose and given that most people desire to live a long and healthy life, I will also be understood as recommending that they do not smoke. Similarly, argue some naturalists, if I say that some action is rational I will be understood as attempting to influence behaviour even though their semantic content is just the proposition that some action has some natural property. It is assumed that in making a normative claim I speak for some reason, and it is assumed that I want people to act in ways that fulfil their goals and desires, so my utterance exerts a kind of conversational pressure on my audience to perform that action.

These ways of explaining why normative utterances constitute advice may or may not stand up to scrutiny. All we need to note here is that since Parfit cannot offer any better explanation he cannot use this issue to establish that naturalism fails.

Conclusion

---

[240] Gibbard (2003), Copp (2001)
[241] Finlay (2014).

Parfit's triviality argument has attracted a lot of attention, but for the most part philosophers have misunderstood his intention. They have overlooked the requirement Parfit places on the naturalist to explain why normative claims are positive. Naturalists cannot meet that requirement, but this does not matter because Parfit has not established that normative claims are positive.

<u>Chapter 5: The Argument from Motivational Internalism</u>

As mentioned at the end of the last chapter, motivational internalism (from hereon I will drop the word 'motivational') presents a potential problem for any cognitivist theory of normative judgement, including naturalist theories and provides support for non-cognitivist theories. Cognitivists usually answer this challenge by denying that internalism is true and that is just what I will do in this chapter. First, I tackle some preliminaries. Then I give a sketch of the existing debate around internalism, showing why it has reached an impasse. Since the debate has reached an impasse, there cannot be drawn from it a decisive argument against naturalism. I then try to break the deadlock by arguing against internalism in an oblique way.

<u>§1</u>

Internalism is the view that there exists a necessary connection between the normative judgements people make and the actions they are motivated to perform.[242] For instance, according to one form of internalism, anyone who sincerely judges that some action is right is necessarily motivated to perform that action, and anyone who judges that some action is wrong is necessarily motivated to refrain from performing that action.

According to cognitivist theories, normative judgements are just beliefs. Judging that some action is right, for instance, is, on cognitivist views, just to believe that that action has the normative property *is right*. It is difficult for cognitivist views to accommodate internalism. Beliefs, it is widely thought, cannot by themselves motivate us to act. It is often thought that beliefs have a mind-to-world *direction of fit*, meaning that the aim of belief is to have its content match that of the world.[243] To desire to do something is, meanwhile, to be motivated to do it.[244] Desires have a world-to-mind direction of fit; their aim is to be realised, or to make the world fit

---

[242] It is important to distinguish this view from Williams-style reasons internalism which I talked about in the last two chapters.

[243] Hume (1739), Anscombe (1957), Platts (1979, p.257), Smith (1987).

[244] Smith (1988), Dancy (2000, ch.2), Toppinen (2015).

the way that the desire would have it be. Both beliefs and desires are required to motivate us to act; without desires we have no aim for how the world is to be, and without beliefs we cannot know whether the world needs to change in order to match that aim. If this picture is right, then cognitivists cannot claim that normative judgements can by themselves motivate us to act.

That does not quite show that cognitivists cannot accommodate internalism; there remains the possibility that anyone who believes that some action has the property *is right* necessarily also has a desire to perform that action. However, it is also widely thought that there are no necessary connections between distinct existences. If that is correct then the cognitivist cannot account for a necessary connection between rightness-judgements and motivation, since on their view the judgement that some action is right and a desire to perform that action are distinct mental states. If two entities are distinct then it should be possible for us to imagine one existing without the other. Therefore, if cognitivism is correct it should be possible to imagine someone who has the belief that some action has the property *is right* but does not have any desire to perform that action. However, such scenarios are, according to internalists, conceptually impossible. So, the truth of internalism would present a serious problem for any cognitivist view of normative judgement. It would also provide strong support for non-cognitivist views, which hold that normative judgements just are desires. Non-cognitivists can explain why normative judgements are necessarily accompanied by a motivation to act because on their view that normative judgements just are motivational states.

Some philosophers have argued that cognitivism can be shown to be compatible with internalism. Ralph Wedgwood, for instance, has argued that it is part of the *inferential role* of normative concepts, that we must intend to perform a particular action if we are to count as judging that that action what we ought to do.[245] Other philosophers have argued that the claim

---

[245] (2007) ch.4. Wedgwood is not a naturalist, but I do not see why naturalists could not adapt his view. Jackson (1998, pp.160-1) briefly makes a suggestion which is along these lines. Laskowski (2019) explores a view which is similar to this.

that someone is motivated to do something is *entailed* or *conventionally implicated* (in Grice's sense[246]) by normative judgements.[247] According to all these views, though normative judgements are just beliefs that some action or object has some natural property, in order to understand them specifically as *normative* judgements it must be the case that we are disposed to infer that the person who makes the judgement has a certain motivation. Interesting though these views are, I will not go into them here.

Since I am defending naturalism and not cognitivism *per se*, I could argue that naturalism is compatible with internalism even if cognitivism is not. Internalism is a conceptual claim about normative *judgements* whereas naturalism, as I have understood it, is the *metaphysical* claim that normative properties are nothing over and above natural properties. That being so, the truth of internalism would not immediately discount naturalism. At the same time, it is very natural to think that naturalism and cognitivism go hand-in-hand. If some action's being right consists in its having a particular natural property, it is hard to see how the *judgement* that some action is right could not simply be the belief that that action has that property.

Philosophers have recently challenged this idea in a number of interesting ways. For instance, some have argued that while normative properties are natural properties, normative judgements consist of both a belief that something has a natural property and a desire to act in a particular way.[248] Allan Gibbard has even argued that naturalism is compatible with the view that normative claims just express desires.[249] There is a lot to say about these views but I will put them to one

---

[246] In *Studies in the Way of Words*, Grice stipulates that by 'S implicates X' he means that X is an implicature of S.

[247] Tresan (2006) is an example of entailment view while Copp (2001) is an example of the conventional implicature view. I do not mention *conversational implicature* views (e.g. Finlay 2005) because conversational implicature is *cancellable* and so difficult to square with there being a *necessary* connection between judgement and motivation. See also, Radcliffe (2006).

[248] e.g. Boisvert (2008), Hay (2014). Schroeder (2009) offers an excellent, critical survey of many of these views. So-called 'Besirest' theories (e.g. McNaughton (1988), Wiggins (1991), Zangwill (2008)) would also fall under this heading since they think that the mental states that constitute normative judgements have a world-to-mind and a mind-to-world direction of fit.

[249] Gibbard (2003), pp.29-33. He prefers to say that normative judgements are *plans* rather than 'desires', but since I understand 'desires' in terms of their direction-of-fit, 'plans' would count as desires as I understand them.

side also and assume that naturalists are cognitivists. Instead, I will argue that internalism is false. First, I need to say a bit more about internalism.

<u>§2 - Internalism</u>

Saying that someone is motivated to perform an action is ambiguous. It could mean that that person would perform that action if the opportunity arose, or it could merely mean that they would feel some pull to perform the action but not necessarily enough motivation to actually perform it. According to, what David Brink, calls "strong internalism", anyone who makes a normative judgement will necessarily act in accordance with that judgement (or at least try to).[250] For example, the strong internalist would think that someone who sincerely judges that visiting the dentist today would be the right thing to do will necessarily at least try to visit the dentist today.

Strong internalism is an implausible theory and very few people now defend it. The idea that we sometimes fail to attempt to do that which we judge to be the right thing to do is not only coherent, but also familiar. For example, it is not uncommon for a student to believe that starting working on their essay immediately would be the right thing to do, only for them to do something fun instead. The strong internalist might argue that the fact that the student does not start work on their essay shows that *really* what they think the right thing to do is to have fun. However, this seems more like an *ad hoc* attempt to save strong internalism than it is a plausible claim about our psychology. It is hard to think of any justification for such a claim that does not assume the truth of strong internalism. According to strong internalism, it is impossible for us to fail to attempt to do what we think would be right for us to do due to lethargy, selfishness, cowardice, etc. Since the theory rules out the possibility of such commonplace, widely recognised

---

[250] Brink (1988, ch.3)

phenomena, we would need some very strong reasons to accept it. No such reasons have been found, so in what follows I ignore the possibility that strong internalism might be true.

Nowadays, internalists almost all opt for weak forms of internalism. According to weak internalism, anyone who makes a sincere normative judgement will necessarily have at least some motivation to act as that judgement requires. For instance, a weak internalist might claim that anyone who judges that some action would be the right thing for them to do will necessarily be motivated, at least to some degree, to perform that action. The motivation to do what they judge to be right could be outweighed by a motivation to perform a different action. It is consistent with weak internalism that I might fail to give £10 to Oxfam, even though I judge that it would be right to give £10 to Oxfam, if my motivation not to give £10 to Oxfam is stronger. Weak internalism merely requires that in that scenario I would have at least a tiny amount of motivation to give £10 to Oxfam. From here on when I use the label, 'internalism', I will usually be talking about weak forms of internalism unless otherwise stated.

Internalist theories can also be grouped into *unconditional* and *conditional* varieties.[251] So far I have been talking as if all internalist theories were unconditional. According to unconditional internalism anyone, without exception, who makes a normative judgement must have a corresponding motivation. Not all forms of internalism are like this. Conditional internalists argue that someone who makes a normative judgement will necessarily have a corresponding motivation *as long as they meet condition C.* Meeting C usually requires that the person making the normative judgement is free from the kinds of quirks or defects which are thought to prevent people from doing what they judge to be right. For example, meeting C might require that the speaker is not a psychopath.

---

[251] See the introduction to Björnsson, Strandberg, Olinder, Eriksson, & Björklund (2014).

Since, according to conditional views, there is still a necessary connection between judgement and motivation, the cognitivist has as much trouble accommodating these forms of internalism as they do unconditional internalism. On the cognitivist view normative judgements and the mental states which motivate us to act on those judgements are distinct so, regardless of which conditions the person making the normative judgement meets, it should be possible to imagine that person having one mental state without having the other.

The main exceptions to this are conditional forms of internalism which require that the person meets certain *normative* conditions. For example, some philosophers argue that anyone who makes a normative judgement necessarily has a corresponding motivation *as long as they are fully rational*.[252] The cognitivist can accommodate this form of internalism because they can plausibly argue that the normative property *is fully rational* is had only by people who would be motivated to act on their normative judgements. This would not be an *ad hoc* move; that someone is fully rational only if they are motivated to act on their normative judgements is an independently plausible claim. It does seem to plausibly be a defect of rationality not to be motivated to do what you judge to be right. That being so, when the naturalist comes to identify which natural property is identical to *is fully rational*, they could restrict their search only to natural properties had by agents who are motivated to do what they judge is right and in so doing capture this kind of internalism. It has often been said, rightly in my view, that introducing the requirement that the person making the normative judgement be fully rational has the effect of making internalism trivial since most metaethical views are able to accommodate it. The same is true, *mutatis mutandis*, for conditional forms of internalism which require that the person making the normative judgements is *fully virtuous* or *maximally good*. Since these conditional forms of internalism are not much threat to cognitivism I do not say any more about them.

---

[252] e.g. Smith (1994), Korsgaard (1996), Wedgwood (2007).

I give some reasons for thinking that we should reject the forms of conditional internalism that are incompatible with cognitivism later in this chapter. First, however I briefly discuss the debate around unconditional internalism. I argue that the main arguments that have been given for or against unconditional internalism are not decisive, but I then go on to provide what I think is a better argument against it. To set up these argument we need to know a bit more about the main alternative theory of normative motivation.

§3 – Externalism

Cognitivists are very often *externalists*. They believe that our motivation to do what our normative judgements require is supplied *externally* to the judgements themselves. On externalist views we might be motivated by a motivational state such as a desire (either *de dicto* or *de re*) to do the right thing, rather than by the normative judgement itself.[253] Unconditional internalists feel that if, for example, someone says that they believe that it would be right for them to do something but they have no motivation whatsoever to do it, then they do not *really* believe that that is the right thing to do. Externalists can accept that this is *usually* the case. Failing to be motivated to do what you say you think is right might be a reliable indicator that the speaker is being insincere. There are a few ways that the externalist can try to explain this fact.

First, they might argue that normative properties are natural properties which usually motivate us to perform actions that we believe exemplify them.[254] A utilitarian could, for instance, argue that the reason we are usually motivated to perform actions we judge to be right is that we are usually motivated to perform actions which would maximise happiness.

Second, they might argue that the content of normative claims makes essential reference to facts about our motivation and desires. For instance, according to one view judging that something is right involves judging that we would desire to desire that it were done if we thought about the

---

[253] See e.g. Svavardóttier (1999)
[254] ibid.

issue as fully as we could.[255] This view is not quite compatible with internalism, since it is conceivable that someone might not desire to do what they believe they would desire to desire to do if they considered the matter fully. But this view could explain why we *usually* are motivated to do what we judge to be right.

Third, there are views which try to explain why normative judgements are usually accompanied by a particular motivation by drawing on the pragmatics of normative language.[256] According to these views normative judgements presupposes or implicates a connection between that judgement and the speaker's motivation. If, for instance, we are happy to accept that people tend to speak in a way that fulfils the function of the conversation, and if we assume that normative conversation function to settle questions of how we are to act, then it will not be surprising if we can usually correctly infer from a speaker's normative utterances that they have particular motivations.

Externalists explain why normative judgements are usually accompanied by a motivation in these kinds of ways. Internalists think that this is insufficient for having a satisfactory account of normative motivation. They believe that it is *necessarily* the case that anyone who makes a normative judgement has a corresponding motivation. Externalists demur. They think that it is at the very least conceivable that someone might not be motivated to do what they judge to be right. I now move on to give a brief overview of how the debate between them has gone.

§4 – The Internalist-Externalist Stalemate

§4.1 – The Externalist Case against Internalism

The debate between internalists and externalists has largely been focused on the possibility of *amoralists* – people who have no motivation to do what they judge to be right. Externalists think

---

[255] Lewis (1989)
[256] Finlay (2005, 2014).

that there are clearly are some amoralists and that this shows that the connection between normative judgements and motivation is not a necessary connection. [257]

Psychopaths are very often cited as examples of amoralists. Psychopaths are people with sociopathic personality disorders who are prone to violence and who seemingly feel unmoved by moral considerations. Nevertheless, it has been found that, when questioned, psychopaths will say that they agree that violent and harmful behaviours are morally wrong; they just do not care that these behaviours are morally wrong. Some externalists consider this to be good evidence that people can make normative judgements while having no motivation to act on them.

Internalists have some good replies to this objection. One reply is to question whether psychopaths really make moral judgements, or if they just use terms like 'right' in what R.M. Hare called a "conventional sense".[258] The idea is that when psychopaths say that something is 'right', they mean merely that society generally *deems* it to be right. These kinds of judgements are not moral judgements in any interesting sense so the fact that they do not entail that the speaker has any particular motivation is not an objection to internalism. This seems to be a plausible account of what psychopaths mean when they make moral judgements. Studies have shown that psychopaths do not generally see as much of a distinction between what is conventional and what is moral as regular adults.[259] Psychopaths will, for instance, usually judge that moral norms are contingent on their society's conventions. Furthermore, they view moral transgressions and transgressions of conventions as being very close in terms of seriousness. A good explanation of these data is that psychopaths simply mean to describe what is conventional when they use moral terms. With a viable alternative account on the table, internalism is at no disadvantage as regards explaining the lack of moral motivation in psychopaths.

---

[257] Stocker (1979), Brink (1989), Mele (1996).

[258] (1952, pp.125-6). Internalists usually talk instead about the 'inverted commas sense' when referencing Hare but for Hare the 'conventional sense' and the 'inverted commas sense' of a word are different and the 'conventional sense' fits better with what internalists want to say about psychopaths.

[259] e.g. Blair, (1995, 2008).

People suffering from depression are another case that externalists sometimes use as an example of amoralists. These externalists claim that depression can totally sap our motivation to do what we judge to be morally right. In such cases it seems intuitive that the depressed person still counts as making a moral judgement, and not just in a 'conventional sense'. Depressed people often feel extreme frustration at their own inability to get motivated to do what they judge that they ought to do. If their moral judgements were merely conventional then it would be unclear why they should feel this frustration. Moreover, there is no evidence to suggest that, unlike psychopaths, people with depression are unable to clearly distinguish the conventional from the moral.

The best response to the case of depressed people (though this could be applied to psychopaths too) is to question whether they really have *no motivation whatsoever*. That depressed people feel so much frustration that their listlessness prevents them from doing what they think they ought to do suggests to me that they experience some pull to do what they judge they ought to. Jamie Dreier has a helpful way of framing what the externalist requires for their argument,

> "[The externalist asks us to] imagine a man who would not lift a finger, as it were, to achieve what he believes to be good - for if he would make even some minimal effort, then the relevant motivation is exposed."

It seems reasonable to suppose that depressed people (and maybe psychopaths also, insofar as they can make sincere moral judgements) might be motivated to at least lift a finger to bring about what they judge to be right. That is enough to show that there is at least some motivation, which is enough for internalists.[260] It is also worth mentioning that the empirical research suggests that depressed people do feel some motivation to do what they believe is morally right.[261]

---

[260] Eggers (2015).

[261] Roberts, 'Mental Illness, Motivation, and Commitment,'. O'Conner et al. 'Empathy and Depression.' L.E. O'Connor, J.W. Berry, J. Weiss, and P. Gilbert, 'Guilt, Fear, Submission, and Empathy in Depression,'. All cited in Cholbi (2011).

A third class of cases that externalists consider concerns what Brink calls *amoralist sceptics*.[262] These are people who claim to recognise that certain actions are morally right but are sceptical about whether this gives them any reason to do those actions. Some philosophers fall under this category. In *Ethics without Morals*, Joel Marks argues that we have excellent reason not to guide our action by moral concerns.[263] Morality, he argues, can lead to terrible consequences, it makes us feel bad, it leads us to living poorly, etc. In sum Marks thinks that we should reject moral concerns because we would better maximise the fulfilment of our desires.

I think that there are two ways for internalists to account for these kinds of claims. The first is to point out that these philosophers are plausibly talking about morality in the 'conventional sense'. These philosophers are moral error theorists so they are not really making moral judgements when they say things are right or wrong; they do not believe that anything is, *strictly speaking*, right or wrong.

The second reply is to question whether these philosophers are really rejecting morality *tout court* or whether they are just offering an alternative morality. In Marks' case, we might think he is offering a moral account according to which what we ought to do is that which maximises the fulfilment of our desires.[264] Marks would reject this, for him, morality, by its very nature is composed of short, absolute, objective rules which apply to all people at all times and which override any other considerations. What he calls, "ethics", on the other hand, is a far superior way to regulate behaviour in that its rules could apply contingently on the addressee's desires. But at this point the internalist could stipulate that they are using 'morality' in a broad sense such that it is coextensive with what Marks calls 'ethics'. (This seems a legitimate move to me; there is no real difference between 'morality' and 'ethics' to my ear). If Marks wants to insist on his

---

[262] Brink (1989, ch.3)
[263] Marks (2012).
[264] A proper examination of Marks' view would probably give a more complicated picture than this, but it will serve well enough as an example.

distinction, internalists can just redefine their position in terms of 'ethical judgements' rather than 'moral judgements' without any real loss.

The externalist might argue at this point that facts about what sceptics, psychopaths and people with depression are usually like are not crucial for showing that internalism is false. As long as we can *imagine* people who have no motivation whatsoever to do what they judge to be right then unconditional internalism fails, since it is a claim about what is *conceptually* possible rather than what is in fact the case.[265]

However, I do not see why internalists should accept that amoralists are imaginable either. Brink cites as examples of possible amoralists, Thrasymachus in *The Republic*, Hume's knave and Mike the conman from David Mamet's *House of Games*.[266] He argues that we can imagine that these characters are real. However, while it is true that we can imagine these characters, the internalist can argue that what we imagine is someone who would at least raise a finger to do what is right. Alternatively, for each of these cases it is also plausible that we might adopt a 'conventional' reading of their use of moral terms.

Matthew Kramer argues that this response would not work for cases of fictional *immoralists*.[267] He uses the examples of Aaron from *Titus Andronicus* and Iago from *Othello*. These characters have no motivation to do what he judges to be right, quite the opposite; they are motivated to do what is wrong *because it is wrong*. Aaron declares shortly before his execution, "If one good deed in all my life I did I do repent it from my very soul". We have no difficulty in believing in these characters and they do not seem in any way impossible. So, argues Kramer, it cannot be a conceptual truth that we are always motivated to do what we judge to be right.

