NATURALISING THE UNITY OF DESIRES: MICHAEL SMITH ON THE ANALYSIS OF 'RIGHT'

by Padriac Amato Tahua O'Leary

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Philosophy, the University of Auckland, 2015.

Abstract:

This thesis explores a neglected facet of Michael Smith's meta-ethics in *The Moral Problem*. Although Smith claims to be a naturalist, he thinks that some popular accounts of a defensible naturalism in ethics fail. Thus he argues that a network analysis and reduction in the manner of Frank Jackson's moral functionalism is vulnerable to a permutation problem and should be rejected. He also argues that a natural kind treatment of ethical terms forces the inappropriate categorisation of some possibilities as acceptable examples of moral relativism when they are not. By identifying and reconstructing the crucial role played in Smith's argument by the notion of desiderative unity, the thesis argues that Smith's own summary-style analysis and reduction of ethical terms either faces the collapse of his rationalist meta-ethical naturalism into a version of Jackson's moral functionalism or requires that he treat ethical kinds as natural kinds, despite his arguments against such a possibility.

Acknowledgements

It is my sincere wish to acknowledge my deep debts of gratitude to the many people who helped in the long and oft times tumultuous course of my Ph.D. First and foremost Megan Hall and Áine Ngahuia Patricia Hall O'Leary without whose love, support, and sacrifice this thesis simply would not have been possible. Fred Kroon, my supervisor, who has gone to extraordinary lengths to help me finish producing this document. David Braddon-Mitchell, Denis Robinson, Justine Kingsbury and Rosalind Hursthouse who all helped me do at least one of the following: understand the material, formulate my ideas for the thesis, edit documents expressing these ideas, and learn one way or another how to write it. And finally the various people working for the University of Auckland managing arts faculty postgraduates whose help and forbearance has permitted me to complete this thesis.

TABLE OF CONTENTS

Chapter 1 Smith, Jackson and the Moral Problem

This thesis is an extended examination of (an aspect of) the view Michael Smith presents in 'The Moral Problem' (1994)[1]. Though the focus will be on Smith's meta-ethical view I will also contrast and compare it with Frank Jackson's position, primarily because Smith supposes that his meta-ethical position is not part of reductive framework Jackson presents in 'From Metaphysics to Ethics A defence of Conceptual Analysis' (1998)[2]. The main issue I will be examining is whether or not Smith's anti-Humean meta-ethics can sustain this claim of methodological distinctness. In this chapter I will be giving an outline of the relevant parts of both Smith and Jackson's theories.

1.1 Smith

Smith has two goals in TMP. One is to present a defensible form of reductive naturalism. The other is to solve what he calls 'the moral problem'. The moral problem is this. Moral judgements look as if they should be judgements about moral facts, and, at the same time, we suppose that coming to know the moral facts should have some kind of impact on our motivations. Now, the belief desire model of our psychology makes coming to know facts a matter of acquiring true beliefs about those facts, and coming to be motivated is a matter of acquiring desires with the relevant contents. But particular states of belief in general do not entail particular desires. This

[1] I will henceforward use the acronym TMP to refer to Michael Smith's "The Moral Problem" 1994.

[2] I will henceforward use the acronym FMtE to refer to Frank Jackson's "From Metaphysics to Ethics A Defence of Conceptual Analysis" 1998.

problem is expressed by Smith in terms of the following inconsistent triad of claims:

> "1 Moral judgements of the form 'It is right that I ϕ' express a subject's beliefs about an objective matter of fact, a fact about what it is right for her to do
>
> 2 If someone judges that it is right that she ϕs then *ceteris paribus*, she is motivated to ϕ
>
> 3 An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences." (TMP, pg. 12)

1 through 3 locate moral judgements as a species of belief necessarily connected to motivations a species of desire which, if 3 is true, cannot be the case since beliefs should not have necessary connections to desires. In this thesis I am indifferent to the arguments for Smith's particular rationalist solution to the moral problem. I will simply grant it to him. I am interested rather in whether or not Smith's position is really the distinctive form of reductive naturalism he argues that it is. However, to do this we will have to understand how his proposed theory works so I will give a brief description of it here. The details of this theory will be subject to more detailed scrutiny in latter chapters.