---

[265] Nichols (2004)
[266] Brink (1989)
[267] Kramer (2009)

Pace Kramer, I think intuitively the 'conventional sense' reply is available in these cases. What we are imagining, an internalist could argue, are deeply anti-social and contrarian characters. When Aaron for example claims that he repents any good actions he may have performed, he could be just evincing his rejection of conventionally accepted values. Kramer argues that this kind of response will not work for Iago at least. He writes,

> "Iago is not seeking to reproduce the moral judgements of ordinary people when in his soliloquies he utters his verdicts on the monstrousness of his own actions. When he declares to himself that his contrivances are diabolical, he is delightedly expressing his own view rather than anyone else's view. For his purposes, he needs to grasp what is morally right and obligatory. Only be apprehending what is actually right and obligatory can he fulfil his objective of acting in defiance of what is right and obligatory."[268]

However, I am not sure that this is a fatal case for internalists either. It might be true that when Iago describes something as 'bad' he is not talking about what is deemed bad by some person in particular but that does not mean that he must be using terms like 'bad' in a genuinely normative way. It seems open to the internalist to argue that Iago is a moral nihilist who delights in betraying others and causing them harm but without really believing that those things have moral significance. Iago might use 'bad' to refer to actions which cause harm, not because he thinks that these actions really are bad, but because he thinks that they are the kinds of actions that would be deemed bad *at the end of inquiry*, or which are deemed bad *by God*, or something along those lines. In other words, while it might be true that Iago is not just referencing the moral judgements of those around him and it might be true that he uses words like 'bad' to try to pick out those things that cause harm and that violate people's rights, it is not obvious that he really views these things as morally bad. It is not unreasonable for the internalist to say that someone who believes that moral facts have no normative authority whatsoever (and are motivated accordingly) does not really believe that there are moral facts at all.[269]

---

[268] ibid. pp.278-9
[269] See Smith (1994 pp.68-70)

So, for all the examples I have looked at so far the internalist has a plausible enough story to offer. I do not claim that the internalist story is always as plausible as the externalist alternative, just these cases are, at most, only slightly damaging to the internalist. In other words, if they cost the internalist any 'plausibility points' at all, the do not cost very many.[270]

A way we might try to break this deadlock is to ask ordinary language users. If it can be shown that ordinary language users overwhelming regard amoralists as conceptually possible then that might tip the scale in the externalists favour. Shaun Nichols and Caj Strandberg & Frederik Björklund conducted studies on students which suggest that, by and large, these students thought that amoralists were possible.[271] Nichols argues that facts about actual people are not crucial, since were competent users of normative concepts to regard amoralist psychopaths as possible that would also undermine internalism. Since ordinary concept users do think that people can make normative judgements without having a particular motivation, internalism cannot be a conceptual truth, or so Nichols argues.

However, things are not so clear cut. First of all, Strandberg & Björklund's tests delivered the opposite results for psychopaths (though they did get externalist-seeming results for cases of depressed people, apathetic people and even normally functioning people). Secondly, Gunnar Björnsson et al. and Jesse Prinz have found that we can get results which seem to support internalism simply by asking the same kinds of questions in a different way.[272] Third, there are methodological worries one could have with all of these studies.[273] For instance, it is not clear that the subjects of these studies were made sufficiently aware of the distinction between genuinely moral judgements and conventional judgements. Nichols did ask his subjects whether they thought that psychopaths could judge things to be "really" right and wrong in an attempt to

---

[270] The 'plausibility points' metaphor comes from Enoch (2011).
[271] Nichols (2002), Strandberg & Björklund (2013).
[272] Björnsson, Strandberg, Olinder, Eriksson & Björklund (2015), Prinz (2007, p.99)
[273] Björnsson, Strandberg, Olinder, Eriksson & Björklund's (ibid) interesting discussion of empirical testing of folk intuitions on internalism also throws up several ways that results from these studies could be misleading.

make the subjects focus only on genuinely moral judgements, but this may have not been enough. Using the qualification, "really", might merely have prompted them to judge what the psychopath thinks people are "really" committed to as oppose to what the psychopath thinks that people might say.[274]

So far, we have not seen a decisive reason to reject internalism. I think it is viable in the face of these examples for the internalist to, as it were, dig in their heels. I move on now to look at the main internalist arguments against externalism.

§4.2 – The Internalist Case against Externalism

Internalists have put forward their own examples designed to show that externalism is false.

First, they argue that if internalism were not true then it would be possible for there to be whole communities, or even whole worlds, in which people make moral judgements but are never motivated to act upon them. They judge that, first instance, it is wrong to cheat or to lie but this gives them no motivation whatsoever to refrain from cheating and lying. Internalists think that intuitively such a world is not possible. [275]

Second, internalists argue that there could be whole worlds in which people make moral judgements but are reliably disposed to act *contrary* to them. For example, people might sincerely judge that causing pain has the property *is wrong* and be motivated to cause pain because it is wrong.[276] Such people, internalists argue, could not count as making sincere moral judgements, *contra* externalism.

Third, internalists argue that if there were no necessary connection between normative judgements and motivation then there could be *amoral saints*; people who judge that all kinds of

---

[274] Ridge (2014), pp. 52-3.
[275] Lenman (1999), Ridge (2014).
[276] Dreier (1990)

terrible things are permissible, or even obligatory, but who nevertheless behave impeccably because they do not care to act in accordance with their horrible moral judgements.[277]

Such examples do not settle the debate. First of all, it is not clear that it follows from the falsity of internalism that these examples would be possible. Some cognitivist theories have it that the content of normative judgements is, in some way or other, tied to the desires, attitudes or plans of normative agents or their communities. For instance, on David Copp's view judging that we have a reason to do something is, very roughly, to judge that some rule or norm which would promote certain desires and interests of our community would permit or prescribe it.[278] This view is compatible with there being one person who does not have motivation to do what they judge that they have most reason to do. It does not, however, follow that if Copp's view were correct then a whole community could make such judgements without ever having any motivation to act in accordance with them. Since the content of normative judgements is, on this kind of view, a function of the community's desires and interests it is not surprising if we cannot imagine a whole community of people who are never motivated to act in accordance with them, or who are reliably motivated to act contrary to them. The same might be said of lots of other cognitivist theories.[279] So, we should not assume that if internalism is false then there could be no tie between the motivations of a whole community worth of people and their normative judgements.

Second, the cognitivist can insist that such examples are conceivable. They could try to explain away some of our resistance to the examples on the grounds that a community who had no motivation to act in accordance with what they judged to be right, would have no real reason to bother making moral judgements in the first place. At the same time, the cognitivist can maintain

[277] Lenman (2003b)
[278] Copp (2003)
[279] e.g. Indexical relativist accounts like that of Jamie Dreier, Aristotelian contextualist accounts like that of Judith Jarvis Thomson or Phillipa Foot, end-relational theories like those defended by Stephen Finlay and Daan Evers, and sophisticated subjectivist accounts argued for by the likes of Peter Railton.

that the examples are not logically impossible. It is hard to come up with a reason, or at least one that does not just assume some form of internalism, why this would be mistaken. We seem perfectly capable of conceiving of worlds of amoralists in fantasy and science fiction novels. The internalist will reply that what we actually imagine is a world of people who make moral judgements only in a 'conventional sense', but the externalist is not forced to accept this. The externalist reading seems as viable as the internalist's alternative.

Third, cognitivists might even argue that some examples of amoralist communities support externalism more than internalism.[280] In a very interesting paper, Gunnar Björnsson and Ragnar Francén Olinder get us to imagine that psychological egoism turns out to be true. According to this cynical hypothesis we only ever act for self-interested reasons and never for moral reasons.[281] Though we delude ourselves that we are sometimes altruistic, in fact we really just look to maximise our self-interest. The delusion that we are altruistic and moral is, according to this hypothesis, a convenient psychological device for promoting our own self-interest because people who consciously act in only self-interested ways do not tend to prosper over the long term. The fact that we often behave in accordance with moral rules is not, according to this view, because we are concerned to do what is right (either *de dicto* or *de re*), rather, it is because we want to be seen as being morally upstanding to reap various certain social benefits. Though we might be generally motivated to do what we judge to be right, on this picture, we are not motivated to do what we judge to be right *because we judge it to be right*; deep down we are really just looking out for number one.

This cynical hypothesis is, Björnsson and Oliner acknowledge, not likely to be true. The chance that it is true is, however, greater than zero. If internalism were true then discovering that the cynical hypothesis were true should lead us to judge that we had never made any genuine moral

[280] Gert and Mele (2005).
[281] Björnsson and Olinder (2013)

judgments. If anyone who makes a genuine moral judgement necessarily has at least some motivation to act on that judgement then we could not have made genuine moral judgements if the cynical hypothesis were true, since, according to that hypothesis, we never have *moral* motivations, only self-interested ones. Björnsson and Oliner contend that this is counterintuitive. It seems much more reasonable to suppose that, if we discovered the cynical hypothesis was true then we would judge that we do make moral judgements but that these just do not have the kinds of links to motivation that the internalist supposes. In other words, discovering the cynical hypothesis were true might lead us to reconsider many of our beliefs about the nature of moral judgements, but it would not lead us to conclude that we had never made moral judgements. This suggests that it is not *conceptually necessary* that anyone who applies the moral concept <right> to some action is *necessarily* motivated to perform that action, since it is conceivable that we might find out that this is not true of us.

There is more that could be said in response to these examples, but we have seen that externalists seem to have viable rebuttals to the internalist's arguments.

§4.3

There are many more examples put forward by internalists and externalists in the literature. Often, these are introduced as examples which are finally going to break the deadlock, but this invariably proves not to be the case. Whether or not we think that examples of amoralists are possible typically depends on whether we already have internalist or externalist sympathies. Since the debate largely boils down to this clash of intuitions, it has reached an impasse. That being so, cognitivism's failure to accommodate internalism cannot be seen as a reason to discount it at this stage. So, the naturalist is safe from the argument from motivational internalism.

Leaving matters there would, however, be unsatisfying, so in what follows I argue against internalism in an indirect way. I accept, at least for the sake of argument, that internalism's truth

would count heavily against cognitivism and heavily in favour of non-cognitivism.[282] Since there are no necessary connections between distinct existences, if it is the case that there is a necessary link between normative judgements and desires, it must be the case that normative judgements are (at least partly) constituted by desires.[283] Non-cognitivists think that the normative judgement that some action is right necessarily provide us with some motivation to perform that action by itself simply because that normative judgement consists in a motivation to perform that action. If non-cognitivism is correct then the *strength* of that motivation ought to covary with the *strength* of the normative judgement that that desire constitutes.[284] However, the strength of our normative motivation does not seem to covary with the strength of our normative judgements, so normative judgements cannot be motivational states, or so I shall argue. This provides us with a premise in the following *modus tollens*:

(1) If internalism were true then normative judgements would be motivational states.
(2) Normative judgements are not motivational states.
(3) Therefore, it is not the case that internalism is true.

My auxiliary argument for the second premise has a lot in common with an argument that has been made by Nick Zangwill.[285] However, unlike my argument, non-cognitivists have the resources to respond to Zangwill's argument and to a similar argument that Zangwill could have made. It will be helpful for setting up my argument if I first show why this is the case.

§5

§5.1 – Zangwill's Argument

According to non-cognitivists, normative judgements are desires. Desires vary in degrees of strength. We desire some things more than others and we desire those things to different degrees

---

[282] I do not actually think that cognitivism is necessarily ruled out by internalism since there remains the possibility that a theory like Tresan's or Copp's might be shown to work. That is fine from my point of view since these theories are broadly naturalistic.
[283] Brink (1989, chapter 3) also argues that the fates of non-cognitivism and unconditional internalism are joined.
[284] I will explain what it means for normative judgements to be weak or strong shortly.
[285] Zangwill (2008, 2009).

at different times. Normative judgements must also vary in degrees of strength if non-cognitivism is true. One way that normative judgements seem to vary in strength is in terms of the degree of *certainty* we have that those normative judgements are correct.[286] I might judge both that giving money to Oxfam is right and that giving money to the Labour party is right, but be much more confident in the former judgement than I am in the latter. Nick Zangwill argues that the non-cognitivist must explain this in terms of the strength of the desires that constitute those normative judgements, the idea being that we are more confident that some action is right the stronger the desire to do it is. Recall now that for non-cognitivists the desires that constitute moral judgements are also what explain the normative motivation to perform those actions.[287] If non-cognitivism is true then, desire strength is doing double duty; it explains a normative judgement's motivational efficacy and it explains the certainty with which we hold that judgement. This means that the non-cognitivist would be committed to saying that the certainty of normative judgements necessarily covaries with the normative motivation the agent has to perform that action. The level of normative motivation to perform an action could not change while the certainty remained the same because the level of moral motivation and the level of certainty are both explained by desire strength. Let us call this claim the *Proportionate Certitude Determination Thesis* (PCDT).[288]

There are plausibly counterexamples to PCDT:

> *Example 1*: John judges that it is right to pick his kids up from school and he is fairly certain that his judgement is correct. On Monday, John is not looking forward to picking up his kids because he hates driving but he also thinks that it is the right thing to do. John recognises that his kids will be stranded at school if he does not pick them up and he believes that he has a duty to pick them up having previously agreed to. Since John is a decent person he feels a fairly high level of motivation overall to pick up his kids. On Tuesday, however, John is in a depressed state. He realises that picking up his kids is the right thing to

---

[286] Zangwill (2008, p.95)

[287] I say 'normative motivation' because we can of course have other motivations to perform those actions that we judge to be right. For example, I might judge that it is right to give to Oxfam but be motivated to give to Oxfam largely by some other mental state which is distinct from that normative judgement, such as a desire to be well thought of by others.

[288] I am using, with slight tweaks, Zangwill's (2008) terminology here.

do but his judgement that it is the right thing to do moves him much less than it did on Monday. Intuitively, John need not be any less certain on Tuesday than he was on Monday that picking up his kids is the right thing to do, but he has less normative motivation to pick them up.

The point of this example is not that on one day John is motivated enough to act as he judges he ought to and the other he is not, nor is the point that on one day John's normative motivation is outweighed by other motivations. The point is that John is motivated less by his normative judgement on one day than the other, regardless of other motivations, despite his confidence in the truth of those judgements staying the same. Such examples seem conceptually possible at the very least. So, we have a counterexample to PCDT. Since, Zangwill argues, non-cognitivism entails PCDT, non-cognitivism must also be false. Because non-cognitivists try to explain both the certitude of our normative judgements and their motivational efficacy in terms of desire strength, they have no resources to explain how the certitude of our normative judgements and the motivational efficacy of normative judgements can come apart.

Cognitivists meanwhile do not face this problem. They can explain why John is just as certain on Monday as he is on Tuesday that picking up his kids is right by saying that he gives equal credence to his belief that picking up his kids has a certain property on both days. The fact that John is less motivated to pick up his kids on one day than on the other does not conflict with that explanation. John's level of normative motivation is explained, on this account, by the strength of John's desires (*de re* or *de dicto*) to do what he judges to be right. Since John's normative judgements and the desires that explain John's normative motivation are separate mental states, on this account, they need not have the same strength.

§5.2 – An Argument Zangwill Could Have Made

Another way that moral judgements vary in strength is the degree to which we a moral property is instantiated in an action. For example, we typically think that lying and killing are both wrong, but we think that killing is wrong to a greater extent. Michael Smith calls this feature

'importance'.[289] If a judgement that some action is wrong has high 'importance' then that action is judged to be wrong to a high degree. From now on when I speak about a judgement's 'importance' or of a judgement being 'important' I will always be using these terms in this sense. If the importance of a judgement is also to be determined by the strength of the desire that constitutes the normative judgement, then the agent's normative motivation must also covary with the importance of the judgement. Let us call this claim the *Proportionate Importance Determination Thesis* (PIDT).

Zangwill does not consider PIDT a problem for non-cognitivists, he says that this claim is "relatively uncontroversial", but I think that a commitment to PIDT would be just as troubling as a commitment to PCDT. It is often the case that when one action is right to a greater degree than some other action, we will be proportionally more motivated to perform the first action than the second. If the judgement that the right thing to do is to help my neighbour move house has less importance for me than the judgement that the right thing to do is to help my mum move house, then I will usually be more motivated to help my mum than to help my neighbour. Additionally, the degree to which I am motivated to help my mum move and the degree to which I am motivated to help my neighbour move will usually be proportional to the degree to which I think those acts are right. However, this will not always be the case as the following example illustrates:

> *Example 2:* John judges that it is right to pick his kids up from school, he judges that picking up his kids is morally right to a high degree and he judges that nothing else he could be doing that day is as normatively pressing as his duty to pick up his kids. John's normative judgement has high importance. On Monday, the fact that he judges that it is the right thing to do supplies John with a significant amount of motivation to pick up his kids, but on Tuesday John is barely motivated at all by the judgement that picking up his kids is right because he is in a depressed state. Nevertheless, on both Monday and Tuesday, John judges that picking up his kids is right to the same degree.

---

[289] Smith (2002)

If this example is conceptually possible, and it seems that it is, then we have a counterexample to PIDT. The degree to which we judge something to be normatively important does not necessarily covary with our motivation to do the right thing. We can become more or less motivated to perform some action because it is right without also changing our mind at all concerning the degree to which that action is right. The non-cognitivist would be committed to denying that this is possible if they are also committed to PIDT.

So, the cognitivist has two potentially powerful arguments against non-cognitivism:

Argument from PCDT

    1. If non-cognitivism is true then PCDT is true

    2. Someone could remain as certain that an action is right but become more or less normatively motivated to do it.

    3. Therefore, PDCT is false.

    4. Therefore, non-cognitivism is false.

Argument from PIDT

    1. If non-cognitivism is true then PIDT is true

    2. The importance of a person's judgement that an action is right could stay the same but they could become more or less normatively motivated to do it.

    3. Therefore, PIDT is false

    4. Therefore, non-cognitivism is false.

The problem for the non-cognitivist is that in identifying normative judgements with the same states of mind that motivate us to act on those judgements, they make the link between judgement and motivation too close. The strength of normative judgements is generally less variable than our normative motivation. I will now suggest that both of these arguments fail, but then go on to argue that a similar argument is telling.

§6

I think that non-cognitivists can refute both Zangwill's argument and the one that I have constructed. Quite aside from issues regarding normative motivation, a commitment to both PCDT and PIDT would be a problem for non-cognitivists. This point becomes clear if we consider Michael Smith's paper *Evaluation, Uncertainty and Motivation* and the discussion that has followed from it.[290]

Smith argues that normative judgements cannot be desires because normative judgements vary along more dimensions than desires do. While desires only vary in strength, normative judgements can vary in terms of certitude and importance. Desire strength alone cannot be used to explain all the different ways that normative judgements can vary in strength. If the strength of the desire that constitutes a normative judgement explains both its certitude and importance, then non-cognitivists cannot explain why normative judgements can change in certitude without changing in importance, and vice versa. We can come to be less sure that some action is wrong without also coming to think that the action in question is less wrong than we previously did, and we can come to think that some action is not as wrong as we used to without being any less sure that it is wrong. Non-cognitivists need to accommodate these facts if they are to have a satisfactory account of normative judgement.

In response, non-cognitivists have argued that neither certitude nor importance were to be explained in terms of the strength of the desire that constitutes the normative judgement.[291] If this is right then it undercuts both arguments considered in the last section, since the first premises of those arguments would be false.

---

[290] Smith (2002)
[291] Lenman (2003a), Ridge (2003).

Non-cognitivists have argued that being very certain that some action is right is, in effect, to be very certain that we would judge that action to be right if we had a perfect normative character. [292] A classic statement of this idea comes from Simon Blackburn:

> The problem comes with thinking of myself (or of us or our tradition) that I may be mistaken. How can I make sense of fears of my own fallibility? Well, there are a number of things I admire: for instance, information, sensitivity, maturity, imagination, coherence. I know that other people show defects in these respects, and that these defects lead to bad opinions. But can I exempt myself from the same possibility? Of course not (that would be unpardonably smug). So I can think that perhaps some of my opinions are due to defects of information, sensitivity, maturity, imagination and coherence.[293]

There are certain traits that we think make people better at making normative judgements. An *ideal normative agent* would have these traits to an ideal extent and so could not be bettered when it comes to making normative decisions. Judging that it is very certain that some action is right is, on this view, to judge that an ideal normative agent would judge it to be right. Because this is a non-cognitivist account, judging that an ideal normative agent would judge some action to be right is understood as judging that the ideal normative agent would desire that that action was performed. So, for example, if I think that it is right to give £10 to Oxfam then I judge that an ideal normative agent would desire that I give £10 to Oxfam.

The judgement that some hypothetical agent counts as normatively ideal is also normative judgement. This judgement must then be a desire if non-cognitivism is true, since non-cognitivists think that all normative judgements are desires. What they say is that judging that a hypothetical agent is an ideal normative agent is to desire to act as that hypothetical agent would desire that we do. For example, to judge that Martin Luther King is an ideal normative agent, is, on the picture, to desire to act as Martin Luther King would desire me to.

For Blackburn, an ideal normative agent is one who is well-informed, sensitive, mature, imaginative and coherent but other people might think that ideal normative agents have

---

[292] Blackburn (1998), Lenman (2003a), Ridge (2003), Gibbard (2011).
[293] Blackburn (1998), p. 318

completely different traits. Some people might, for instance, judge that maturity has a corrupting influence and that only those with a childlike naïveté could be ideal normative agents. We can also differ in our views about the extent to which ideal normative agents have particular traits. For instance, someone might think that an ideal normative agent must be sensitive, but think that they would be sensitive to a lesser degree than Blackburn does. The fact that different people judge different hypothetical agents to be ideal is understood, on this view, as reflecting the fact that different people desire to act as different agents would desire them to.

On this account then, the judgement that some action is right consists of both a belief that some hypothetical agent would desire that I perform that action and a desire to do as that hypothetical agent would desire me to. We can be more or less certain that hypothetical agents would have certain desires. Suppose I think that a normatively ideal agent is just one who is fully-informed on all non-normative matters, i.e. I desire to do the things a fully-informed agent would desire that I do. When I judge that it is right for me to give £10 to Oxfam I judge that a fully-informed agent would desire that I give £10 to Oxfam. I could be more or less certain about that judgement. For all I know, a fully-informed agent might desire that I give money to a different charity since they would know which charities most effectively use money that is donated to them. A fully-informed agent might even desire that I keep the money for myself. So, the certainty of normative judgements is understood on this view in terms of the certainty of our belief regarding what a particular hypothetical agent would desire them to do.

The importance of a normative judgement, meanwhile, is most naturally explained in terms of the strength of the desire that I believe that an ideal normative agent would have towards the action in question. Suppose again that I think that ideal agents are just fully-informed agents. In that case for me to think that it is better to give £10 to Oxfam than to Shelter is, on this view, to judge that a fully-informed agent would desire that I give £10 to Oxfam more than they would desire me to give £10 to Shelter.

We can apply this picture to the example of John I gave earlier. Let us suppose that, according to John, an ideal normative agent is just one who is fully-informed. John's normative judgement that it is right to a very high degree to pick up his kids could be said to be made up of three component judgements:

> 1. John's judgement that a fully-informed agent is normatively ideal.
> 2. John's judgement that a fully-informed agent would judge that it is right for him to pick up his kids.
> 3. John's judgement that a fully-informed agent would judge that it is right to a high degree for him to pick up his kids.

The first of these is a normative judgement so it must be understood by the non-cognitivist as a desire. The other two judgements are not normative judgements, they are just judgements concerning what a particular hypothetical agent would think, so the non-cognitivist can understand these judgments as being ordinary beliefs about a hypothetical agent's judgements. The resulting picture looks a bit like this:

> 1. John's judgement that a fully-informed agent is normatively ideal ≈ John's *desire* to act as a fully-informed agent would desire that he does.
> 2. John's judgement that a fully-informed agent would judge that it is right for him to pick up his kids ≈ John's *belief* that a fully-informed agent would desire that he picks up his kids.
> 3. John's judgement that a fully-informed agent would judge that it is right to a high degree for him to pick up his kids ≈ John's *belief* that a fully-informed agent would strongly desire that he picks up his kids.

The certitude and importance of John's normative judgement that it is right that he picks up his kids is explained by credence he gives to the second and third component judgments respectively. We can explain why John is less motivated by his judgement that picking up his kids is right on Tuesday than he is on Monday by the fact that the strength of his desire to do as a fully-informed agent would desire that he does has gone down due to his depressed state. This need not affect the certainty that John has on Monday or on Tuesday that picking up his kids is right. The certainty of his judgement is explained by how certain he is that a fully-informed agent would desire that he picks up his kids. Since the credence that he gives to that belief has not changed, neither has the certainty he has that it is right to pick up his kids.

John could also think that picking up his kids is right to do the same degree on Tuesday as it is on Monday because his depressed state need not change the credence he gives to his belief concerning the strength that a fully-informed agent's desire that he pick up his kids would have. So, the non-cognitivist can explain Example 2 as well.

The second and third component judgements are just beliefs. Since beliefs are motivationally inert, they do not also explain the motivational efficacy of John's normative judgement that it is right that he picks up his kids. So, the motivational efficacy of John's judgement can become stronger or weaker without the importance or certitude of his judgement also becoming stronger or weaker. This shows that non-cognitivists are not committed to PCDT or PIDT after all. Therefore, the arguments considered in the last section both fail. Those arguments presuppose that non-cognitivism must take the very simple form that normative judgements are just desires, when actually non-cognitivists are entitled to say that normative judgements are more complex than that. The non-cognitivist can explain importance and certitude in terms of the belief-components of normative judgements so neither certitude nor importance need co-vary in strength to the motivational component of normative judgements. However, I think that we can tweak these arguments so that they apply to this more complex non-cognitivist picture.

§7

There are plenty of potential problems with this non-cognitivist account but I want to focus on just one.[294] The desire that we have to do as some particular agent does is an important component of the normative judgement that some action is right on the non-cognitivist picture we are considering. That desire constitutes the component judgement that a particular agent is normatively ideal but it also explains why we necessarily feel some motivation to perform the action we are judging to be right (since it is the only motivational component of that judgement).

---

[294] See especially Egan (2005), but also Olson and Bykvist (2009), Scanlon (2009), Parfit (2011).

That being so, the strength of the component judgement that a particular agent is ideal should be proportional to the level of normative motivation we feel to perform the action we judge that to be right (or in other words the action that we believe that the hypothetical agent would desire that we do). The problem for the non-cognitivist is that we can run the same arguments we just considered against this claim.

Since desires come in degrees of strength, the judgement that a particular agent is ideal should also admit of degrees of strength. Our *certainty* that a particular agent is normatively ideal seems to come in degrees. I am much more certain that a fully-informed agent is normatively ideal than I am that a fully-misinformed agent is normatively ideal, and I am much more certain that an empathetic agent is normatively ideal than a sadistic one, to take just two examples. The only resource that the non-cognitivist has to explain this kind of certainty is in terms of the strength of the desire that constitutes our judgement that a particular agent is normatively ideal. The fact that I am more certain that a fully-informed agent is normatively ideal than I am that a fully-misinformed agent is normatively ideal could then be explained by the fact that I desire much more strongly to do as a fully-informed agent would desire that I do than as a fully-misinformed agent would desire that I do. If my desire to do as a particular agent would desire that I do explains both my judgement that that agent is normatively ideal *and* the level of motivation I feel to do what I judge to be right, then the level of *certainty* I feel that a particular agent is ideal should covary with the level of motivation we feel to perform the action we judge to be right. Let us call this claim the *Proportionate Certitude Determination Thesis 2* (PCDT2). PCDT2 is vulnerable to the kinds of examples we have previously looked at. Consider the following scenario:

> *Example 3*: John thinks that the ideal normative agent is just a fully-informed agent. He also thinks that a fully-informed agent would desire that he picks up his kids from school. As a result, John judges that the right thing to do is to pick up his kids. On Monday, he is very motivated to pick up kids because it is the right thing to do, but on Tuesday, due to his being in a depressed state, he is barely moved by the fact that

picking up his kids is the right thing to do. John is no less convinced on Tuesday than on Monday that fully-informed agents are normatively ideal.

Example 3 looks like a coherent possibility but it should not be if the non-cognitivist is right. If John's judgement that full-informed agents are normatively ideal is a desire to do as a fully-informed agent would desire him to do, and if that judgement is a component of John's judgement that picking up his kids is the right thing to do, then his normative motivation to pick up his kids should covary in strength with his judgement that fully-informed agents are normatively ideal. This seems not to be the case. Our level of normative motivation seems to be less stable than our judgements concerning which kinds of agents are normatively ideal.

The non-cognitivist might reply at this point that the certainty we have that a particular normative agent is ideal is not to be explained in terms of the strength of the desire that constitutes that judgement. The problem is that if non-cognitivists want to hold that the normative judgement that a particular agent is ideal is constituted by a desire then they need to explain how that judgement admits of degrees of certitude. They cannot respond as they did to Michael Smith's challenge, i.e. they cannot bring in the desires of hypothetical agents to explain how certain we are that a particular agent is normatively ideal. They could not say that our level of certainty that a hypothetical agent is ideal is explained by the credence we assign to the belief that that hypothetical agent would judge *themselves* to be ideal. Being certain that some agent is ideal cannot be understood as being very sure that they would desire that we desire to act as they would want us to. I might be very sure that a particular hypothetical agent would want me to act as they would desire me to, while being very unsure whether that agent is ideal. A maximally self-regarding agent would presumably want me to act as they desire me to, but I am very confident that they are not ideal normative agents.

Alternatively, the non-cognitivist might try to say that our certainty that a particular hypothetical agent is ideal can be explained by the credence of our belief that a *different hypothetical agent* would

judge the first agent to be ideal. This could not work either. The first agent is, *ex hypothesi*, one that we judge to be normatively ideal. We could not explain our confidence that that agent is ideal in terms of what some other, non-ideal agent desires; that agent's desires are irrelevant to us, normatively speaking. I might be quite unsure whether agent X is normatively ideal while being very sure that that agent Y would desire that I act as agent X would desire that I do.

Additionally, even if we did say that judging that some agent is normatively ideal is to judge that some other agent would desire that we desire as the first agent does, we would then have to explain our level of certainty in the judgment that the second agent is an appropriate judge of the first agent. If we did that in terms of the strength of our desire to desire as the second agent does, then the same problem would re-emerge. And, if we tried to explain it in terms of the desires of some third hypothetical agent, we would be walking into an infinite regress in which we invoke more and more hypothetical agents in order to explain our certitude that any one of them is normatively ideal.

So, while I think that the non-cognitivist can get round Zangwill's argument by drawing on the belief-components of normative judgements, the same basic argument can then be reapplied to the desire-components of normative judgements. It does not seem that desires could explain both the level of motivation we necessary feel to act and be the component judgement that a particular agent is ideal, since the two can vary in strength. The non-cognitivist ties normative judgement too closely to motivation, such that they end up with a very implausible picture of how our normative motivation works. If, as we have assumed for the sake of argument, the truth of internalism requires that normative judgements are (at least partly) constituted by a motivational state, then the cognitivist can reject internalism too.

The cognitivist thinks that normative judgements and the states of mind that motivate to act on those judgements are distinct states of mind. This allows them to explain why features of normative judgements can change while motivation remains the same, and vice versa. But since

there are no necessary connections between distinct existences, it does require them to reject internalism. This is acceptable since accepting internalism requires that we also accept a flawed picture of normative motivation.

§8 – Conditional Internalism Again

Not all non-cognitivists accept unconditional internalism. Some non-cognitivists say instead that when we judge that something is morally right we will necessarily be motivated to do it *as long as we meet condition C*, where C stipulates that we are suitably normal in some way.[295] For example, C may require that the speaker is free from depression. If the non-cognitivist wants instead to defend a conditional internalism then they may be able to avoid counterexamples like the example of John I gave since his depression is what explains the variability of his motivation.

There are good reasons to doubt that non-cognitivists can avoid a commitment to an unconditional form of internalism.[296] If normative judgements are desires, and desires are motivational states, then everyone, including people who meet condition C, who makes a normative judgement must surely be in a motivational state. Maybe the non-cognitivist can elaborate on the kind of desire they have in mind. Perhaps normative judgements are kinds of conative mental states that only motivate if you are not depressed. On this picture, desires are not necessarily motivating, they only motivate some of the agents who have them; those agents that are 'normal' in some stipulated sense. But without any motivation being present it is difficult to see what the content of 'abnormal' agent's normative judgements would be on this picture. In other words, if it is not motivation, then there is nothing for the desires that constitute 'abnormal' agents normative judgements to consist in. This problem seems to extend to any other form of conditional internalism. If ('abnormal') agents who fail to meet condition C can make sincere normative judgements without having any motivation to act in accordance with

---

[295] e.g. Gibbard (1993), Blackburn (1995).
[296] Standberg (2012)

them it becomes very difficult to see in what way those judgements could be said to consist in a motivational states of mind.

I think we can put that issue to one side however, because I think that the problem I considered for unconditional internalism, applies just as much to conditional internalism. In other words, it seems likely that even the most 'normal' normative agents will be agents whose motivation to do what they judge to be right will not covary with the certitude of their judgement that a particular agent is ideal. Depressed people like John seem like a particularly clear example but I think that for any of us the level of our motivation to do what we judge to be right does not covary with certitude we have that a particular agent ideal. Our views about which kinds of agents are normatively ideal are, I submit, fairly stable. Sometimes we might re-examine our commitments and wonder whether we should be acting as a more spontaneous or a more thoughtful or a more mature agent would desire for us to but for the most part these judgements are not in flux. Our motivation to do what we judge to be right meanwhile is rather less stable.

Studies have shown that we seem to be less motivated to help others if the environment we are in is noisy than if they are quiet.[297] But on reflection we don't tend to think that the level of noise is a relevant factor, i.e. we do not tend to think that a normatively ideal agent would desire to a greater degree that we help if surroundings are quiet than if they are noisy.[298] This seems like a case where the amount we are moved by judging that something is right does not match the strength of our judgement that a particular agent is normatively ideal. For other examples, we might point to the fact that fluctuations in mood can affect how much our normative judgements motivate us without changing how sure we are that certain kinds of hypothetical agents are normatively ideal. Since the strength of even normal agents' motivation to do what we judge to be right because it is right does not covary with the strength of their judgement that a

---

[297] Mathews, K. E., & Canon, L. K. (1975). "Environmental noise level as a determinant of helping behaviour", *Journal of Personality and Social Psychology, 32*(4)

[298] Though there may some cases where we think this is so. Failing to help an old lady cross the street in deafeningly loud conditions is less wrong than failing to help her in peaceful and safe auditory conditions.

particular kind of normative agent is ideal, our normative judgements cannot be motivational states. Since normative judgements cannot be motivational states, and since we are assuming that the only way that there could be necessary connections between normative judgment and motivation would be if normative judgements were motivational states, we can reject that there are necessary connections between the normative judgements of even the most 'normal' people and their motivation.

§9 – More Problems Relating to Normative Motivation

§9.1 – Which Motivational State?

The arguments considered so far tell against the claim that normative judgements are constituted by motivational states. I now want to consider some other, admittedly less decisive, problems for the non-cognitivist explanation of normative motivation in order to bolster my case.

When it comes to many normative judgements there is an obvious suggestion for what motivation the agent must have if internalism is true. If normative judgements are desires, and if we understand desires in terms of their having a mind-to-world direction of fit, then a normative judgement of the form 'X is right' will be constituted by a disposition to change the world so as to match the content of the desire. For example, when John judges, 'Picking up my kids now would be the right thing to do', the internalist might say that John would necessarily be motivated to pick up his kids. To put it another way, in making the normative judgement he does John will necessarily be disposed to change the world so that it becomes the case that he picks up his kids.

In other cases, however, this account looks implausible. Suppose Dan judges it would be morally right for John to pick up his kids. It is not plausible that Dan *necessarily* is motivated to pick up John's kids. So what action must Dan be motivated to perform if internalism is true? We do not make normative judgements just about actions that we ourselves could perform; we judge the

actions of others, actions of our past selves, the actions of fictional characters, etc. Non-cognitivists need to find some action that an agent will plausibly necessarily feel some motivation to perform for each of these judgement types.

Simon Blackburn offers one potential answer in his reply to an objection from Sigrun Svavarsdóttir, which is along the same lines as the one I am making.[299] Blackburn suggests that in these kinds of cases what we are necessarily motivated to do is to *encourage* the action in question.[300] This suggestion preserves a link between judgement and motivation but I worry that the move is *ad hoc*; why should some moral judgements require motivation to perform the action that is the object of the judgement, while others motivate us to adopt certain attitudes towards people who perform those actions?

A more serious problem is this: the normative judgements, 'The right thing for John to do is to pick up his kids', and, 'Encouraging John to pick up his kids is the right for me to do', are obviously different judgements. However, if normative judgements are constituted by a desire and if that desire is what motivates the corresponding action, then these judgements come out as being identical since both judgements would consist in a desire to encourage John to pick up his kids. 'Encouraging John to pick up his kids is the right for me to do', is a first-personal normative judgement, so on the story we are considering, the desire-state that it consists in is a desire to encourage John to pick up his kids. "The right thing for John to do is to pick up his kids", meanwhile, is not a first-personal moral judgement, so it must consist in a desire to encourage John to pick up his kids too. It would be a problem for non-cognitivism if these judgements come out as being the same judgement since they clearly are not.

Problems multiply; consider the following argument:

---

[299] Svavarsdóttir (2002), Blackburn (2002)
[300] His answer is a little more complex than this, but not in a way that affects the point that I am making.

1. It was right for me to pick up my kids.
2. It was right for John to pick up his kids.
3. Therefore, we both did something that was right.

The first premise, on this picture, expresses a desire for me to do something, whereas the second premise expresses a desire to encourage John to do something. What kind of desire does the use of 'right' in the conclusion then express? Since 'right' is predicated of both one of my actions and one of yours it does not seem that 'right' can express either a desire for me to do something or a desire to encourage you to do something. 'Right' in the conclusion however needs to have the same meaning (which means that it must express the same kind of attitude in non-cognitivist semantics) that it does in the premises, otherwise the argument will not be valid, which it clearly is.

To get round these issues we might then suggest that *all* judgements concerning rightness, including first-personal judgements, are necessarily linked to a motivation to encourage the action in question. On this view, the judgement, 'Encouraging John to pick up his kids would be the right thing for me to do', would consist in a desire to encourage myself to encourage John to pick up his kids. This is a different attitude to the desire to encourage John to pick up his kids, so this suggestion avoids the problems just raised.

However, now the link between judgement and motivation is becoming quite remote. If *that* is supposed to be the link between judgement and motivation, then it seems of little to no cost for externalists to deny that there is any such link. Or to bring this back to the issue of naturalism, if the sense in which normative judgements are normative is that anyone who makes them necessarily desires to *encourage* the subject to follow them, then it does not seem to be much of a cost to the naturalist to deny that normative judgements are normative *in this sense*.

More to the point, this account of the link between judgement and motivation seems to be vulnerable to counterexamples. Here is one: suppose an evil demon causes everyone lots of pain

if anyone *encourages* you to visit your mother but does nothing if you visit your mother. A strict utilitarian could judge that it is right for you to visit your mother without seemingly having any motivation to encourage you to do it. In that case, the judgement, 'The right thing for you to do is to visit your mother', cannot be constituted by a desire to encourage you to visit your mother.

If that example seems too fantastical, we can bring up examples that are more realistic. I might judge that it would be right for John to donate money to charity, while not having any motivation to encourage John to donate; I might want John to donate entirely at his own behest or I might think that by giving John a social incentive I will taint the moral worth of his actions. It is not difficult to come up with examples of actions we judge to be right that we are not motivated to encourage. Since this non-cognitivist picture identifies our rightness-judgements with motivations to encourage certain actions this is a problem for that picture.[301]

§9.2 – Frege-Geach

Sometimes Blackburn and other non-cognitivists are a bit more circumspect when it comes to identifying what the link between normative judgements and motivation is. In *The Flight from Reality*, Blackburn compares normative judgements with the attitude of loving someone.[302] We do not expect, for every individual and for all times, that if someone counts as loving someone else then they will have any particular level of motivation to perform any particular actions. The picture is more complicated and nuanced. Love can drive someone on one day to deeds of heroism but on another day drive them to deeds of cruelty, and on other days have little effect at all on their will.

A problem with this approach is that the non-cognitivist needs to be precise about which attitudes constitute which normative judgements if they are to have a satisfactory semantic account. Normative judgements stand in the same kinds of semantic and logical relations to one

---

[301] Toppinen (2015) provides a compelling and more detailed case against this kind of view.
[302] Blackburn (1995).

another as do non-normative judgements. Cognitivists think that the contents of both normative and non-normative judgements are just *propositions*. They can explain why normative judgements stand in the semantic and logical relations that they do fairly easily, since propositions stand in the same kinds of logical relations to one another that judgements do. Non-cognitivists meanwhile defend non-propositional semantic theories. They think that the semantic contents of normative claims are desires, not propositions.[303] That being so, it needs to be the case that those desires enter into the same semantic and logical relations that normative judgements do if non-cognitivism is to succeed. For instance, it needs to be the case that the desires that are expressed by the normative judgement, 'Giving to charity is right and volunteering is right', entail the desire that is expressed by, 'Giving to charity is right'. Non-cognitivists need to show why the logical rules, such as conjunction elimination, apply to normative claims, even though they just express desires. This is known as the *Frege-Geach Problem*.[304]

Non-cognitivists typically try to solve this problem by arguing that desires to perform certain actions commit the speaker to desiring to perform other actions. For example, they would argue that 'Giving to charity is right and volunteering is right' entails 'Giving to charity is right' because the first claim expresses a desire to give to charity and to volunteer. Having this desire commits the agent to having the desire to give to charity, it is argued, and this is just the attitude that is expressed by 'Giving to charity is right'.

This kind of explanation is also used to account for the validity of *modus ponens*. 'Giving to charity is right' and 'If giving to charity is right then volunteering is right' entails 'Volunteering is right' because accepting the first judgment involves having a desire to give to charity, and accepting the second judgement involves having a conditional desire to volunteer if one desires to give to

[303] Blackburn (1988), Gibbard (2003), Schroeder (2008a)
[304] The literature on this problem is vast but some sources worth mentioning are Geach (1960, 1965), Blackburn (1984, 1993, chapter 10), Hale (1993), Gibbard (2003), Schroeder (2008a, 2010).

charity. Together these desires would commit the agent to desiring to volunteer; the attitude that is the semantic content of the judgement, 'Volunteering is right'.