1.1.1 Smith's anti-Humean rationalism

Smith adopts the view that reasons to act are about desires you have or would have after being subject to rational criticism (TMP, pp. 154-161). This

is a relatively uncontentious view since without modification it amounts to little more than supposing it would be by anyone's lights better (and perhaps more coherent) to desire in accord with your means-ends beliefs. Smith, however, adds the idea that non-derivative desires – that is, desires that are not merely dependent on prior existing desires in the way desires for means to ends depend on desiring the end for which they are a means – are subject to rational criticism. Smith supposes that rationality potentially has a mechanism for selecting one set of desires over another. Though he does not give a full account of this mechanism he does give some details on what he supposes we know about it or at least what we know about how we conceive it. It is[3] in part a relationship between the contents of the desires in a set of desires that tracks or constitutes the justificatory relationships between the contents of these desires. The relevant feature of desires, what ever it is, is named 'desiderative unity'. Fully rational agents maximise the desiderative unity among their desires. If there are normative reasons – that is reasons to act in particular ways relative to circumstances that are the same for all agents in those circumstances[4] - then it will be because maximising desiderative unity will lead from a population of all sub-fully rational agents as starting points to an idealise fully rational population whose desires overlap in the right way. That is, if there in a normative reason to ϕ in a circumstance C for an agent S then all agents in C will have reason to ϕ and this will be because all fully rational idealisations of these agents will have a

[3] Desiderative rationality is, so far as we conceive it in the way that Smith anti-Humean rationalist meta-ethics argues we do. To be clear there are two ideas in play– one is whether or not anti-Humean rationalism captures the concepts of morality – this is to make as explicit as possible the implicit folk theory of morality and desiderative rationality. The other is whether or not anything in the actual world serves to realize morality so conceived. Moral concepts are a way the world could be (perhaps).

[4] Smith supposes that distinguishing the possible world of evaluation (the one where we are idealized) from the possible world being evaluated (the one where we are sub-fully rational) we can include relevant features of our actual capacities or incapacities where they are morally relevant in specifying circumstances. This is made more explicit in (Smith: 1996a).

desire that any agent in circumstance C will ϕ. Moral reasons will be that subset of normative reasons that have the right kind of contents – effectively, that they recognisably concern what we would now call 'moral matters'.

This view is refined in TMP, and Smith shows that it avoids many relatively obvious criticisms. I take up these details in later chapters as they become relevant. Smith's rationalist position is anti-Humean in the sense that he supposes rationality conditions non-derivative desire formation. But it remains consistent with a belief desire model of psychology and in particular of motivation because only fully rational agents must be able to condition their desires to match their moral beliefs.[5] Actual people who acquire a moral belief would, if they were rational, acquire the relevant motivation or desire but all that Smith's rationalist meta-ethics requires is that they recognise that if they fail to be appropriately motivated then, by their own lights, they count as irrational. So moral judgements are beliefs – about the overlapping desires of a population of fully rational idealisations of all relevant agents. These beliefs will either lead you to be motivated to act in accord with them or you must count yourself as irrational by your own lights. And since the connection between your beliefs and motivations is defeasible in this way there are no necessary connections between beliefs and desires in sub-fully rational agents. When we continue the examination of Smith's theory, we shall see that this lack of necessary connection between beliefs and desires will be preserved in fully rational agents as well (modulo certain complications to be discussed in chapter 6).[6],[7]

[5] I argue in later chapters that even in fully rational agents this capacity to change desires has to be distinct from the features that constitute the norms of desiderative rationality i.e. the desiderative unity feature.

[6] The distinction will be maintained by the fact that the normative component of rationality – desiderative unity maximisation – is not a causal theory. Facts about desiderative unity alone, even in a fully rational agent don't cause that agent to acquire the relevant desires.

1.12 Smith's goal of a defensible naturalism

The positions Smith adopts in relation to the issue of a defensible naturalism is the main focus of this thesis.[8] What a defensible reductive naturalism amounts to is where Smith's anti Humean meta-ethics contrasts with Jackson's moral functionalist framework according to Smith. I will outline Smith's position on defensible reductive naturalism here. Smith supposes that there are two candidate naturalist positions and he provides arguments to reject both of them as inadequate. These positions are definitional network analysis reduction[9] and metaphysical-but-not-definitional naturalism. He offers an alternative that he calls summary style analysis and reduction.