Whether or not this story is ultimately plausible is not our concern here. What is important is that this account requires that the motivational states that constitute normative judgments are related to one another in very particular, systematic ways. However, if Blackburn is right that judging that something is 'right' can, for different people and at different times, consists in motivations to perform any number of different actions then this would not be the case. If, at some times, 'Volunteering is right', does not consist in a motivation to volunteer, but instead consists in, say, a motivation to encourage volunteering then it will be difficult to explain why, 'Volunteering is right', is logically entailed by, 'Giving to charity is right', and, 'If giving to charity is right then volunteering is right'. If the premises of this *modus ponens* express, say, a desire to give to charity and a desire to volunteer conditional on desiring to give to charity, but the conclusion expresses a desire merely to *encourage* volunteering then argument is not valid, since desiring to do something does not commit me to also desiring to *encourage* doing it.

Solving the Frege-Geach problem also requires showing that if I say that something is right and you say that the same thing is right then we agree with each other. However, if Blackburn is right that one and the same normative judgement can motivate different actions then it becomes difficult to see how this could be. If it could be the case that my judgement that giving to charity is right is constituted by a motivation to perform a particular action whereas your judgement that giving to charity is right is constituted by a motivation to perform some other action, it is hard to see in virtue of what we would count as agreeing that giving to charity is right. The semantic content of our judgements would not in any way match. We know that Blackburn thinks that the conative attitudes that constitute moral judgements are fundamentally to be understood in terms of their mind-to-world direction of fit.[305] This means that normative judgements have the same

---

[305] e.g. Blackburn (1993, p.185)

content only if they both consist in a disposition to change the world in the same particular way.[306] In other words, two agents who make the same normative judgement must have the motivation to perform the same kind of action if they are to count as agreeing with one another on a non-cognitivist semantics.

All of this is a way of saying that since non-cognitivists identify the semantic contents of our normative judgements with motivational states, they cannot afford to be vague or evasive when it comes to identifying which actions agents are motivated to perform when they make normative judgments. Motivational states need to stand in very particular, systematic relations to one another if non-cognitivist semantics are to get off the ground, but if this is the case then the actions I am necessarily motivated to perform if I make a normative judgement should not be able to vary day to day in the way that Blackburn imagines.[307]

§9.3

A different way to think about this issue is along the lines that Allan Gibbard suggests in *Thinking How to Live*.[308] His idea is that when we judge other people's actions, we judge what we should do if we were in their place. More precisely, he says that when we judge that "the thing to do" for some person is to perform a particular action then we *plan* to perform that action for the contingency of being that person in their circumstances. Similar claims will be true, *mutatis mutandis,* for other normative terms too. When Dan judges that it would be right for John to pick up his kids, he plans to pick up John's kids for the contingency of being John. Having a plan is a motivational state on this view because it is a kind of *intention*. Gibbard says that thinking that something is the thing to do is to decide to do that thing.

---

[306] This is a potentially important disanalogy between conative and affective attitudes like love, since affective attitudes, we might think, cannot be understood simply in terms of the actions that they dispose us to perform. Scarantino and De Sousa (2018) provide a good overview of these issues.

[307] I think that Gibbard (1990) plausibly suffers from a similar problem.

[308] Gibbard (2003)

I think that Gibbard offers a more plausible and general account than Blackburn of what actions the person making a moral judgement will be motivated to perform if internalism is true, but it faces problems of its own. Particularly, it seems to bring the connection between normative judgements and action too close. Gibbard's position borders on the kinds of *strong internalism* which I wrote off at the start of this chapter. He argues that when I plan to do something then, unless I change my mind, I will do that thing when the time comes. Gibbard writes:

> "For a crucial sense of 'ought' , I say, the following holds: if you do accept, in every relevant aspect of your mind, that you ought right now to defy the bully, then, you will do it if you can."[309]

Since Gibbard thinks that normative judgements are *plans,* he is committed to saying that if I judge that I ought, *all things considered*, to do something then I will do it. However, it seems perfectly possible for someone to believe that they ought to defy the bully and not defy the bully. Gibbard argues that in these cases people are necessarily of two minds as to what we ought to do. However, this seems false. I could be wholly decided that I ought to defy the bully and not do it. Gibbard could say that the fact I do not defy the bully suggests that I believe that there are some *pro tanto* reasons not to do it, e.g. it might lead to me being hurt. However, this is not the same thing as being in two minds concerning whether, *all things considered*, I ought to do it. I can be wholly certain that some action is what I ought, *all things considered*, to do while also acknowledging that there are some weighty *pro-tanto* reasons not to do it.[310]

A second problem is that it just does not seem plausible that all normative judgements involve forming plans to act in some way. We can judge that some action is wrong even if we are absolutely certain that we will never be faced with that situation. If I judge that it was wrong for Brutus to betray Caesar then, according to Gibbard, I plan to refrain from betraying Caesar for the contingency of being Brutus. This seems a bit farfetched. I have no reason to plan for the contingency of being Brutus, I know that I will never be a Roman politician existing more than

---

[309] Gibbard (2003, p.153)
[310] Scanlon (2006).

two millennia ago. The range of circumstances I could possibly be in is an extremely small sliver of the total range of possible scenarios that I can make normative judgements concerning.

Gibbard argues that in many cases, we make plans in order to "practice" for similar situations we might face. While I might know that I will never be in Brutus' position, I could well end up in position which shares some of its characteristics, e.g. I might be in a position to betray my tyrannical boss. Gibbard thinks that I will plan for being in Brutus' position so that I can then apply that plan, *mutatis mutandis*, to similar situations I come across.

This explanation has a ring of plausibility but it is not fully satisfying. First, the range of situations of which I can make normative judgements is very wide indeed, including situations that are so dissimilar to situations I will face that it begins to look strained to suggest that I am engaging with these scenarios because it is instrumentally useful for my future self. Jerry Fodor's sardonic response to Steven Pinker seems appropriate here:

> "What if it turns out that, having just used the ring that I got by kidnapping a dwarf to pay off the giants who built me my new castle, I should discover that it is the very ring that I need in order to continue to be immortal and rule the world? It's important to think out the options betimes, because a thing like that could happen to anyone and you can never have too much insurance."[311]

Second, there are potential situations for which we have no need of normative 'practice' but about which we can nonetheless make normative judgements. If asked whether the Holocaust was wrong I will of course judge that it was. There might not seem much point in my planning for the contingency of being a Nazi in Germany circa WW2 because, aside from anything else, time travel to the past is impossible. Gibbard will presumably answer that here I am getting in 'practice' for situations in which I might commit genocide. The thing is, I do not need any 'practice' when it comes to planning whether to commit genocide. I was already fully against committing genocide before I was asked whether the Holocaust was wrong. When asked whether the Holocaust was wrong it does not seem plausible to me that I am planning, for the

---

[311] Fodor (1998b)

contingency of being a Nazi in Germany circa WW2, so as to apply any lessons I may have learned to future situations.

Non-cognitivists are in something of a bind. They need to argue that the motivational states that (they say) constitute our normative judgements are related to each other in precise and systematic ways in order to have a plausible semantic account, but succeeding in this enterprise commits them to an implausible and contrived picture of the necessary relationship between our normative judgements and our motivation. As we have seen the non-cognitivist has no easy way of explaining which motivational states we are necessarily in when we make normative judgements. If even the non-cognitivist cannot explain this then that should lead us to doubt that normative judgements are accompanied necessarily by motivational states. At the very least, if capturing internalism requires saying that the link between normative judgement and motivation is as the non-cognitivists are committed to saying that it is, then it does not seem very costly to the naturalist to deny internalism.

§10 - Conclusion

In this chapter I have circumvented the traditional internalist-externalist debate – which has reached an impasse – and argued against internalism in a roundabout way. I accepted that, since there are no necessary links between distinct existences, if internalism is true then normative judgements must be (at least partially) constituted by a motivational state. I then argued that internalism could not be true because normative judgements cannot be constituted by motivational states. The strength of normative judgements can change while the strength of the normative motivation to act on those judgement does not change, so normative judgements cannot be constituted by the normative motivation we have to act on them.

Additionally, there is not a plausible account of which motivational state we are necessarily in when we make normative judgements that covers all the kinds of normative judgements we can

make. Or, at least, the necessary link that must exist between normative judgements and motivation if internalism is true, turns out to be so strange and contrived that it is not at all costly to naturalists if they cannot accommodate it.

<u>Chapter 6: The Moral Twin Earth Argument</u>

It has been argued that naturalists struggle to accommodate certain kinds of normative disagreement. The most influential of these arguments come from Terry Horgan and Mark Timmons.[312] They argue that their Moral Twin Earth thought experiments show that naturalists cannot explain why we morally disagree with members of certain hypothetical extra-terrestrial communities. In this chapter, I argue that Moral Twin Earth scenarios present problems for most metaethical views. That makes more palatable the responses to the Moral Twin Earth arguments that I go on to develop. First, I suggest that we could widen our understanding of 'disagreement' in such a way that naturalists could claim that there is disagreement going on in Moral Twin Earth scenarios. Second, I argue that we can debunk the intuitions that Moral Twin Earth scenarios are supposed to pump. I start by setting up the Moral Twin Earth arguments.

<u>§1</u>

Take a form of naturalism which says that the moral property *right* is identical to the natural property *maximises pleasure*. Now suppose two people disagree about whether voting for the Labour party is good. These two people might have conflicting beliefs about whether voting for the Labour party maximises pleasure, in which case their moral disagreement is easily explicable. Equally though, they might both think that voting for the Labour party maximises pleasure but still disagree about whether or not it is right. How is such a form of naturalism to accommodate this?

In chapter 1, I argued that a normative term could refer to a natural property even if it was not obvious to competent language users that that term refers to that property. If that argument was sound, then the problem of explaining why, if *right* is identical to *maximises pleasure*, two people can have a moral disagreement, against a background of broad agreement concerning the natural

---

[312] Horgan and Timmons (1991, 1992a, 1992b, 1996, 2000, 2009, 2015), Timmons (1998).

facts, becomes tractable. If competence with 'right' does not require that I know that *right* is *maximises pleasure*, then competent moralisers can disagree about whether voting Labour is right while having the same views, or no views, about whether voting Labour maximises pleasure. Just as two people could seemingly disagree about whether some liquid is water without disagreeing about whether it is H_2O, so two people could disagree about whether something is right without disagreeing about whether it maximises pleasure.

One way to flesh out this kind of move is with a causal semantics. According to what Terry Horgan and Mark Timmons call "Boyd/Brink view" - a position which combines elements of Richard Boyd's semantic theory with David Brink's - terms refer to the particular properties that they do because our use of those terms is *causally regulated* by those natural properties. [313] On this kind of account, 'water' refers to H_2O, because water's being composed of H_2O molecules causally regulates the sorts of things we say and do using the term 'water'. A property can causally regulate our use of a term without our being aware that it does. H_2O, for instance, guided our use of 'water' before the discovery that water is composed of H_2O molecules. Because of this, 'water' can refer to H_2O even though competent speakers could doubt or disagree with claims like, 'water is H_2O'.

Similar considerations apply in the moral case, argue Boyd and Brink. Suppose that the use of a term like 'right' is, by and large, causally regulated by the natural property *maximises pleasure*. In that case, 'right' refers to the property *maximises pleasure*. Competent language users need not know which properties causally regulate our use of 'right', which explains why it is possible to disagree with claims like, 'Maximising pleasure is right', even if *right* is identical to *maximises pleasure*.

---

[313] Especially, Boyd (1988) and Brink (1989).

Terry Horgan and Mark Timmons cast doubt on this suggestion with their Moral Twin Earth thought experiment.[314] In Hilary Putnam's original Twin Earth thought experiment, we are asked to imagine a hypothetical planet Twin Earth which is just like Earth except that the water-like liquid that fills the lakes and rivers on Twin Earth is not composed of H_2O, but is instead composed of some other chemical compound called 'XYZ'. Other than that fact, Twin Earth is as close to Earth as possible; all the plants, animals, hills, oceans, buildings, etc. on Earth have exact duplicates on Twin Earth. The inhabitants of this planet, the Moral Twin Earthlings, have a language which is very like English, Twin English, containing a term 'water*' which they use when talking about the liquid that fills their lakes and rivers. If they were to say something like 'Water* is XYZ' and we were to say 'Water is not XYZ' then, despite surface appearances, we would not really be disagreeing. 'Water*' denotes *XYZ*, whereas 'Water' denotes *H₂O*, so we would really just talking past one another.[315]

The same does not apply in the moral case, argue Horgan and Timmons. We can imagine *Moral Twin Earth* which, again, is just like Earth in almost all respects. On this planet the humanoid inhabitants have terms like 'right*' and 'good*', which have the same kind of 'practical roles' that the English terms 'right' and 'good' respectively have. For instance, Moral Twin Earthlings typically apply 'right*' to actions in order to convince others to perform those actions. Generally, they feel some shame, guilt or regret if they fail to perform actions they judge to be 'right*'. They also censure and punish those who do not do what is deemed 'right*'. Most of the philosophers in the community might believe that if some potential action can be accurately called 'right*' then it must be possible to perform that action. In other words, they are largely disposed to use 'right*' in the kinds of ways and for the same reasons that we use our term 'right'. When I talk about the 'practical role' of a normative term in what follows I will be referring to these kinds of features of a normative term's usage.

---

[314] Particularly in Horgan and Timmons (1991, 1992a, 1992b)
[315] Putnam (1973, 1975)

The key difference between English and Twin-English is that the Twin-English term 'right*' is causally regulated by a different natural property to the property that causally regulates of our use of 'right'. 'Right', Horgan and Timmons assume for the sake of argument, is causally regulated by *maximises pleasure*, whereas 'right*' is causally regulated by the property had by those actions that Kantians on Earth would describe as 'right'. Following Horgan and Timmons, let us call the property that causally regulate the use of 'right*', the *Kantian property*.

Next, Horgan and Timmons ask whether our moral terms, 'good', 'bad', 'right', 'wrong' have the same meaning as the Twin English terms 'good*', 'bad*', 'right*', 'wrong*', and they submit that it is intuitively very plausible that they do. Suppose that we and the Moral Twin Earthlings agree that an act of sacrificing one to save five has the property that causally regulates our use of 'good', but that it does not have the property that causally regulates the use of the Twin English term 'good*'. Now imagine that we said to a Twin Earthling, 'Sacrificing one to save five is right', and they say, 'Sacrificing one to save five is not *right'. Intuitively, Horgan and Timmons claim, in making these claims we would count as disagreeing with one another, which shows that 'right' and 'right*' must have the same meaning.

According to the Boyd/Brink view, however, the English term 'right' picks out a different property, and so has a different meaning, to the Twin English term 'right*'. In that case, the Boyd/Brink view is committed to saying, falsely, that we are merely talking past the twin-earthlings. Additionally, if the Boyd/Brink view were correct then it should seem clear to both parties that what the other has said is true. But intuitively we would not be inclined to think that what the other had said was true. So, the Boyd/Brink view must be wrong about the meaning of moral terms.

This is not just a problem for the Boyd/Brink view, it is argued. Horgan and Timmons, as well as several other philosophers think that this problem, or a problem very much along the same lines, applies to all naturalist positions.[316] For instance, according to Frank Jackson's view a term like 'right' refers to the descriptive property that, were it the referent of 'right', would make as many mature folk platitudes surrounding the use of the word 'right' come out as true as possible.[317] Another way to put this is to say that, 'right' refers to whichever property fills the 'rightness-role'. Mature folk platitudes are the claims that we would regard as platitudinous in certain ideal conditions. Suppose that there is a mature folk platitude that says that attacking a passer-by with a sword in order to test its sharpness is not right.[318] If 'right' referred to *maximises pleasure* then that platitude would come out as true. If more of these kinds of mature folk platitudes surrounding 'right' come out as true on the supposition that 'right' refers to *maximises pleasure* than on the supposition that 'right' refers to some other property, then 'right' refers to *maximises pleasure* on this theory.

A crucial detail for this theory is how a naturalist can spell out what makes mature folk morality 'ideal'. Jackson is a bit quiet on this issue but there has been plenty of discussion of it in the literature; possessing full non-moral information, having undergone a full course of cognitive psychotherapy, having full imagine acquaintance, being fully coherent and being behind a 'veil of ignorance' are just some of the criteria that have been suggested.[319] But however the naturalist understands 'ideal conditions', we can imagine that the property that would make true the Moral Twin Earthling's mature folk platitudes surrounding the use of their term 'right*', is different to the natural property that makes true our mature folk platitudes surrounding 'right', or, put another way, the property that plays the rightness*-role is different to the property that plays the

---

[316] Horgan and Timmons (2000, p.149). Ridge (2014), Streumer (2017) and Enoch (2011) are three philosophers who think that MTE arguments present serious problems for naturalists.
[317] Jackson (1997). Horgan and Timmons' talk about Frank Jackson's view specifically in their (2009).
[318] The example is adapted from Midgley (1981)
[319] Firth (1952), Lewis (1989), Brink (1989), Merli (2003), Rawls (1971).

rightness-role. Suppose that *maximises pleasure* plays the rightness-role while the *Kantian property* plays the rightness* role. If that were the case then Jackson's theory would have it that we would be merely talking past the Moral Twin Earthlings if we were to say something like, 'Killing one to save five is right', and they were to say, 'Killing one to save five is not right*'. Intuitively, however, we would be disagreeing with one another, or so it is alleged. So, Jackson's semantics is also found wanting.

Horgan and Timmons think that this kind of problem generalises.[320] For any naturalist account, we can imagine a community of Moral Twin Earthlings of whom it is the case that (i) they have terms with a similar 'practical role' to our moral terms; (ii) intuitively, they can enter into moral disagreements with us using claims in their language containing those terms, and (iii) according to the naturalist position under examination the referents of those terms would be different to the referents our moral terms. Given (iii), the naturalist cannot account for (ii), so naturalist accounts of moral terms must be incorrect.[321]

The problem is a particularly acute one for naturalists, it is argued, because other metaethical accounts can easily explain why we seem to disagree with the Moral Twin Earthlings. In the next part of this chapter, I show that this claim can be challenged; all major metaethical theories have problems providing a satisfactory explanation of why there appears to be a disagreement going on in these kinds of cases. After that, I survey two ways we might solve the problem. First, we might argue that we disagree with Moral Twin Earthlings even though the semantic content of

---

[320] Horgan and Timmons (2000).

[321] This is a reconstruction of, what Laurence, Margolis and Dawson (1999) call, Horgan and Timmons' "direct argument" against naturalism. Horgan and Timmons also use the Moral Twin-Earth thought experiment to try to resurrect the open question argument Moore (1993), Horgan and Timmons (1992a)) and John Mackie's queerness argument (Mackie 1977, Horgan and Timmons (1992b)). Though these arguments are in many ways similar to each other it is nevertheless important to treat them separately since what refutes one argument will not necessary refute all three. For instance, I addressed Horgan and Timmons' revival of the open question argumnent in Ch.1, but what I said there would not defeat their 'direct argument' also, so I feel the need to address it in this chapter. I do not have anything to say about their revival of Mackie's argument that has not already been said by Laurence, Margolis and Dawson (1999), so I pass over it.

what we say is consistent with the semantic content of what they say. Second, we could argue that there is nothing to explain; appearances are deceptive and in fact there is no disagreement taking place.

§2 - Why Explaining Moral Disagreement is a Shared Problem 1

I argued in chapter 1 that the open question argument was a problem for most metaethical theories. The open question argument affects all analyses of normative concepts, and most metaethical theories involve the analysis of some normative concepts. While non-naturalists, error theorists and non-cognitivists might not analyse normative concepts in natural terms, they do tend to analyse some normative concepts in terms of other normative concepts. For example, according to Parfit's theory when we call something good we mean, roughly, that something about it gives us a reason to respond in a certain positive way to it such as wanting it, preserving it, choosing it, etc.[322] Since it is possible to doubt claims like, 'Something is good iff something about it gives us a reason to respond to it positively', the open question argument also applies to these views. The same is true, *mutatis mutandis*, of the Moral Twin Earth argument.

Suppose that in ideal circumstances we would be disposed towards accepting a kind of value fundamentalism, like the one defended by G.E. Moore. Particularly, we would apply 'right' to actions only if we believe that they would produce the greatest amount of good in the universe. Now suppose that on Moral Twin Earth they have terms 'right*' and 'good*' which play very similar respective practical roles to our terms 'right' and 'good'. We can even imagine they are disposed to apply 'good*' in exactly the same circumstances that we would use 'good'. Let us suppose that both 'good' and 'good*' would be used to refer, in ideal circumstances, to things that cause pleasure. However, Moral Twin Earthlings, in ideal conditions, would apply their term 'right*' to some actions which they would not also be disposed to apply their expression

---

[322] Parfit (2011a, p.34)

'produces the greatest amount of good* in the universe' to. For instance, they could think that if an action involves killing someone then it cannot be 'right*' even if it produces the greatest amount of 'good*' in the universe. We, however, would say that whenever an action counts as 'right' it must be the case that 'produce the greatest amount of good' can also be applied to it.

Do we disagree with Moral Twin Earthlings about what is right in this case or are we merely talking past each other? I maintain that, intuitively, it just as plausible that we genuinely disagree here as it is in Horgan and Timmons' original Moral Twin Earth example. This is a problem for these views. 'Right', in the present example, refers to the property *produces the greatest amount of good*, whereas 'right*' cannot plausibly be said to. We should be merely talking past Moral Twin Earthlings, but (assuming the intuitions pumped by Moral Twin Earth examples are what Horgan and Timmons claim) intuitively this is not the case.

We can apply the same line of argument to those views that reduce evaluative concepts to reason concepts. According to T.M. Scanlon, to say that something is 'good' is to say that it has the property of having other properties which provide reasons to choose, prefer or recommend it.[323] Suppose that Scanlon is correct about what 'good' means. Now suppose that on Moral Twin Earth they have terms like 'good*' and 'reason*' which work much like 'good' and 'reason'. A difference however is that, even in ideal conditions, they would not think that something is 'good*' only if it has the property of having other properties which provide 'reasons*' to choose, prefer or recommend it. They would be swayed by arguments made by Twin-Roger Crisp and other Twin-philosophers, who say that we can sometimes have a 'reason*' to choose, prefer or recommend something that is not 'good*', and that we can sometimes lack a 'reason*' to choose, prefer or recommend something that is 'good*'.[324] Now suppose we were to say, 'If something is

---

[323] Scanlon (2000, pp. 96-97. 2014, p.2).
[324] Crisp (2000, 2005, 2008). D'Arms and Jacobson (2000a, 2000b). Rabinowicz and Rønnow-Rasmussen (2004).

good then we must have a reason to choose, prefer or recommend it', and they were to say, 'It is not the case that if something is good* then we must have a reason to choose prefer or recommend it'. Intuitively it is about as plausible that there is a disagreement going on here as it is in the original Moral Twin-Earth case. But the Moral Twin earthlings could not be using 'good*' to pick out the property *has properties which provide reasons to choose, prefer or recommend it*. So, we cannot explain why there seems to be disagreement in terms of our and their claims having inconsistent semantic contents.

We could apply the same line of argument, *mutatis mutandis*, to other views which hold that we can analyse normative concepts in terms of other normative concepts or reduce normative concepts to other normative concepts. For most mainstream views we consider of the relation between concepts of, say, reasons and obligation, value and desirability, wrongness and blameworthiness, etc. we will be able to imagine a Moral Twin Earth scenario on which it appears to be the case that they accept some other conflicting view. This means that, if Horgan and Timmons' argument is sound, then it would rule out most views of normative concepts. This fact should make us suspicious that the argument is a telling one.

§3 - Why Explaining Moral Disagreement is a Shared Problem 2

§3.1

I just showed that any views which analyse normative concepts (be they normative or non-normative analyses) will be potentially weak to Moral Twin Earth arguments. Most non-naturalist, error theorist and non-cognitivist theories involve at least some analyses of normative concepts, so naturalists are not uniquely vulnerable to these worries. However, it is possible for non-naturalists, error theorists and non-cognitivists to deny that there are any analytic or conceptually necessary links between normative concepts. To bolster my line of argument then I

show that non-naturalists, error theorists and non-cognitivists are vulnerable to Moral Twin Earth scenarios in other ways.

§3.2 Non-Naturalism (and Error Theory)

Non-naturalists and error theorists appear, at first blush, to have an easy time explaining why we seem to disagree with Moral Twin Earthlings. Non-naturalists and error theorists will say that both 'right' and 'right*' attempt to refer to the non-natural, *sui generis* property *right*.[325] Error theorists will think that neither term succeeds in referring to *right*, since they think that there is no such property, while non-naturalists think that 'right' or 'right*' could successfully refer to *right*. In either case, 'X is right', and, 'X is not right*', will express inconsistent propositions. In what follows I will focus on just the non-naturalist, but what I say for them goes for the error theorist too.

The non-naturalist is not entitled to simply help themselves to the claim that 'right' and 'right*' refer to the same property, they must provide an explanation of it. It clearly would not be acceptable for the naturalist to refuse to give any theory of reference-determination, and simply insist that 'right' and 'right*' refer to the same natural property, in order to rebut Horgan and Timmons' challenge. Likewise, the non-naturalist cannot just insist that 'right' and 'right*' must refer to the same non-natural property. In the Moral Twin Earth thought experiment, in ideal circumstances, Earthlings apply 'right' to actions which maximise pleasure, whereas in the same ideal conditions Twin Earthlings apply 'right*' to actions which have the Kantian property. The non-naturalist needs to tell us what makes it the case that 'right*' denotes the same thing that 'right' does, and not some other property, like, for instance, the *Kantian property*.

---

[325] Non-naturalists and error theorists may say instead that 'right' and 'right*' denotes some other normative property that *rightness* reduces to such as the property *is what we have most reason to do*. In that case, we could then reapply the same argument to whatever they take to be the most fundamental normative properties that normative terms refer to.

This will be difficult to do because the hypothesis that 'right' and 'right*' denote the same property is very uncharitable. If in ideal circumstances we and Moral Twin earthlings denote *right* with 'right' and 'right*', but we apply those terms to different actions, then either we make a lot of false claims containing 'right' or they make a lot of false claims containing 'right*'. Assuming that we are reasonably confident in the truth of our own moral judgements, if they denote *is right* with 'right*' then many of their claims containing 'right*' would come out as false. This contravenes a commonly accepted principle of interpretation, which roughly has it that we ought, *ceteris paribus*, to translate terms such that sentences in which those terms are embedded come out true.[326] The hypothesis that Moral Twin Earthlings use 'right*' to denote the *Kantian property* conforms perfectly with this principle, unlike the hypothesis that 'right*' denotes the same non-natural property that 'right' does. Since the non-naturalist's translation of the term 'right*' contravenes a basic principle of interpretation they need to provide us with a strong countervailing reason to accept it.

Non-naturalists could argue that the similarities in 'practical role' show that we and Moral Twin Earthlings are trying to denote the same property. In the thought experiment both Moral Twin Earthlings are disposed to use 'right*' in the same kinds of ways that we use 'right', e.g. they are motivated to do what they judge to be 'right*', they reliably feel guilty for failing to do what they think is 'right*', they shame others if they do not do what is deemed 'right*', etc.

However, I think that this explanation fits uncomfortably with other parts of the non-naturalist account. Non-naturalists typically believe that all that is required to judge that some action is, say, right is to have the belief that that action has a particular *sui generis* property. So, on the non-naturalist account, someone can believe that some action is right without necessarily being motivated to do it, without feeling any guilt for failing to do it and without satisfying the other criteria that non-cognitivists take to be necessary for judging that something is right. In fact, it is

---

[326] The *loci classici* for discussions of this principle are Quine (1960) and Davidson (1973).

consistent with the non-naturalist story that I might believe that something is right while being reliably motivated to *refrain* from doing it, if I am (substantively) irrational enough. I might even believe that something is right and consistently praise people for refraining from doing it if I am an *immoralist*.[327] All that being so, the fact that the Moral Twin Earthling term 'right*' has a particular 'practical role' tells us nothing about its denotation if the rest of the non-naturalist story is correct. For all the non-naturalist tells us, Moral Twin Earthlings could be *immoralists* who use their term 'right*' to denote the *sui generis* property *is wrong* and 'wrong*' to denote the property *is right*. On this hypothesis, Moral Twin Earthlings shame people for failing to do what they believe is wrong, and they are usually motivated to do what they believe is forbidden because it is wrong. The non-naturalist lacks grounds for saying that Moral Twin Earthlings use 'right*' to denote *is right* rather than *is wrong*, or any other property, based on the 'practical role' that 'right*' has.

A caveat to that is that non-naturalists can adopt some kinds of motivational internalism. For example, they can say that anyone who judges that something is right is necessarily motivated to do it *if they are fully (substantively) rational*.[328] However, that does not affect the point I am making. We have no guarantee that the Moral Twin Earthlings are fully substantively rational, so the fact that they are typically motivated to do whatever they describe as 'right*', would not show us that it must have the same meaning as 'right', even if the kinds of internalism that are compatible with non-naturalism were true.[329]

Non-naturalists might point out that in many paradigmatic cases 'right' and 'right*' will be applied to the same actions in order to ground the claim that 'right' and 'right*' must refer to the same property.[330] Arguably, utilitarians and Kantians will largely see eye-to-eye when it comes to

---

[327] Dreier (1990), Lenman (2003b).

[328] I discussed this in the last chapter also.

[329] Because of this, I do not think that non-naturalists can use a Ralph Wedgwood-style conceptual role semantics to explain why 'right*' and 'right' denote the same property.

[330] Thank you to James Lenman for this suggestion.

paradigmatic moral cases. They will both think, for instance, that killing is usually not right and that helping someone in need usually is. Since the Moral Twin Earthlings describe as 'right*' those actions which a Kantian would describe as 'right', 'right' and 'right*' will arguably both be applied to many of the same actions. The non-naturalist might try to use this fact to argue that 'right' and 'right*' must have the same referent.

However, I think that there are three problems with this move. Firstly, it is not clear to me that utilitarians and Kantians do in fact agree all that much. For example, utilitarians think that every single action that someone could perform other than the ones that maximises utility are morally wrong. Presumably, that will mean that, even in paradigm cases, for many of the possible actions that utilitarians will call 'wrong', the Kantian will call them 'permissible'. And, for many of the actions that a utilitarian regards as 'obligatory', the Kantian is going to call it either 'wrong' or 'superogatory'. If the Twin Earthling use of 'right*' mirrors the Kantians use of 'right' then, we should not assume that 'right' and 'right*' will be applied in the same way even in paradigm cases.

Second, it is arguably not crucial that Moral Twin Earthlings do apply 'right*' to the same actions we apply 'right' to, in order to generate the intuition that we disagree with them. Years before Horgan and Timmons introduced Moral Twin Earth, R.M. Hare pumped more or less the same intuitions with his Missionary and the Cannibals example. Hare imagines that we encounter a community of cannibals who have a word 'good*' which they use to refer to the taking of scalps.[331] Naturalist theories, Hare argued, would have it that we would not really disagree with the cannibals if we were to say, 'Scalp hunting is not good', and they were to say, 'Scalp hunting is good*', rather, each of us would be saying something true in our own language. But, intuitively we seem to disagree with the cannibals about what is good, or so it is claimed.

---

[331] Hare, (1952, ch.5). Horgan and Timmons (1996) and Timmons (1998) acknowledge the similarity between their argument and Hare's, but they deny that they were influenced by it.

If those intuitions count against the naturalist then, the non-naturalist also has a problem. Even if utilitarians and Kantians will agree about the moral status of many actions (e.g. Killing, torturing and theft are usually wrong), the same cannot be said of us and the cannibals. Since 'good' and 'good*' will be applied to such different kinds of actions, the non-naturalists will have a hard time convincing us that both terms must refer to the *sui generis* property *good*.

Finally, as we just saw, for all the non-naturalist says, Moral Twin Earthlings could be *immoralists* who use 'right*' to denote *is wrong*. If for every action that we judge to be right, they judge it to be wrong, and vice versa, then they could apply 'right*' to exactly the same actions that we apply 'right' to, without 'right' and 'right*' having the same referent. That being so, the non-naturalist needs some reason other than the fact that 'right' and 'right*' have the same application in paradigm cases to justify their claim that 'right' and 'right*' both refer to the non-natural property *right*.

I cannot rule out that the non-naturalist and error theorist could come up with some other explanation of why 'right' and 'right*', or 'good' and 'good*' must refer to the same *sui generis* moral property, but it is hard to see what it could be. Until they provide one, the non-naturalist has no better explanation of why we intuitively seem to disagree with Moral Twin Earthlings than the naturalist does.

§3.3 Non-Cognitivism

At first blush, non-cognitivists seem to fare better than non-naturalists. Since non-cognitivists think that the semantic contents of normative terms are, in a sense, determined by the 'practical role' of those terms, they can give a more plausible-seeming explanation of why 'right' and 'right*' must have the same semantic content.[332] If it is the practical role of normative terms to express desires, and if the content of normative terms is determined by their practical role, then

---

[332] Gibbard (1992) gives a particularly good discussion of this.

since 'right' and 'right*' have the same kind of practical roles they will have the same semantic

content. If the similarity in the practical roles of the terms 'right' and 'right*' suffices for their

having the same semantic content then we and the Moral Twin Earthlings would disagree if we

were to say, 'Sacrificing one to save five is right', and they were to say, 'Sacrificing one to save

five is not right*'. This is Horgan and Timmons' position.

However, this explanation has implausible consequences. Consider the following example:

> Moral Twin Earth 2
>
> Moral Twin Earth 2 is in most respects just like Earth and Twin-Earth. On Moral Twin Earth 2, however,
> the humanoid inhabitants have a term 'right**' which they apply to actions only if those actions result in
> hedgehogs being looked at in the light of the moon.[333] They do not call actions which result in hedgehogs
> being looked at in the light of the moon 'right**' because they believe that looking at hedgehogs in the
> light of the moon has certain effects, e.g. they don't call actions 'right**' because they think that hedgehog-
> watching is fun or that it improves health. They say that actions which result in hedgehogs being looked at
> in the light of the moon can be described as 'right**', simply because it is 'right**' in and of itself for
> hedgehogs to be looked at in the light of the moon and that no further explanation can be given. At the
> same time, 'right**' has the same kind of 'practical role' in the lives of the inhabitants of Moral Twin Earth
> 2 that 'right' and 'right*' has in the lives of Earthlings and Moral Twin earthlings respectively. For instance,
> they are typically motivated to do what they deem 'right**' and they shame others who fail to do actions
> that they deem 'right**'.

I submit that when we say, 'Giving to charity is right', and the inhabitants of Moral Twin Earth 2

say, 'Giving to charity is not right**', we are not intuitively disagreeing about what is morally

right.[334] The inhabitants of Moral Twin Earth 2 are too different and too strange to really count

as engaging in substantive moral discussion with us or being within the same moral community

as us. In other words, the inhabitants of Moral Twin Earth 2 simply do not seem to count as

having any interest in what is morally right at all and they really are talking past us. However,

according to the non-cognitivists account we have considered so far we disagree with the

inhabitants of Moral Twin Earth 2 because both 'right' and 'right**' have the same practical role.

Since sameness of practical role does not mean that we and the inhabitants of Moral Twin Earth

---

[333] This example is taken from Foot (1979, p.102).
[334] Rubin (2014).

2 disagree, it cannot explain why we disagree with the inhabitants of Moral Twin Earth either. So, the simple non-cognitivist explanation of why we seem to disagree with Moral Twin Earthlings does not work either.

One way to get around the problem would be for the non-cognitivists to say that the semantic content of normative terms is determined by more than simply 'practical role'. They could, for instance, agree with Philippa Foot that a term can only be a moral term if that term is governed by standards that regulate harms and benefits.[335] On this view, when we say that some action is 'right' we primarily express a desire to perform that action, but we also express a background belief that that action is appropriately related to the prevention of harms or the promotion of benefits. This kind of strategy could be used to get round cases like Moral Twin Earth 2. Since the use of 'right**' is not governed by standards that regulate harms and benefits it does not count as a moral term, so this non-cognitivist account is not forced to say that 'right' and 'right**' have the same content, nor that we and inhabitants of Moral Twin Earth 2 count as being in a moral disagreement.

At the same time, the descriptive constraints that are placed on moral terms are relaxed enough that non-cognitivists can still explain why we and the inhabitants of the Moral Twin Earth disagree with another in the original thought experiment. Both 'right' and 'right*' are plausibly constrained by standards that regulate harms and benefits. The actions these terms are used to describe may promote different benefits or promote the same benefits to different degrees but the constraints on what counts as a moral term are minimal enough that these differences are not sufficient to show that they are not both moral terms. That being so, the non=cognitivist can say that, given the sameness of their 'practical role', and given that they are both meet the same

---

[335] This is the strategy that Rubin (2014) pursues. See also Hare (1952) and Gibbard (1992).

minimal descriptive constraints, 'right' and 'right*' have the same meaning, and so they can explain why we and the original Moral Twin Earthlings disagree.

However, I think that this more nuanced non-cognitivist account is vulnerable to further examples like the following:

> Moral Twin Earth 3
>
> Moral Twin Earth 3 is in most respects just like Earth, Moral Twin Earth and Moral Twin Earth 2. However, the humanoid inhabitants of Moral Twin Earth 3 have a term 'right***' which has the same kind of 'practical role' as 'right', but whose use is governed by some natural property other than the one that governs our use of 'right'. One important difference is that 'right***' is not governed by standards that regulate harms and benefits, instead it is governed merely by standards that regulate harms. Inhabitants of Moral Twin Earth 3 do not care about benefits, except insofar as they are privations of harm. They really want to minimise harm, but are indifferent to benefiting others beyond the point of their being free from harm.

Suppose that we were to say, 'Giving birthday cards to friends is right', and inhabitants of Moral Twin Earth 3 say, 'It is not the case that Giving birthday cards to friends is right***'. In my view, it is about as plausible that there is a disagreement going on about what is right as it was in the original Moral Twin Earth case. This is a problem for the more nuanced kind of non-cognitivism I am considering because 'right***' should not come out as a genuine moral term, since its use is not governed by standards that regulate harms *and* benefits, only harms. By adding even minimal descriptive content to their account of normative judgement to get round examples like Moral Twin Earth 2, non-cognitivists end up being vulnerable to the same kinds of worries that affect ordinary descriptivist accounts.

That might seem to merely suggest that the kinds of descriptive constraints that Foot suggested for moral terms are too demanding. A non-cognitivist might argue that the semantic constraints should be much lesser than this, such that inhabitants of Moral Twin Earth 3 do count as using moral terminology but inhabitants of Moral Twin Earth 2 still do not. To settle the matter we would need to hear in more detail what the descriptive constraints are on moral terms, but I

suspect that similar problems will arise for any constraints we suggest. Whatever the constraints on moral terms like 'right', we will be able to imagine another community who have a term 'right****' which has the same practical role as 'right' but *slightly different* descriptive constraints, and with whom we can intuitively disagree about what is right. Since 'right****' will not meet the descriptive constraints necessary for counting as a moral term, it should not have the same meaning as 'right', so we will be at loss for how to explain why we intuitively disagree when we say 'X is right' and they say 'X is not right****'.

So, while I think that non-cognitivists can get round problems arising from scenarios like Moral Twin Earth 2 by adding descriptive content to their account of normative terms, they are then unable to explain why we intuitively disagree with some hypothetical speakers who have terms with the same practical role as our normative terms but slightly different descriptive content, as in *Moral Twin Earth 3*.

The best way to explain cases like Moral Twin Earth 2 then, may be for the non-cognitivist to simply bite the bullet. This seems costly to me (though admittedly not fatal). Intuitively, the inhabitants of that world do not have anything that could be considered a moral concept. It seems odd to think that they are trying to find out what is right and good and they have spectacularly missed the mark, rather than thinking that they are interested in other things.

Having established that there are some potential costs for any account arising from Moral Twin Earth-style scenarios I move on to sketch two ways that the naturalist might solve the problem.[336]

---

[336] It might be objected that my *partners in guilt* style argument is incomplete because I have not talked about, what Max Kölbel (2004) calls, 'genuine relativism'; a position motivated largely by its supposed ability to handle disagreement. I have not talked about this position in the main text because I do not think that relativist explanation of moral disagreement even gets off the ground. I will say why briefly now.

Genuine relativists argue that moral claims express propositions which are never absolutely true but only ever true *relative to a circumstance of evaluation*. According to one form of genuine relativism, when we, who are utilitarians, say, 'An action is right iff it maximises pleasure', we express the proposition - *An action is right iff it maximises pleasure* – which is not simply true or false, but is true *relative to their own normative outlook* and false relative to, the Twin

§4 – Non-Content Based Disagreement

Naturalism entails that the semantic contents of our moral claims like, 'An action is right iff it maximises happiness', are not inconsistent with the semantic contents of Moral Twin Earthling claims like, 'It is not the case that an action is right* iff it maximises happiness'. This makes it difficult to explain why we intuitively seem to be disagreeing with Moral Twin Earthlings. One way to solve the problem would be to argue that we disagree with Moral Twin Earthlings despite the fact that the semantic contents of our respective claims are not inconsistent. Let us call cases where we disagree, despite the fact that the semantic contents of our utterances could both be true, *non-content based disagreements*.

There are plausibly many examples of *non-content based disagreement*. Some of the most discussed cases of potential non-content based disagreement concern taste. Many people think that two people can disagree about what is, say, tasty or funny without either of those people necessarily saying something that is false.[337] If I say that aubergines are tasty and you say that they are not then we seem to disagree, but, so goes the thought, neither of us necessarily say something that has false semantic content.

Earthlings' outlook. Relativists think that they can explain easily why we disagree in these cases since we have inconsistent beliefs about the proposition - *An action is right iff it maximises pleasure*; we think that it is true and Moral Twin Earthlings think that it is false.

However, I do not think that this account really explains why we and Twin-Earthlings disagree, rather, it just presents the superficial appearance of explaining disagreement. Either, we and they disagree about whether one and the same proposition is absolutely true, or we and they disagree about whether that proposition is true relative to a particular circumstance of evaluation. The first option is incompatible with relativism since they think that moral propositions are never absolutely true or false. The second option does not explain the disagreement between us and Twin-Earthlings, since we might agree entirely about which circumstances of evaluation the proposition - *an action is right iff it maximises pleasure* – is *true relative to*, while still disagreeing. We and they might agree that it is true relative to a utilitarian outlook but false relative to a Kantian outlook that an action is right iff it maximises pleasure, but disagree about whether an action is right iff it maximises pleasure, so whence the disagreement? (James Dreier makes the same kind of point well in his (2009)).