Nor do the beliefs about these facts. Though a fully rational agent must somehow come to have the relevant desires, having the relevant beliefs is not sufficient. So a Humean theory of motivations where beliefs and desires are not necessarily linked is preserved.

[7] It might also occur to the reader that this severs the link needed between the desires of the full rational and the motivations of the agents whose idealisations constitute this fully rational population. A solution, or at least the start of one, is given when the role of 'actual desiderative unity' is discussed in chapter 3, 4, and 5 below.

[8] It initially appears that Smith is explicitly a reductive naturalist and so a defensible naturalism just is a reductive one. However in chapter 4 I consider how Smith argues that fully rational agents can be seen as naturalistic and so allow the squaring the reduction of right to natural properties of acts (properties identified as right because they are desired by the fully rational relative to our circumstances) with a broader naturalism his commitment to this becomes unclear. I argue that the cost of non-reductive naturalism is prohibitive making it an unlikely position for Smith. All forms of ethical naturalism are assumed to be reductive unless explicitly stated otherwise in this thesis.

[9] Reduction can be the reduction of properties or the reduction of terms. Smith discusses the reduction of properties based on the claim that concepts specify ways that properties have to be to be the referents of the concepts. Jackson explicitly focuses on the way representations and their contents (see section 1.2.1 where this is discussed) convey information and express the assertoric components of concepts. Both refer to and in similar ways use the Lewis-Ramsey-Carnap sentence approach to reduce properties, but in so doing accept, implicitly in Smith's case and explicitly in Jackson's, that there is an account of the substantial move from term reduction to the idea that this reliably and systematically constrains property reduction. So though term and property reduction are not equivalent I will along with Smith and Jackson take it that there is a systematic relationship between the two and (except where this might cause confusion) talk of property reduction and term reduction interchangeably.

Smith uses contrasts between summary style analysis and reduction methods and definitional network analysis reductions to explain both so we will do the same.

The dispositional analysis[10] and reduction of colours is the case Smith uses to show what a summary style analysis and reduction is. The dispositional analysis of colour is this: the property of being red (in an object) is the property that causes objects to look red to normal perceivers under standard conditions. Smith says that this analysis is best seen as

> "..an attempt to *encapsulate*, or to *summarize*, or to *systematize*, as well as can be done, various remarks we come to treat as platitudinous in coming to master the term 'red'." (TMP, pgs. 31-32)

Notably the analysis is circular[11]. In part what differentiates Smith's position is his supposition that this circular interdefinition does not need to be removed to permit a viable reductive naturalism in ethics. The summary style analysis found in a dispositional analysis of colour can be used to

> "…readily construct a two-stage argument of the following kind …

[10] The position on colour terms that Smith describes here is very similar to, but also importantly different to the one Jackson gives in FMtE, chapter 4: The Primary Quality View of Colour (see section 2 of this chapter). An important thing to note is that the dispositional theory of colour that Jackson objects to in FMtE chapter 4 is not the theory that Smith describes as a dispositional analysis and reduction of colours being discussed here.

[11] In this thesis we will be discussing circular interdefinitions and non-circular interdefinitions. Since we are frequently considering the interdefinitional relationships between more than just a pair of terms, 'circular' interdefinition will come to mean that the network of interdefined terms of a certain type (for example colour, normative, mental) will, from the conceptual or a priori point of view, have insufficient definitional connections to terms outside the type to secure a unique reduction to naturalist terms. This notion of circularity will serve as a first approximation (various complexities will be discussed later in the thesis).

Conceptual claim:	the property of being red *is* the property that causes objects to look red to normal perceivers under standard conditions
Substantive claim:	the property that causes objects to look red to normal perceivers under standard conditions *is* surface reflectance property α
Conclusion:	the property of being red *is* surface reflectance property α" (TMP pp. 52-53).