To his credit, John MacFarlane, a genuine relativist, recognises that the simple relativist account of disagreement does not work (MacFarlane, 2007). The account of disagreement he offers instead is along the quasi-expressivist lines that I look at shortly. Since that kind of explanation is open to naturalists I do not need to talk about genuine relativism any more.

[337] The literature on this topic is large; see the references in Zeman (forthcoming).

We do not, however, need to commit ourselves to any position on these controversial topics in order to show that there are non-content based disagreements. Uncontroversial examples of this kinds of disagreement are the following:

> A: John has one leg.
> B: I disagree. John has both his legs.

> A: John said that Jane was from London.
> B: I disagree. Jane is from Leeds.

What A says in the first example has true semantic content iff John has one leg. That is not logically inconsistent with B's claim that John has two legs; in fact B's claim entails A's claim. However, B will assume based on what A has said, that A believes that John has *only* one leg. The reason, according to Paul Grice's classic account, is that it is assumed that when we converse with someone we try to communicate a sufficient amount of information given the context of the conversation.[338] It is more informative to say that John has two legs than to say that John has one leg, since John having two legs entails that John has one leg but not vice versa. Since B assumes that A wants to provide sufficient information with their utterance, B will assume that A does not believe that John has two legs, for if they did then they would have said so. Since B disagrees with the belief that John does not have two legs, B disagree with A, even though they believe that what A has said is strictly speaking true. The disagreement here is a result of the *pragmatics* – the practical purpose for which the speakers make their utterances – rather than the *semantics* – the literal meaning, or truth-value of the utterance – of what has been said.

The second example works in a similar way. An indirect speech report can function not just to convey the information that someone said something, but also to imply or suggest that the content of what that person said is correct. In a context where it will be taken for granted that a speaker will be reporting what someone else has said faithfully, expressions of disagreement can be understood as contesting the content of the utterance that is being reported. This is what we

---

[338] Grice (1989)

see in the second example. Here B disagrees with the content of what John is reporting to have said and not the claim that John said it.

Similarly, some naturalists argue, we can disagree with someone's normative claim where we disagree with some information or content that that claim communicates, which is distinct from the semantic content of that claim. This presents us with an opportunity to explain why we disagree with Moral Twin Earthlings when we say, 'X is right', and they say, 'X is not right*'. There are several different suggestions for what the disagreement in these cases concerns. I will briefly sketch three suggestions, without stopping to assess their relative merits.

According to one view, moral disagreements are often *metalinguistic disagreements* concerning what certain terms mean or how they are, or should be, applied.[339] The classic example of metalinguistic disagreement is the following:

> A: Secretariat is the greatest athlete of the 20th Century.
> B: I disagree, Secretariat is not an athlete.

The disagreement here is not about the Secretariat's achievements or physical prowess, rather, it concerns whether 'athlete' is the kind of term that can correctly be applied to horses as well as humans. Some naturalists want to say that moral disagreements are often just metalinguistic disagreements concerning the meaning of moral terms. Applied to Moral Twin-Earth, the thought would be that we disagree with Moral Twin Earthlings when we say, 'X is right' and they say 'X is not right*' because our claim expresses the judgement that 'right' and 'right*' are, or should be, used in such a way that they can both be applied to X.

According to another view, two people can count as disagreeing if their normative claims attempt to *update the conversational common ground* in inconsistent ways.[340] The conversational common ground is, roughly, the set of things which are taken for granted by participants to a

---

[339] Plunkett & Sundell (2013) and Silk (2016).
[340] Khoo and Knobe (2018). This is a common strategy when accounting for disagreements concerning epistemic modals, see e.g. Stalnaker (2014).

conversation. Joshua Khoo and Justin Knobe suggest that *norms*, as well as propositions or beliefs, are part of the conversational common ground. In other words, they think that participants to a conversation tacitly acknowledge certain rules concerning how they should or should not behave. When we make a normative claim like, 'Stealing is wrong', we try to update the conversational common ground so that it contains a norm like *Don't steal*. On this view, we would disagree with someone who says, 'Stealing is not wrong', because, by making that utterance, they are trying to update the common ground so that it contains a conflicting norm. We can apply this view to Moral Twin Earth too. Because 'right' and 'right*' have the same kind of practical roles, utterances containing them might be said to attempt to update the norms of the conversational common ground. If we were to talk with them and we were to say that something is 'right' and they were to say that the same thing is 'not right*', we will disagree with each other because we would be trying to update the norms of the conversational common ground in inconsistent ways.

Finally, we have the *quasi-expressivist view*.[341] This is an attempt by naturalists to co-opt the non-cognitivists account of moral disagreement. Very roughly, the idea is that someone who says, 'X is right' disagrees with someone who says, 'X is not right' because their claims express *inconsistent attitudes*. According to non-cognitivists, 'X is right' expresses a kind of desire, which is inconsistent with the desire expressed by 'X is not right'. While naturalists disagree with non-cognitivists that the *semantic contents* of moral claims are desires, naturalists can say that moral claims express desires due to the *pragmatics* of moral discourse. There are multiple ways to spell this out, but one thought goes that part of the *function* of moral discourse is to influence people's behaviour and attitudes. For instance, if someone says, 'Stealing is wrong', then part of their reason for speaking is to convince others not to steal. Because of this, someone who makes a moral claim will be understood as expressing a desire that people act in such-and-such a way and

---

[341] Merli (2004), Björnsson and Finlay (2010), Björnsson and Almér (2009, 2011), Evers (2011), Finlay (2014, 2017).

if this desire is inconsistent with the desires of others, they will be in disagreement with them. Since we know that Moral Twin Earthling terms like 'right*' have the same 'practical roles' as terms like 'right' they will also express desires. Since Moral Twin Earthlings would, under ideal conditions, apply 'right*' differently to how we apply 'right', they will express desires which are inconsistent with our desires, so we will disagree with them.

Naturalists then have several options for how to explain why we can be in a moral disagreement with the inhabitants of Moral Twin Earth, even though the semantic contents of what we say is not inconsistent with the semantic content of what they say. I do not go into which is the best theory because they all share one very serious problem which I will talk about now.[342]

§4.1 – Disagreement and Truth-Assessments.

Since these naturalist accounts locate the source of our disagreement in the *pragmatics* and not the semantics of moral claims, they are all vulnerable to a powerful objection. They cannot explain why, when we morally disagree with someone, we will also be inclined to say that what they have said is false.[343] If, according to naturalist semantics, the Twin English term 'right*' picks out the Kantian-property then we should be disposed to say that Moral Twin Earthlings speak truly when they say things like, 'It is not the case that an action is right* iff it maximises utility'. But, assuming again that we are utilitarians, we would not do this. We would say that they have spoken falsely. It is not enough for the naturalist to locate any old source of conflict and use that to explain why we disagree with Moral Twin Earthlings, they need to also explain why we regard moral claims we disagree with as false.[344]

---

[342] I also do not need to discuss which is the best strategy because Finlay (2017) has already convincingly shown that the rivals to the quasi-expressivist theory are either clearly inferior to it or that they collapse into a form of it.

[343] Ridge (2014), MacFarlane (2007), Richard (2011). This objection would also apply to other naturalist accounts which have it that normative disagreements are non-content based.

[344] This problem, arguably, does not affect the ordinary non-cognitivist since they think that to say that a moral claim is true only ever to amounts expressing the same attitude that is the semantic content of that claim. See Dreier (1996) for why this tactic is difficult for the non-cognitivist to pull off.

Some naturalists have challenged this.[345] They point out that we can sometimes attribute falsity to someone's utterance even where the literal semantic content of what they have said is a true proposition[346], such as in the following cases:

> A: John has one leg.
> B: That's false, John two legs.
>
> A: John said that Jane was from London
> B: That's false, Jane is from Leeds.

The thought goes that our truth or falsity attributions can sometimes be used, not to assess the truth-value of the literal semantic content of what is said, but to signal a disagreement with something that is pragmatically expressed by the utterance. Where the pragmatic content, rather than the semantic content, is more salient in certain conversational contexts, phrases like 'That's false', will be understood as targeting that pragmatic content. The point of moral discourse is primarily about influencing action and aligning our attitudes, so our truth and falsity attributions to moral claims can be understood not as assessments of the literal truth or falsity of the proposition semantically expressed, but, rather, as a means of expressing pragmatically expressed content which conflicts with the pragmatically expressed content of the other party's utterance.[347]

Michael Ridge objects that even if this were the case, we should still be able to assess the literal truth of the semantic content of an utterance by using phrases like 'literally true' and 'strictly speaking false' or by simply clarifying what we mean.[348] Consider this example:

> A: John said that Jane was from London.
> B: What you said is literally true; that is indeed what John said. However, John was lying, Jane is not from London.

---

[345] Björnsson and Finlay (2010), Björnsson and Almér (2011), Evers (2011).
[346] Björnsson and Finlay (2010) p.19
[347] Björsson and Almér (2011), Evers (2011).
[348] Ridge (2014, ch.3)

If the naturalist semantics were correct then we should be able to do something similar in the moral case. We should be able to have discussions like the following:

> Moral Twin earthling: It is not the case that an action is right* iff it maximises utility.
> Earthling: What you have said is literally true. Though to be clear I disagree with you; I desire to perform actions only if they maximise utility.

But intuitively such claims are very strange. This suggests that naturalist semantics are incorrect, since we do not regard the Moral Twin Earthlings claims as being even 'literally' or 'strictly speaking' true, but according to naturalists we should.

Addressing this issue, Gunnar Björnsson and Alexander Almér suggest that in the moral case, unlike the indirect speech case, ordinary speakers just do not ever concern themselves with the literal semantic content of what is said; that is primarily of interest only to philosophers and linguists.[349] Because the role of moral talk is to coordinate action and align attitudes, assessments of the truth or falsity of moral claims only ever target the conative attitude expressed, even if qualified by expressions like 'really' or 'strictly speaking'. This means that, there is no context in which we will say that the Moral Twin Earthlings moral utterances are true, since saying that would express a desire *not to* maximise utility at all times; a desire which, assuming we are utilitarians, we do not have.

This reply seems to me to be a bit of a stretch. The semantic content of moral claims now appears to be doing almost no work in explaining how and why we use moral language in the way that we do. If the literal semantic content of moral claims cannot be assessed by ordinary speakers with claims like, 'That's true', then it seems hard to see in virtue of what that content constitutes the *meaning* of the claim in question. If agreeing with a moral claim or believing that claim to be true only ever amounts, in the everyday contexts that ordinary speakers make moral claims, to expressing a particular conative attitude, then we begin to lose our grip on why this

---

[349] (2009, 2011).

account is naturalist and not just non-cognitivist. In other words, then the naturalist's semantic account seems to have fallen away altogether.[350]

That worry might seem a bit nebulous so allow me to sharpen the point. A good reason not to take Almér and Björnsson's approach is that by doing so we will incur the problems that beset non-cognitivists. I will mention just one problem. Take the following *modus ponens*:

> Giving to charity is right.
> If giving to charity is right then volunteering is right.
> Therefore, volunteering is right.

If I agree with the premises then I am committed to agreeing with the conclusion also. Standard naturalist accounts can explain this easily; since the truth of the propositional content of the premises entails the truth of the propositional content of the conclusion on those accounts, anyone who believes the premises to be true must also think that the conclusion is true. According to Almer and Björnsson's story however, in ordinary contexts, agreeing with moral claims or believing them to be true only ever amounts to having the same conative attitude as the attitude expressed by those claims. In that case, they need to give a different story of why someone who agrees with the premises is committed to agreeing with the conclusion. They need to tell us why someone who desires to give to charity, and desires to volunteer if they desire to give to charity, is rationally bound to desire to volunteer. In other words, they need to solve a version of the *Frege-Geach* problem that plagues non-cognitivists; a problem I mentioned in the last chapter. More specifically, they need to solve a version of the Frege-Geach problem put forward by Mark Schroeder, which challenges the non-cognitivist to explain why moral claims have the *inference-licensing property*.[351] This is the challenge of explaining why, if accepting or believing moral claims to be true just amounts to having a particular desire, accepting certain moral claims rationally binds us to accepting other moral claims, such as when we accept the

---

[350] See Ridge (2014, p.97)
[351] Schroeder (2009)

premises of a moral *modus ponens*. Despite many years of trying, arguably, no solution to the Frege-Geach problem has been found. That would be a serious problem for Almér and Björnsson to take on board, and if this problem were solved then one of the main dialectical advantages that naturalists have over non-cognitivists would have been lost.

So, it might be unwise for naturalists to argue that agreeing or disagreeing with moral claims, and believing moral claims to be true or false, only ever amounts to expressing a particular conative attitude. To have a fully satisfactory account of disagreement they would have to show that there could be a logic of attitudes, which has for decades proved a vexing problem for non-cognitivists.[352]

There is another option. While philosophers often assume that ordinary speakers are tacitly moral realists, anyone who has taught undergraduates or done some public engagement knows that this is often not the case. Ordinary speakers very often seem much more sympathetic towards some kind of moral anti-realism than they are towards realism. Studies have backed this up. One that is particularly relevant for our purposes was undertaken by Khoo and Knobe.[353] They found that ordinary speakers tend to think that in hypothetical scenarios in which we appear to have moral disagreements with extra-terrestrials, both of us might speak truly. This means that, the content of what we say need not be inconsistent with the content of what the extra-terrestrials say if we are to make sense of why we can disagree with them.

Moreover, these findings, Khoo and Knobe argue, turn the normal dialectic on its head. Previously, it was thought that anyone who is committed to saying that 'right' and 'right*' have different content cannot explain why Earthlings and extra-terrestrials (like Moral Twin Earthlings) intuitively disagree. Now we see that it is those who think that 'right' and 'right*'

---

[352] I think that many of the other problems that face non-cognitivist semantics would also apply (see the objections presented by van Roojen (1996), Jackson and Pettit (1998) Dorr (2002), Egan (2007) Schroeder (2008a, 2008b 2010)).

[353] Khoo and Knobe (2016)

must have the same content who fail to respect ordinary speakers' linguistic intuitions. Other experimental findings have also supported this claim.[354] In order to respect these data, our account of disagreement must accommodate cases in which both parties to a disagreement speak truly.

These findings also provide the naturalist with a potential response to Ridge. While Ridge assumes that where I disagree with someone I will also believe what they have said to be false, the experimental data challenge this. Ordinary speakers seem to think that two people can disagree even where neither party has said something false or has the belief that the other has said something false. The naturalist can simply reply to Ridge that there is nothing to explain, and that in fact, naturalist accounts which have it that the disagreement between us and Moral Twin Earthlings is non-content based, actually come out better than rival accounts which explain the disagreement in terms of inconsistent sematic contents. These rival accounts are committed to saying that whenever we disagree with someone, we will say that they have spoken falsely, but the experimental findings show that this is not the case.

However, for reasons that are beyond the scope of this chapter, I am unsure whether the kinds of surveys that Khoo and Knobe rely on are very useful for semantic theorising.[355] That being so, I prefer to canvas another approach.

§5 – Debunking the disagreement intuition

Another strategy that the naturalist could adopt is simply to deny that there is really a disagreement going on in cases like Moral Twin Earth. As discussed in a previous chapter, people have mistaken intuitions about all manner of things, so we could potentially debunk these intuitions pumped by Moral Twin Earth. Since every major metaethical theory has difficulty

---

[354] Bolinger (ms), Ross and Schroeder (2013), Sarkissian et. al. (2011). These findings are discussed briefly in Finlay (2017).

[355] Knobe and Nichols (2017, §3) provide a nice, concise overview of the potential issues associated with the use of surveys in philosophy.

explaining how we could disagree with hypothetical speakers in some Moral Twin Earth-style scenarios, that might suggest to us that it would be best to just accept that we do not really disagree. However, ordinary speaker's linguistic intuitions are the raw data with which we do semantics, so we cannot disregard them lightly. We need a principled reason to debunk our intuitions in this case.

I think, however, that we can make a very good case for that denying that, *strictly speaking,* we do disagree with Moral Twin Earthlings. First, the intuition that we do disagree with them is not universally held. Nor is it, I think, that strongly held by many of those that do have the intuition. Something does seem off about these cases, they do not seem like other more paradigmatic disagreements. Having considered that Moral Twin Earth-style scenarios are problems for everyone, we might well accept that we do not *really* disagree with Twin earthlings, nor regard them as having made some kind of error, so much as are just want and value different things to them. If we would apply 'right' to different actions than those that Twin Earthlings apply 'right*' to, even when we were both fully-informed on non-normative matters, made no logical errors, had fully-reflected, had undergone a full course of cognitive psychotherapy, were maximally coherent, made our judgements behind a veil of ignorance, etc., then it does not seem unreasonable to think that we simply do not disagree when we say, 'X is right', and they say, 'X is not right*'. If, in these kinds of ideal conditions, Moral Twin Earthlings use 'right*' to refer to actions only if they believe has the Kantian property, then it seems to me to be plausible, regardless of its 'practical role', that 'right*' might denote *the Kantian property*, and not whatever is denoted by our term 'right' (if it denotes anything at all).

Second, the way that Moral Twin-Earth scenarios get presented to us are often very misleading which should make us wary of trusting the intuitions that have been pumped by them. Above I set out the thought experiment in a fair way by omitting the many misleading features that it

typically comes with. Here are a few of the ways that Horgan and Timmons mislead their audience[356]:

1. They tell us that the Moral Twin Earthling terms 'good*', 'bad*', 'right*, 'wrong*' etc. are *moral terms*[357] and they tell us that Moral Twin Earthlings make *moral claims*. [358] Since we know that moral terms denote moral properties, this will instantly lead us to believe that they must be using terms like 'good*' to talk about *goodness*. If the naturalist is right then, 'good*' actually denotes some natural property other than *goodness*, so Horgan and Timmons unfairly influence the reader against naturalism. Precisely what is at stake in the debate is whether or not 'good*' and the like, are really moral terms or whether (given the difference in referent) they have merely similar practical roles to moral terms. Similarly, by telling us that in their thought experiment members of both communities are making *moral* claims, Horgan and Timmons prejudice the reader into thinking that there is a shared subject matter. If we are told that their claim, 'X is good*' is a *moral* claim, then it is hard to avoid the conclusion that our moral claim 'X is not good' must be inconsistent with it.

2. We are also told, that Moral Twin Earthlings have 'folk morality'[359], 'moral sensibilities'[360], 'moral educations',[361] etc. By telling us, at the outset, that Moral Twin Earthlings are moralisers in these ways, Horgan and Timmons lead us to judge that they morally disagree with us.

3. Relatedly, we are told that terms like 'good*' and 'right*' are used to *evaluate* actions. That presumably makes them evaluative terms, which are a subset of normative terms. If 'good*' and

---

[356] Laurence, Margolis and Dawson (1999), Copp (2000), Merli (2002) and Dowell (2016) mention more ways that Horgan and Timmons mislead us which I do not cover.
[357] e.g. Horgan and Timmons (1991a, p.460)
[358] e.g. Horgan and Timmons (1996) and (2015, p.358).
[359] Horgan and Timmons, (2009,p.9).
[360] ibid. p.20, fn.5
[361] ibid.

'right*' are normative terms then they presumably share a subject matter with our terms like 'good' and 'right'.[362]

4. They repeatedly stipulate that explorers from Earth who visited Twin-Earth would translate 'good*' as 'good'. [363] If they would translate the terms this way then presumably those terms must have the same meaning. This stipulation prejudices the reader against the view that these terms have different meanings.

5. They tell us that were we to say that 'good' and 'good*' denote different properties then a kind of moral relativism would be true, according to which each of us make moral claims which are only ever true *relative to our community*.[364] But in fact, if relativism were true, then the claims 'X is good' and 'X is good*' would have the same *semantic content* or meaning (though those contents would be true or false only *relative to a communities normative outlook*). By telling us that if their view is not correct, then relativism must be correct, Horgan and Timmons present us with two views both of which have it that 'good' and 'good*' have the same meaning. But this is a false dichotomy. If 'good' and 'good*' denote different natural properties then we and the Moral Twin Earthlings would each say something (absolutely) true in our own language, if naturalism is true. By inaccurately restricting the field of possible views to only those where 'good' and 'good*' have the same meaning, Horgan and Timmons mislead the audience. Additionally, since many readers are hostile to moral relativism, being told, incorrectly, that it is the only alternative view, unfairly pushes the reader towards Horgan and Timmons' preferred theory.

6. Unlike me, Horgan and Timmons do not use a '*' to distinguish our moral terms from the relevant Moral Twin Earthling terms. They say things like this: "Moral Twin Earthlings … use the terms "good" and "bad," "right" and "wrong"".[365] The most natural way to understand a

---

[362] e.g. Horgan and Timmons (1991a, p.459)
[363] ibid.
[364] e.g. Horgan and Timmons (1996).
[365] Horgan and Timmons (1991)

phrase like 'Moral Twin Earthlings use "good"', is as asserting that Moral Twin Earthlings use the *English* term 'good'.[366] If Moral Twin Earthlings used the English term 'good' then they would obviously disagree with us if we say, 'X is good', and they say, 'X is not good'.

Given that the thought experiments are usually set out in these misleading we should be wary of the intuitions that have been drawn from them.

Third, we have other potential explanations of why our intuitions could be mistaken. As was pointed out in the last subsection, in these cases there is a kind of tension or conflict resulting from the mutually unsatifiable desires that moral claims plausibly express even on a naturalist account. If that does not amount to a disagreement proper (as we are currently supposing), it could at least be mistaken for one by ordinary speakers. These cases have many of the hallmarks of disagreement, and, since ordinary speakers do not have very considered views about what disagreement consists in, they can potentially mistake them for genuine disagreements.

Finally, as Janice Dowell argues in a brilliant paper, for the intuitions pumped by Twin-Earth scenarios to have probative value it would need to be the case that competence with a term like 'right' requires knowing just which similarities in the use of 'right' and the term 'right*', from any other hypothetical language, are sufficient for 'right' and 'right*' to have the same meaning.[367] This is much more than we usually assume is the case when constructing semantic theories from people's linguistic intuitions.[368] Ordinarily, we construct semantic theories based on how competent speakers are disposed to *deploy* expressions from their own language. Whether those speakers can also accurately tell how to *translate* those terms into any hypothetical language is a different matter. In this regard, the Moral Twin Earth thought experiment requires of competent speakers more than Putnam's original Twin Earth experiment. Putnam asks us whether we, who know that water is H_2O, would apply our term 'water' to XYZ were we to visit Twin Earth. He

---

[366] A point raised by Dowell (2016)
[367] ibid.
[368] ibid. pp.12-14.

thinks that we would not, suggesting that XYZ does not fall under the extension of 'water'. Putnam did not ask whether competent speakers would translate 'water' as 'water*' – a Twin English term used to denote XYZ. If he had, it is not obvious that the answer to this question would have told us much about the semantics of 'water'.

If the assumption that ordinary speakers can translate terms they know into any hypothetical language is not true, then there would be no reason for us to think that ordinary speakers' intuitions about which hypothetical alien-language speakers we do or do not disagree with, tell us anything decisive about the meaning of terms in our language.

And it seems quite implausible that all competent speakers have this knowledge. I am competent with a term like 'hammer', for example, but I do not think I know which similarities in use between 'hammer' and a term from any hypothetical language make those terms have the same meaning. If a hypothetical community uses a term from their language, 'hammer*', when pointing at a tool they use to drive nails into wood with, is that enough of a similarity in use for 'hammer*' to have the same meaning as 'hammer'? Perhaps it is. But suppose that what they call 'hammers*' are naturally occurring objects on their planet. Suppose also that 'hammers*' are typically different in shape from hammers and that they are typically made from rock. Do those differences in usage mean that it is better to say that they use some other tool, and not hammers, to drive nails into wood? Does 'hammer*' translate better as 'bludgeon' or 'rock' than it does 'hammer'? Or is it more accurate to say that they have hammers which have different shapes and manufacturing origins to our hammers? Whatever the answers to these questions, ordinary speakers do not necessarily have to know what they are in order to count as competent with the term 'hammer'.

Similarly, ordinary speakers need not know just which similarities in usage between 'right' and 'right*' are sufficient for them to have the same meaning, so their intuitions about whether the person who says 'X is right' and the hypothetical speaker who says 'X is not right*' disagree

would not show us that 'right' and 'right*' do, or do not, have the same meaning. Competent users of moral vocabulary (who can seemingly be very young[369]), do not, as Dowell points out, plausibly all possess "the skills of a field linguist engaged in radical translation".[370] So, even if ordinary speakers do have the intuition that we disagree with Moral Twin Earthlings, we are safe to question those intuitions without undermining the foundations on which we build semantic theories.

All in all then, it seems reasonably safe to reject the claim that we and Moral Twin earthlings disagree. 'Good' and 'good*' might be orthographically very similar and they might serve similar practical roles, but that is not reason enough to think that they must have the same meaning. It is not intuitively obvious that we do disagree with the Moral Twin Earthlings and, even if it were, those intuitions would not be decisive.

Conclusion

In this chapter, I looked at Horgan and Timmons' 'direct' Moral Twin Earth argument. I argued, first, that it is difficult, on anyone's account, to explain why we intuitively disagree with the hypothetical speakers in these kinds of scenarios. I then sketched two solutions. One solution involved widening our understanding of disagreement so that it including *non-content based* disagreements. This solution, I argued, faces some serious issues when it comes to explaining why expressions of disagreements seem to go hand-in-hand with truth and falsity attributions. The alternative solution was to reject the claim that we disagree with Moral Twin Earthlings. The intuitions guiding this claim are dubious in multiple ways, so I think that this is a perfectly defensible route for a naturalist to take.

---

[369] See Kagan (2007).
[370] Dowell (2016), p.21

Conclusion

Derek Parfit claims that it is "amazing" that we still need to defend the claim that normative properties are not natural properties.[371] There is nothing amazing about it. Naturalism provides a vindicatory account of our ordinary normative claims with appeal to nothing more than garden-variety natural properties. Unlike non-cognitivist positions, naturalism can do all this while remaining compatible with orthodox theories of semantics and truth. Naturalism is also well positioned to explain why normative properties supervene on natural properties.[372]

The anti-naturalist argument that normative properties could not be natural properties because we can doubt any naturalistic analysis is not convincing. As I showed in chapters 1, we are able to doubt analyses that are correct and, even if we were not, normative terms need not necessarily be analysable in terms of natural terms if normative properties are natural properties. I showed in chapter 2 that there are no metaphysical laws either requiring or forbidding normative properties from exemplifying themselves which do not also apply to natural properties.

Parfit contends that, regardless of how debates concerning meaning and analysis turn out, we can know that normative properties are not natural properties. Parfit thinks that this is just intuitively obvious to anyone who considers the matter. In chapter 3, I showed that there is no convincing argument against naturalism from such intuitions. I also showed that naturalists can explain why normative properties are normative in the reason-involving sense and that nothing Parfit says shows that natural properties could not both be normative properties and non-causally make actions have those properties.

Parfit also argues that if normative properties were natural properties then normative claims could not be substantive and positive, which they are. This is considered Parfit's master argument against naturalism, but I showed in chapter 4 that it is both unconvincing and

---

[371] Parfit, (2001), p.39
[372] See Dreier (2015).

uninteresting since the key premise of that argument just amount to a flat denial of naturalism. More interesting are Parfit's auxiliary arguments for that premise, but they ultimately also fail.

Naturalists can also respond to arguments from motivational internalism. There do not appear to be any decisive arguments for motivational internalism, so naturalists, as well as other descriptivists are safe from arguments that require it. There are also some good reasons to think that internalism is not true. If there are no necessary connections between distinct existences then internalism requires that normative judgements and motivational states are not distinct mental states. This is very implausible because the strength of normative judgements does not covary with the strength of the motivational states we have to act on those judgements. What is more, there is no very convincing story to tell about just which motivational states we must be in when we make normative judgements. Unlike some rival accounts, naturalists can say that the connection between normative judgements and motivation is as strong as it intuitively is, but are not committed to also saying that it is stronger.

The possibility of fundamental normative disagreements might be thought to be one last spanner that can be thrown in the naturalist works. Seemingly, we can disagree with speakers from hypothetical communities who use terms which would refer to different natural properties to our moral terms if naturalism were correct. I showed in chapter 6 that scenarios involving seeming disagreements with communities of hypothetical speakers provide a difficult explanatory burden for us all. One way to solve these problems is by appealing to theories of non-content based disagreement; theories which are available to the naturalist to adopt. Another, less costly way, to address these problems is just to deny that we really disagree with the hypothetical speakers in these scenarios.

Given the advantages of naturalism, and given that we can remove the roadblocks to naturalism that philosophers have thought are impassable, naturalism should appear a promising theory. However, there is still be a long way to go before we can be confident that naturalism is the

correct metaethical theory. To establish that we would need a particular account of which natural properties are identical to which normative properties, which makes good sense of the way we use normative language. This will be a very difficult task, but none of the arguments that I have looked at from Moore, Parfit, Horgan and Timmons, as well as other philosophers, should deter us from trying to complete it.

References

- Altman, Andrew (2004), "Breathing Life into a Dead Argument: G.E. Moore and the Open Question Argument", *Philosophical Studies*, Vol.117, No.3.

- Anscombe, Elizabeth (1957), *Intention*, Cambridge, MA: Harvard University Press.

- Aristotle (1985), *Nichomachean Ethics*, Terence Irwin (trans.), Indianopolis/Cambridge: Hackett.

- Ayer, A.J., (1936), *Language Truth and Logic*, UK: Penguin Classics.

- Baldwin, Thomas (1993), "Introduction", in G.E. Moore, *Principia Ethica (Revised Edition)*, Thomas Baldwin (ed.), Cambridge: Cambridge University Press.

- Barker, Stephen (1999), "Is Value Content a Component of Conventional Implicature", *Analysis*, 60.

- Bedke, Matthew S. (2012), "Against Normative Naturalism", *Australasian Journal of Philosophy*, 90:1.

- Bex-Priestley, Graham (2018), *Higher-Order Expressivism, The Dyadic Nature of Moral Judgement*, PhD Thesis, University of Sheffield.

- Björnsson, Gunnar (2012), "Do 'objectivist' features of moral discourse and thinking support moral objectivism?', *Journal of Ethics*, 16.

- Björnsson, Gunnar & Finlay, Stephen (2010), "Metaethical Contextualism Defended", *Ethics*, 121:1.

- Björnsson, Gunnar & Almér, Alexander (2009), "Contextualism, Assessor Relativism and Insensitive Assessments, *Logique & Analyse*, 208.
  ---, (2011), "The Pragmatic of Insensitive Assessments", *The Baltic International Yearbook of Cognition, Logic and Communication*, Volume 6: *Formal Semantics and Pragmatics*.

- Björnsson, Gunnar & Olinder, Ragnar Francén (2013), "Internalists Beware – We Might All Be Amoralists", *Australasian Journal of Philosophy*, 91 (1).

- Björnsson, Gunnar, Strandberg, Caj, Olinder, Ragnar Francén, Eriksson, John & Björklund, Fredrik (2014), *Motivational Internalism*, Oxford: Oxford University Press.

  ---, (2015), "Motivational Internalism and Folk Intuitions", *Philosophical Psychology*, 28 (5).

- Black, Max (1944), "The Paradox of Analysis", *Mind,* 53: 211.

  ---, (1945), "The "Paradox of Analysis" Again: A Reply", *Mind*, 54: 215.

- Blackburn, Simon (1984), *Spreading the Word*, Oxford: Oxford University Press.

  ---, (1993), *Essays in Quasi-Realism*, Oxford: Oxford University Press.

  ---, (1995), "The Flight from Reality", in, Rosalind Hursthouse, Gavin Lawrence & Warren Quinn (eds.), *Virtues and Reasons*, Oxford: Clarendon Press.

  ---, (1998), *Ruling Passions*, Oxford: Oxford University Press.

  ---, (2002), "Reply", *Philosophical Books: Symposium on Simon Blackburn's Ruling Passions*, Vol. 42 (1).

- Blair, R. J. R. (1995), "A cognitive development approach to morality: Investigating the psychopath", *Cognition*, 57.

  ---, (2008), "The cognitive neuroscience of psychopathy and implications for judgements of responsibility", *Neuroethics,* 1 (3).

- Block, Ned and Stalnaker, Robert (1999), "Conceptual Analysis, Dualism and the Explanatory Gap", *Philosophical Review*, 108: 1.

- Boisvert, Daniel (2008), "Expressive Assertivism", *Pacific Philosophical Quarterly*, 89 (2).

- Bolinger, Renee (ms.), "Metalinguistic negotiations in moral disagreement".

- Boyd, Richard (1988/1995), "How to be an Ethical Realist", in Paul K. Moser and J.D. Trout (eds.), *Contemporary Materialism: A Reader*, London and New York: Routledge.

- Brandt, Richard B. (1979), *A Theory of the Good and the Right*, Oxford: Clarendon Press.

- Brink, David O. (1989), *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.

- Broad, C.D. (1934), "Is "Goodness" a Name of a Simple Non-natural Quality?", *Proceedings of the Aristotelian Society*, Vol.34.

  ---, (1968), "Certain Features in Moore's Ethical Doctrines", in, P.A. Schilpp (ed.), *The Philosophy of G.E. Moore*, Cambridge: Cambridge University Press.

- Chalmers, David (2004), "Imagination, Indexicality and Intentions", *Philosophy and Phenomenological Research*, Vol.LXVIII: 1.

- Chalmers, David & Jackson, Frank (2001), "Conceptual Analysis and Reductive Explanation", *Philosophical Review*, 110.

- Cholbi, Michael (2011), *Suicide: The Philosophical Dimensions*, Peterborough: Broadview Press.

- Church, Alonzo (1946), "Review", *The Journal of Symbolic Logic*, 11:4.

- Clark, Ernest & Clark, Maria (1941), "What is goodness?, *The Australasian Journal of Psychology and Philosophy*, 19:2.

- Cline, Brendan (2015), "Moral Explanations, Thick and Thin", *Journal of Ethics and Social Philosophy*, Vol. 9 (2).

- Copp, David (1995), "Moral Obligation and Moral Motivation", *Canadian Journal of Philosophy Supplemental Issue*, 25.

  ---, (2000), "Milk, Honey and the Good Life on Moral Twin Earth", *Synthese*, 124: 1.

  ---, (2001), "Realist-Expressivism: A Neglected Option for Moral Realism", *Social Philosophy and Policy*, 18.

  ---, (2003), *Morality, Normativity and Society,* Oxford: Oxford University Press.

  ---, (2012), "Normativity and reasons: five arguments from Parfit against normative naturalism", in *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press.

---, (2017), "Normative naturalism and normative nihilism: Parfit's dilemma for naturalism", in, Simon Kirchin (ed.), *Reading Parfit: On What Matters*, UK: Routledge.

- Crisp, Roger (2000), "Review of *Value… And What Follows* by Joel Kuppeman", *Philosophy*, 75 (3).

  ---, (2005), "Value, Reasons and the Structure of Justification: How to Avoid Passing the Buck", *Analysis*, 65 (1).

  ---, (2008), "Goodness and Reasons: Accentuating the Negative", *Mind*, 117 (466).

- Cullison, Andrew (2008), "Three Ways to Resolve Open Questions", *Journal of Ethics and Social Philosophy*, 3:1.

- Dancy, Jonathan (2000), *Practical Reality*, Oxford: Oxford University Press.

  ---, (2004), *Ethics without Principles*, Oxford: Oxford University Press.

  ---, (2005), "Ethical Intuitionism", in, D.M. Borchert (ed.), *Encyclopaedia of Philosophy,* USA: MacMillan.

  ---, (2006), "Non-naturalism", in, David Copp (ed.), *The Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press.

- D'Arms, Justin & Jacobson, Daniel (2000a), "The Moralistic Fallacy: On the "appropriateness" of Emotions", *Philosophy and Phenomenological Research*, Vol. LXI (1).

  ---, (2000b), "Sentiment and Value", *Ethics*, 110 (4).

- Davidson, Donald (1973), "Radical Interpretation", *Dialectica*, 27.

- Donagan, Alan (1981), "W.K. Frankena and G.E. Moore's Metaethics", *The Monist*, 64 (3).

- Dorr, Cian (2002), "Non-cognitivism and Wishful Thinking", *Noûs, 36:1*.

- Dowell, J.L. (2016), "The Metaethical Insignificance of Moral Twin Earth", in, Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol. 11, Oxford: Oxford University Press.

- Dowell, J.L. & Sobel, David (2017), "Advice for Non-Analytical Naturalists", in, Simon Kirchin (ed.), *Reading Parfit: On What Matters*, UK: Routledge.

- Dreier, James (1990), "Internalism and Speaker Relativism", *Ethics*, Vol.101: 1.

  ---, (2009), "Relativism (And Expressivism) And The Problem of Disagreement", *Philosophical Perspectives*, 23, *Ethics*.

  ---, (2015), "Explaining the Quasi-Real", in, Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics Volume 10*, Oxford: Oxford University Press.

- Durrant, R.G. (1970), "Identity of properties and the definition of 'good'", *Australasian Journal of Philosophy*, 48:3.

- Dummett, Michael (1978), *Truth and Other Enigmas*, Cambridge, MA: Harvard University Press.

- Egan, Andy (2007), "Quasi-Realism and Fundamental Moral Error", *Australasian Journal of Philosophy*, 85:2.

- Eggers, Daniel (2015), "The motivation argument and motivational internalism", *Philosophical Studies*, 172 (9).

- Evers, Daan (2011), "Subjectivist Theories of Normative Language", DPhil Dissertation, Oxford University.

  ---, (2014), "Moral Contextualism and the Problem of Triviality", *Ethical Theory: Moral Practice*, 17.

- Feldman, Fred (2005), "The Open Question Argument: What It Isn't; And What It Is", *Philosophical Issues*, 15, Normativity.

- Feyerabend, P. K. (1956), "A Note on the Paradox of Analysis", *Philosophical Studies,* 7:6.

- Fillmore, Charles (1982), "Towards a Descriptive Account of Spatial Deixis", in, R. Javella and W. Klein (eds.), *Speech, Place and Action*, London: Wiley.

- Finlay, Stephen (2005), "Value and Implicature", *Philosophers' Imprint*, 5.

  ---, (2008), "The Error in the Error Theory", *Australasian Journal of Philosophy*, 86 (3).

  ---, (2009), *"*Oughts and Ends", *Philosophical Studies*, 143.

  ---, (2011), "Errors Upon Errors: A Reply to Joyce", *Australasian Journal of Philosophy*, 89 (3).

  ---, (2014), *Confusion of Tongues*, Oxford: Oxford University Press.

  ---, (2017), "Disagreement Lost and Found", in, Russ Shafer-Landau, *Oxford Studies in Metaethics*, Vol.12, Oxford: Oxford University Press.

- Fleming, Partick (2015), "The Normativity Objection to Normative Reduction", *Acta Analytica*, 30.

- Fodor, Jerry (1981), "The Present Status of the Innateness Controversy", *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge MA: MIT Press.

  ---, (1998a), *Concepts: Where Cognitive Science Went Wrong*, Oxford: Clarendon Press.

  ---, (1998b), "The Trouble with Psychological Darwinism", in, *London Review of Books*, Vol.20: 2.

  ---, (2004), "Having Concepts: A Brief Refutation of the Twentieth Century", *Mind and Language*, 19:1.

- Foot, Phillipa (1979), *Virtues and Vices*, Berkeley and Los Angeles: California University Press.

- Frankena, William (1939), "The Naturalistic Fallacy", *Mind*, Vol.48, No.192.

  ---, (1963), *Ethics (Second Edition)*, New Jersey: Prentice Hall.

- Frankfurt, Harry (1988), *The Importance of What We Care About: Philosophical Essays*. Cambridge: Cambridge University Press

  ---, (2004), *Reasons of Love*, New Jersey: Princeton University Press.

- Frege, Gottlob (1997), "*On* Sinn and *Bedeutung*", in, Michael Beaney (ed.), *The Frege Reader*, Oxford: Blackwell Publishing.

- Fricker, Miranda (2007), "The Value of Knowledge and the Test of Time", *Epistemology, Royal Institute of Philosophy Supplement.*

- Gert, Joshua & Mele, Alfred (2005), "Lenman on Externalism and Amoralism: An interplanetary exploration", *Philosophia*, Vol. 32.

- Gibbard, Allan (1990), *Wise Choices, Apt Feelings*, Cambridge: Harvard University Press.

  ---, (1992), "Morality and Thick Concepts", *Proceedings of the Aristotelian Society Supplementary Volume*, Vol.66.

  ---, (1993), "Reply to Blackburn," *Philosophical Issues*, 4.

  ---, (1999), "Morality as Consistency in Living: Korsgaard's Kantian Lectures," *Ethics*, 110: 140–164.

  ---, (2003), *Thinking How to Live*, Cambridge: Harvard University Press.

  ---, (2011), "How Much Realism? Evolved Thinkers and Normative Concepts," in R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. 6, Oxford: Oxford University Press.

  ---, (2012), *Meaning and Normativity*, Oxford: Oxford University Press.

- Green, Georgia M. (1989), *Pragmatics and Natural Language Understanding*, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

- Grice, Paul (1989), *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

- Hale, Bob (1993), "Could there be a Logic of Attitudes", in, John Haldale and Crispin Wright (eds.), *Reality, Representation and Projection*, Oxford: Oxford University Press.

- Hamblin, C.L. (1973), "Questions in Montague English", *Foundations of Language,* 10 (1).

- Hampton, Jean E. (1998), *The Authority of Reason*, Cambridge: Cambridge University Press.

- Hare, R.M. (1954, Reprint (1972)) *The Language of Morals*, Oxford: Oxford University Press.

  ---, (1963), *Freedom and Reason¸* Oxford: Clarendon Press.

- Harman, Gilbert (1977), *The Nature of Morality*, Oxford: Oxford University Press.

- Harris, Sam (2010), *The Moral Landscape*, United States: Free Press.

- Hay, Ryan (2014), "Attitudinal Requirements for Moral Thought and Language", in, Guy Fletcher and Michael Ridge (eds.), *Having it Both Ways*, Oxford: Oxford University Press.

- Henning, Tim, "Normative Reasons Contextualism", *Philosophy and Phenomenological Research*, Vol. LXXXVIII: 3.