The reason that we are focusing on the colour case instead of the ethics one is that Smith uses it to give a putatively uncontentious illustration of his position on reductive naturalism in ethics and to contrast his summary style analysis and reduction from definitional network analysis reductions (to natural properties). Also Smith makes analogy arguments with the colour case when arguing that in ethics definitional naturalism suffers from a permutation problem, that his summary style analysis and reduction of right to natural properties of acts does not, and that the narrow reduction[12] of right to the natural properties of acts can be squared with a broader naturalism. The colour case is put too a lot of work in the course of Smith's construction of a supposedly defensibly reductive naturalist anti-Humean rationalist meta-ethics. A 'squaring argument' for the 'narrow' reduction of colours is an argument that shows how the 'looks red' component of the conceptual premise in the reductive argument for colour can be reduced to physical

[12] What a 'narrow reduction' is will be discussed at length latter. A good grip on it can be got using the reduction of red in objects to surface reflectance property α in the reductive argument above. This reduction is narrow because though it is successful it relies on the undischarged use of 'looks red'. For the reduction of red to a physical property α to be vindicated it must be squared with a broader naturalism. This means a reductive physicalist account of 'looks red' of some sort has to be given.

properties, and so vindicates a broader naturalism (that is, it 'squares' a non-reductive analysis with a broader naturalism). The squaring argument for colours that Smith points to is the reductive physicalism provided by Lewis, 1972, *Psychophysical and Theoretical Identifications*. It is important to notice that in footnote 15 p. 257 is where the reduction of colours and colour experiences is discussed. That footnote points out that there are two circular essentially causal definitions of colour in terms of colour sensations and of colour sensations in terms of colours. Because of the interdefinition the 'meaning' of only one of these terms can be given but not both. For our purposes that is to effectively suppose that the meaning and reference of colour terms can be appropriately *uniquely* specified so long as you assume you can identify and so differentiate colour experiences of sensations and also vice versa. You must assume one or the other, you cannot assume both, and by assuming one you are explicitly not giving a reductive theory of it. Immediately this feature of the colour case suggests that the reductive physicalism is Lewis 1972 is only a part of a squaring argument for a broader physicalism. It also seems to suggest that sustaining the difference between definitional naturalism Smith's summary style analysis and reductive naturalism in the case of ethics will be difficult. However we will simply follow Smith's arguments and see where they lead in much of what follows.

What is so far characteristic of Smith's proposed method is that he supposes summary style circular definitions don't have to be discharged for a defensible reductive naturalism for ethics to proceed. The circularity of a definition does not prevent its use in a two stage, two-premise reductive argument where the circular definition is the conceptual component. The second premise involves what Smith calls substantive claims. In the case of colour terms these are a posteriori facts about what physical properties play

the relevant causal role, the one alluded to in the conceptual premise. This, for colours, yields an a posteriori identity of colours with physical properties. This only works if you have some good reason to suppose you grasp what makes colour sensations or experiences different so that you can use that difference to differentiate colours. Smith argues at some length that this is the case with colour terms because we suppose, and have numerous supporting reasons to continue believing, that our colour concepts are 'hooked up' in the right way to colour experiences and all the other relevant mental state variables. On first blush this plausible claim looks like a good strategy of differentiating colours by their causal roles relative to colour experiences. And when squaring this with a broader physicalism, on first blush this does not seem to block a general reduction of mental states to physical states. But as I have indicated above and as we will return to in the thesis this position is not as clear-cut as we might like in the case of colours especially if it is to serve to differentiate the summary style analysis and substantive reductive argument method of Smith from a definitional network analytic reduction. The transfer of the strategy to the case of ethics faces more, and, so I will argue, ultimately insurmountable, difficulties.

Smith defines metaphysical-but-not-definitional naturalism in the case of ethics[13] as the view that

> "..we use the word 'right' to refer to the property of acts that is causally responsible for our uses of the word 'right'." (TMP, p. 32).

Smith rejects metaphysical naturalism, using Hare's 'cannibals and missionaries' case to illustrate the point, because metaphysical of naturalism

[13] Hereafter, 'metaphysical naturalism' for short.

permits a *possibility* of relativism that folk morality explicitly forecloses. He reformulates the argument by way of making an analogy between how a natural kind term like 'water' permits the possibility of twin earth water/water* cases and how metaphysical naturalism would have to permit the possibility of twin earth right/right* cases (TMP, p. 205, note 7). A twin earth water/water* possibility is just one where two separate communities, using the same reference fixing description for water find that their terms refer to different substances – H_2O and X_YZ respectively. The right/right* case would be one where different properties of acts count for what makes those acts right. Again the problem is that a metaphysical naturalist should describe this as an instance of relativism when folk morality would disagree with that description. I will argue at some length that both of these claims are disputable and that Smith's actual arguments against metaphysical naturalism in the case of ethics are flawed and ultimately can't be rehabilitated. The heart of my arguments is to notice that the role of reference fixing descriptions from appropriate sources is ignored in Smith's argument. When you include an appropriately rich reference fixing description, in the case of ethics got from an explicit expression of our folk moral concepts, Smith's objection to metaphysical naturalism can no longer be made. This set of arguments is found in chapters 5 and 6.