- Horgan, Terry & Timmons, Mark (1991), "New Wave Realism Meets Moral Twin Earth", *Journal of Philosophical Research*, Vol. XVI.

  ---, (1992a), "Troubles on Moral Twin Earth: Moral Queerness Revived", *Synthese*, 92.

  ---, (1992b), "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived", *Philosophical Papers*, Vol. XXI: 3.

  ---, (1996), "Troubles for Michael Smith's Moral Functionalism", *Philosophical Papers*, 25:3.

  ---, (2000), "Copping Out on Moral Twin Earth", *Synthese*, 124.

  ---, (2009), "Analytic functionalism versus Moral Twin Earth", in, Ian Ravenscroft (ed.), *Minds, Ethics, and Conditionals: Themes From the Philosophy of Frank Jackson*. Oxford: Oxford University Press.

  ---, (2015), "Exploring Intuitions on Moral Twin Earth: A Reply to Sonderholm", *Theoria*, 81: 4.

- Hudson, W.D. (1983), *Modern Moral Philosophy (Second Edition)*, London: MacMillan.

- Huemer, Michael (2005), *Ethical Intuitionism*, UK: Palgrave MacMillan.

- Hume, David (1739-40/1978), *A Treatise of Human Nature* (Second Edition), Oxford: Oxford University Press.

- Hurley, S. L. (1989), *Natural Reasons*, Oxford: Oxford University Press.

- Huvenes, Torfinn Thomesen (2012), "Varieties of disagreement and predicates of taste", *Australasian Journal of Philosophy*, 90:1.

- Jackson, Frank (1998), *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Oxford: Oxford University Press.

  ---, (2011), "On ethical naturalism and the philosophy of language", in, Susan Nuccatelli and Gary Seay (eds.), *Ethical Naturalism: Current Debates*, Cambridge: Cambridge University Press.

  ---, (2016), "In Defence of Reductionism in Ethics", in, Peter Singer (ed.) *Does Anything Really Matter? Essays on Parfit on Objectivity*. Oxford: Oxford University Press.

- Jackson, Frank & Pettit, Philip (1998), "A Problem for Expressivism", *Analysis*, 58:4.

- Johnston, Mark (1989), "Dispositional Theories of Value", *Aristotelian Society Supplemental Issue*, 63 (1).

- Joyce, Richard (2001), *The Myth of Morality*, Cambridge: Cambridge University Press.

  ---, (2006), *The Evolution of Morality*, Cambridge, MA: MIT Press.

  ---, (2011), "The Error in 'The Error in the Error Theory'", *Australasian Journal of Philosophy*, 89 (3).

- Kagan, Jerome (2007), "Morality and its Development", Walter Sinott-Armstrong (ed.), in *Moral Psychology: The Neuroscience of Morality*, Cambridge: MIT Press

- Kalderon, Mark Eli (2004), "Open Questions and the Manifest Image", *Philosophy and Phenomenological Research,* Vol. LXVIII: 2.

- Katz, Jerrold (1966), *The Philosophy of Language*, New York: Harper and Row.

- Khoo, Justin & Knobe, Joshua (2018), "Moral disagreement and moral semantics", *Noûs, 52:1.*

- *King, Jeffrey C. (1998), "What is a Philosophical Analysis?", Philosophical Studies, 90.*

- *Knobe, Joshua & Nichols, Shaun (2017), Experimental Philosophy, Stanford Encyclopaedia of Philosophy.*

- *Korsgaard, Christine (1996), Sources of Normativity, Cambridge, MA: Harvard University Press.*

- *Kramer, Matthew (2009), Moral Realism as a Moral Doctrine, UK: Wiley-Blackwell.*

- Kripke, Saul (1982), *Wittgenstein on Rules and Private Language*, Oxford: Basil Blackwell Publishing.

- Langford, C. H. (1968), "The Notion of Analysis in Moore's Philosophy", in, P.A. Schilpp (ed.), *The Philosophy of G.E. Moore*, Cambridge: Cambridge University Press.

- Laskowski, Nicholas (2019), "The Sense of Incredibility in Ethics", *Philosophical Studies*, 176.

- Laurence, Stephen, Margolis, Eric & Dawson, Angus (1999), "Moral Realism and Twin-Earth", *Facta Philosophica*, 1.

- Lemos, R.M. (1991), "Bearers of Value", *Philosophy and Phenomenological Research,* 51:4.
  ---, (2005), "The Bearers of Intrinsic Value", in, T. Rønnow-Rasmussen M. J. Zimmerman (eds)., *Recent Work on Intrinsic Value*, Netherlands: Springer.

- Lenman, James (1999), "The Externalist and the Amoralist", *Philosophia*, 27.
  ---, (2003a), "Non-cognitivism and the Dimensions of Evaluative Judgement", *Brown Electronic Article Review Service*.
  ---, (2003b), "Moral Deviants and Amoral Saints", *Southern Journal of Philosophy*, 41.
  ---, (2009), "Naturalism Without Tears", *Ratio*, 22.

- Lenman, James & Lutz, Matthew (2018), "Moral Naturalism", *Stanford Encyclopaedia of Philosophy*.

- Levine, Joseph (2001), *Purple Haze: The Puzzle of Consciousness*, Oxford: Oxford University Press.

- Levy, Neil (2011), "Moore on Twin Earth", *Erkenn*, 75.

- Lewis, David (1979a), "Attitudes *De Dicto and De Se*", *Philosophical Review*, 88.

   ---, (1989), "Dispositional Theories of Value", *Proceedings of the Aristotelian Society, Supplemental Issue,* Vol. 63.

- Lewy, Casimir (1976), *Meaning and Modality*, Cambridge: Cambridge University Press.

   ---, (1979b), "Scorekeeping in a Language Game", *Journal of Philosophical Logic*, 8.

- Lillehammer, Hallvard (2004), "Moral Error Theory", *Proceedings of the Aristotelian Society*, Vol.104:1.

- Lycan, William (1986), "Moral Facts and Moral Knowledge", *Southern Journal of Philosophy*, 24.

- McNaughton, David & Rawling, Piers (2003), "Descriptivism, Normativity and the Metaphysics of Reasons", *Aristotelian Society Supplemental Volume*, 77:1.

- Mackie, John (1977), *Ethics: Inventing Right and Wrong*, Great Britain: Penguin Books.

- MacFarlane, John (2007), "Relativism and Disagreement", *Philosophical Studies*, 132:1.

- Margolis, Eric & Laurence, Stephen (1999), "Concepts and Cognitive Science", in Eric Margolis and Stephen Laurence (eds.), *Concepts: Core Readings*, Cambridge, MA: MIT Press.

   ---, (2003), "Concepts and Conceptual Analysis", *Philosophy and Phenomenological Research*, Vol. LXVII, No.2.

   ---, (2004), "Concepts", in Stephen A. Stich and Ted A. Warfield, *Blackwell Guide to Philosophy of Mind*, London: Wiley.

- Marks, Joel (2012), *Ethics without Morals*: *In Defence of Amorality*, New York: Routledge.

- Mates, Benson (1952), "Synonymity", in Leonard Linsky, *Semantics and the Philosophy of Language,* Illinois: University of Illinois Press.

- Mehta, Neil (2013), "How to Explain the Explanatory Gap", *Dialectica*, 67.

- Mele, Alfred (1996), "Internalist Moral Cognitivism and Listlessness", *Ethics*, 106 (4).

- Midgley, Mary (1981), "Trying Out One's New Sword," in, *Heart and Mind*, New York: Palgrave Macmillan.

- Miller, Alexander (2003), *An Introduction to Contemporary Metaethics*, Oxford: Polity Press.

- Millikan, Ruth (1987), *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.

- Moore, G.E. (1968), "Reply to my Critics", in, P.A. Schilpp (ed.), *The Philosophy of G.E. Moore*, Cambridge: Cambridge University Press.

    ---, (1993), *Principia Ethica (Revised Edition),* Thomas Baldwin (ed.), Cambridge: Cambridge University Press.

- Nagel, Thomas (1970), *The Possibility of Altruism*, Oxford: Oxford Clarendon Press.

- Nichols, Shaun (2004), *Sentimental Rules: On the Natural Foundations of Moral Judgement*, New York: Oxford University Press.

- Nuccetelli, Susana and Seay, Gary (2011), "Does analytical moral naturalism rest on a mistake?", in, *Ethical Naturalism*: *Current Debates*, Cambridge: Cambridge University Press.

- O'Connor, David (1982), "Moore and the Paradox of Analysis", *Philosophy*, 57.

- Oddie, Graham (2005), *Value, Reality and Desire*, Oxford: Oxford University Press.

- Olson, Jonas & Bykvist, Krister (2008), "Expressivism and moral certitude", *Philosophical Quarterly*, 59 (235).

- Olson, Jonas (2014), *Moral Error Theory: History, Critique, Defence*, Oxford: Oxford University Press.

- Quine, W.V.O. (1951), "Two Dogmas of Empiricism", *Philosophical Review*, 60.

    ---, (1960), *Word and Object*, Cambridge, MA: MIT Press.

- Papineau, David (2002), *Thinking about Consciousness*, Oxford: Oxford University Press.

- Parfit, Derek (2001), "Rationality and Reasons", in, Dan Egonsson, Björn Petersson, Jonas Josefsson, and Toni Rønnow-Rasmussen (ed.) *Exploring Practical Philosophy: From Action to Values*, Aldershot: Ashgate.

  ---, (2011a), *On What Matters Volume 1*, Oxford: Oxford University Press.

  ---, (2011b), *On What Matters Volume 2*, Oxford: Oxford University Press.

  ---, (2016), *On What Matters Volume 3*, Oxford: Oxford University Press.

  ---, (2017), "Responses", in, Simon Kirchin (ed.), *Reading Parfit: On What Matters*, UK: Routledge.

- Peacocke, Christopher (1992), *A Study of Concepts*, Cambridge, MA: MIT Press.

- Perry, John (1977), "Frege on Demonstratives", *Philosophical Review*, Vol.86: 4.

  ---, (1979), "The Problem of the Essential Indexical", *Noûs*, 13

  ---, (2001), *Knowledge, Possibility and Consciousness*, Cambridge, MA: MIT Press.

- Pinker, Steven (2002), *The Blank Slate: The Modern Denial of Human Nature,* Great Britain: Allen Lane, The Penguin Press.

- Phillips, David (1998), "The Middle Ground in Moral Semantics", *American Philosophical Quarterly*, Vol.35:2.

- Platts, Mark (1979), *Ways of Meaning*, Cambridge, MA: MIT Press.

- Plunkett, David & Sundell, Tim (2013), "Disagreement and the Semantics of Normative and Evaluative Terms", *Philosophers' Imprint*, 13:23.

- Price, Huw (2013), *Expressivism, Pragmatism and Representationalism*, Cambridge: Cambridge University Press.

- Prior, Arthur N. (1949), *Logic and the Basis of Ethics*, Oxford: Clarendon Press.

- Prinz, Jesse (2007), *The Emotional Construction of Reasons*, Oxford: Oxford University Press.

- Putnam, Hilary (1962a), "It Ain't Necessarily So", *The Journal of Philosophy*, 59.

--- (1962b), "The Analytic and the Synthetic", In H. Feigl & G. Max- well (eds.), *Minnesota Studies in the Philosophy of Science*, Volume III, Minneapolis: University of Minnesota Press.

---, (1975), "The Meaning of "Meaning"", *Minnesota Studies in the Philosophy of Science*, Vol. 7.

---, (1981), *Reason, Truth and History*, Cambridge: Cambridge University Press.

- Rabinowicz, Wlodek & Rønnow-Rasmussen, Toni (2004), "The Strike of the Evil Demon: On Fitting Pro-Attitudes and Value", *Ethics*, 114 (3).

- Radcliffe, Elizabeth S. (2006), "Moral internalism and moral cognitivism in Hume's metaethics", *Synthese,* 152.

- Railton, Peter (1989), "Naturalism and Prescriptivity", *Social Philosophy and Policy*, Vol.7: 1.

  ---, (1993), "Non-cognitivism about Rationality: Benefits, Costs and an Alternative", *Philosophical Issues*, 4.

  ---, (2003), *Facts, Values and Norms*, Cambridge: Cambridge University Press.

  ---, (2009), "Practical competence and fluent agency", in, David Sobel & Steven Wall (eds.), *Reasons for Action*, Cambridge: Cambridge University Press.

  ---, (2016), "Two Sides of the Metaethical Mountain", in, Peter Singer (ed.) *Does Anything Really Matter? Essays on Parfit on Objectivity.* Oxford: Oxford University Press.

- Raz, Joseph (1999), "Explaining Normativity: On rationality and the justification of Reason", *Ratio*, 12 (4).

- Rey, Georges (1993), "The Unavailability of What We Mean", in Jerry Fodor and Ernest Lepore (eds.), *Holism: A Consumer Update*, Netherlands: Rolopi B.V.

- Rickless, Samuel (2015), "Plato's Parmenides", *Stanford Encyclopaedia of Philosophy*.

- Ridge, Michael (2003), "Review of Michael Smith's Evaluation, Uncertainty and Motivation", *Brown Electronic Article Review Service.*

  ---, (2014), *Impassioned Belief*, Oxford: Oxford University Press.

- Ridge, Michael and McKeever, Sean (2005), *Principled Ethics*, Oxford: Oxford University Press.

- Robinson, Denis (2009), "Moral Functionalism, Ethical Quasi-Relativism, and the Canberra Plan", in David Braddon-Mitchell and Robert Nola (eds.), *Conceptual Analysis and Philosophical Naturalism*, Cambridge, Ma: MIT Press.

- Ross, Jacob & Schroeder, Mark (2013), "Reversibility or Disagreement", *Mind,* 122: 485.

- Rubin, Michael (2014), "Biting the Bullet on Moral Twin Earth"; *Philosophical Papers*, 43:2.

- Russell, Bertrand (1903), "Review of G.E. Moore, *Principia Ethica (1903)*", *The Cambridge Review 25*, *Lit Supplement.*

  ---, (1990), *The Problems of Philosophy,* Indianapolis/Cambridge: Hackett Publishing Company, Inc.

- Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011), "Folk Moral Relativism." *Mind and Language*, 26 (4).

- Scanlon, T.M. (2000), *What We Owe to Each Other* (Fourth Printing), USA: Harvard University Press.

  ---, (2006), "Reasons and Decisions", *Philosophy and Phenomenological Research*, Vol. LXXII, 3.

  ---, (2014), *Being Realistic about Reasons*, Oxford: Oxford University Press.

- Scaratino, Andrea and De Sousa, Ronald (2018), "Emotion", *Stanford Encyclopaedia of Philosophy.*

- Schroeder, Mark (2005), "Realism and Reduction: The Quest for Robustness", *Philosophers' Imprint*, Vol.5 (1).

---, (2007), *Slaves of the Passions*. Oxford: Oxford University Press.

---, (2008a), *Being For*, Oxford: Oxford University Press.

---, (2008b), "How Expressivists Can and Should Solve Their Problem with Negation", *Noûs, 42:4.*

---, (2009), "Hybrid Expressivism: Virtues and Vices.", Ethics, 119: 2.

---, (2010), *Non-cognitivism in Ethics*, Abingdon: Routledge.

---, (2016), "What Matters about Metaethics", in, Peter Singer (ed.) *Does Anything Really Matter? Essays on Parfit on Objectivity*. Oxford: Oxford University Press.

- Shanklin, Robert (2011), "On Good and 'Good'", PhD. dissertation, University of Southern California.

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

- Silk, Alex (2016), *Discourse Contextualism*, Oxford: Oxford University Press.

---, (2017), "Normative Language in Context", in, Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, Vol.12, Oxford: Oxford University Press.

- Singer, Peter (1993), *Practical Ethics: Second Edition*, Cambridge: Cambridge University Press.

- Smith, Michael (1987), "The Humean Theory of Motivation", *Mind*, Vol. 96: 381.

---, (1988), "Reason and Desire", *Proceedings of the Aristotelian Society*, 88.

---, (1994), *The Moral Problem*, Great Britain: Blackwell.

---, (2002), "Evaluation, Uncertainty and Doubt", *Ethical Theory and Moral Practice*, 5.

- Snook, I.A. (1970), "Katz on 'good'", *Australasian Journal of Philosophy*, 48:1.

- Sobel, David (2016), *From Valuing to Value: Towards a Defense of Subjectivism*, Oxford: Oxford University Press.

- Stalnaker, Robert (1981), "Indexical Belief", *Synthese*, 49.

  ---, (2002), "Common Ground", *Linguistics and Philosophy*, Vol.25: 5/6.

  ---, (2014), "Disagreement and Projection", in, *Context*, Oxford: Oxford University Press.

- Stevenson, Charles L. (1944), *Ethics and Language*, New Haven, CT: Yale University Press.

  ---, (1963), *Facts and Values*, New Haven, CT: Yale University Press.

- Stocker, Michael (1979), "Desiring the Bad: An Essay in Moral Psychology", *The Journal of Philosophy*, Vol.76:12.

- Strandberg, Caj (2004), "In Defence of The Open Question Argument", *The Journal of Ethics*, 8.

- Streumer, Bart (2008), "Are there Irreducibly Normative Properties?", *Australasian Journal of Philosophy*, 86.

  ---, (2011), "Are Normative Properties Descriptive Properties?", *Philosophical Studies,* 154.

- Sturgeon, Nicholas (1988), "Moral Explanations". In Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*, Ithaca and London: Cornell University Press.

  ---, (1986), "What difference does it make whether moral realism is true?", *The Southern Journal of Philosophy*, Vol. XXIV, Supplement.

  ---, (2003), "Moore on Ethical Naturalism", *Ethics,* Vol.113: 3. Centenary Symposium on G.E. Moore's *Principia Ethica.*

  ---, (2005), "Ethical Naturalism", in, David Copp, *Oxford Handbook of Ethical Theory*, Oxford: Oxford University Press.

  ---, (2009), "Doubts about the Supervenience of the Ethical", in, Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics Vol. 4*, Oxford: Oxford University Press.

- Sutton, Peter A. (2014), "Moore's "New" Open Question Argument", *Res Philosophica*, 91:4.

- Svavarsdóttir, Sigrún (1999), "Moral cognitivism and motivation", *Philosophical Review*, 108:2.

   ---, (2002), "On Simon Blackburn's *Ruling Passions*", *Philosophical Books*, 108.

- Syntsma, Sharon E. (1995), "Ethical Internalism and Moral Indifference", *Journal of Value Inquiry*, 29.

- Szabó, Zoltán Gendler (2001), "Adjectives in Context", in, R. Harnish and I. Kenesei (eds.), *Perspectives on Semantics, Pragmatics and Discourse*, Amsterdam: John Benjamins.

- Thomson, Judith Jarvis (2001), *Goodness and Advice*, UK: Princeton University Press.

   ---, (2008), United States: Carus Publishing Company.

- Timmons, Mark (1998), *Morality Without Foundations*, Oxford: Oxford University Press.

- Toppinen, Teemu (2015), "Expressivism and the Normativity of Attitudes", *Pacific Philosophical Quarterly*, 96 (2).

- Tresan, John (2009), "The Challenge of Communal Internalism", *The Journal of Value Inquiry*, 43.

   ---, (2010), "Question authority: in defense of moral naturalism without clout", *Philosophical Studies*, Vol.150: 2.

   ---, (2015), "Naturalistic Moral Realism and Motivational Internalism", in Gunnar Björnsson, Caj Strandberg, Ragnar Francén Olinder, John Eriksson & Fredrik Björklund, *Motivational Internalism*, Oxford: Oxford University Press.

- Unwin, Nicholas (1999), "Quasi-Realism, Negation and the Frege-Geach Problem", *Philosophical Quarterly* 50: 196.

- van Roojen, Mark (1996), "Expressivism and Irrationality", *Philosophical Review*, 105:3.

- von Wright, George Henrik (1963), *The Varieties of Goodness*, London: Routledge & Kegan Paul.

- Wedgwood, Ralph (2007), *The Nature of Normativity*, Oxford: Oxford University Press.

- Williams, Bernard (1981), *Moral Luck*, Cambridge: Cambridge University Press.

- Wittgenstein, Ludwig (1965), *"*A Lecture on Ethics", *Philosophical Review*, Vol. 74: 1.

- White, Morton (1945), "Note on the "Paradox of Analysis"", *Mind* 54: 215.

  ---, (1948), "On the Church-Frege Solution of the Paradox of Analysis", *Philosophy and Phenomenological Research*, Vol. 9:1

- Yetter-Chappell, Helen and Yetter Chappell, Richard (2012), "Mind-Body Meets Metaethics: A Moral Concept Strategy", *Philosophical Studies*, 165:3.

- Zangwill, Nick (2008), "The Indifference Argument", *Philosophical Studies*, 138.

  ---, (2009), "Non-cognitivism and Variable Motivation", in, C. Sandis (ed.), *New Essays on the Explanation of Action*, New York: Palgrave MacMillan.

- Zeman, Dan (forthcoming), "Faultless Disagreement", in, Martin Kusch (ed.), *Routledge Handbook of Relativism*, UK: Routledge.

- Ziff, Paul (1960), *Semantic Analysis*, Ithaca, New York: Cornell University Press.