Definitional network analysis reduction in the case of ethics[14] is according to Smith the view

> "…that we can define moral terms exclusively in terms apt for describing the subject matter of the natural and social sciences. The

[14] Here after definitional naturalism for short.

> catch-cry of definitional naturalists is therefore not just analysis, but *reductive* analysis." (TMP p. 35)

This is refined to the claim

> "As I see it, definitional naturalism is best understood as the view that we can come up with a Ramsey-Carnap-Lewis-Jackson style thoroughly explicit and reductive analysis of our moral concepts, a 'network' analysis, as I will call it form here on." (TMP, pp.44-45)

It is against this style of reductivism that Smith formulates the permutation problem objection. The objection to definitional naturalism is that the schema that results from using a Ramsey-Carnap-Lewis-Jackson style reductive analysis underdetermines the referents of ethical terms. Smith uses the underdetermination or reference problem found for the reduction of colours to physical properties when using this approach to illustrate what a permutation problem is and then argues that definitional naturalism in ethics has that problem.

It is worth pointing out here that it is uncontentious that using the Ramsey-Carnap-Lewis-Jackson style of reductive analysis in particular ways will, in the colour case, fail to provide unique referents for colour terms – and given what we know this is a flaw in that approach not in our colour concepts. This permutation problem for colours is solved in Lewis 1972 by preserving the distinction between colour experiences as simply understood in the reduction of colours to physical properties and thereby preventing a permutation problem. In a similar vein Jackson's discussion of what he calls the primary qualities theory of colours is explicit about how the permutation problem that

a Ramsey-Lewis-Carnap-Jackson style reductive analysis is solved by the effectively pre-existing distinction between the mental states that are colour experiences or sensations. This will be picked up in latter chapters. I will call the style of reductive analysis the Ramsey sentence approach. In it you suppose that you have (or can in principle) gather all the platitudes about morality together and state them explicitly. Platitudes effectively express facts about our concepts and so far as these concepts apply to things we can formulate assertoric statements to this effect.[15] The process of a Ramsey sentence analysis then begins by turning all of these statements of our concepts into a format where all the moral terms are reformulated in a property-name way.

> "…'If someone judges her ϕ-ing to be right, then, other things being equal, she will be disposed to ϕ' becomes 'If someone judges ϕ-ing to have the property of being right, then other things being equal, she will be disposed to ϕ'…" (TMP, p. 45).

All these reformulated property-name style sentences are conjoined and effectively form an expression of how our moral concepts made explicit can yield a kind of theory of the moral way the world is.

> "We can represent this as a relational predicate 'M' true of the various moral properties. Where the property of being right and the rest are

[15] This, among many other things is common ground between Smith and Jackson. Unless I flag otherwise we will assume that there is a great deal of congruence between the positions formulated in TMP and FMtE especially concerning the role and nature of a priori analysis and the conditions for a reductive metaphysics. Successful or defensible reductive naturalism for ethical properties in Smith's terms just is the location problem for ethics in Jackson's terms.

represented by the letters 'r', 's', 't' and the like, the conjunction can be represented by:

M[r s t…]" (TMP, p.45)

Then the moral property names are replaced with free variables transforming M to M[x y z…]. And then, as Smith describes it

"…we can say, that, if there are any moral properties, then the following must be true:

∃x∃y∃z…M[xyz…] & (x*) (y*) (z*)…M[x*y*z*…] iff (x=x*,y=y*, z=z* …)" (TMP, p. 45).

This is a way, neither simple nor uncontentious but useful none the less, of representing the idea that we could understand what it would take for a collection of moral concepts, where moral properties are effectively given a variety or roles and relationships by these concepts, to be uniquely realised. This approach is useful when considering reduction because if our concepts are structured in the right way we can use this Ramsey sentence approach to capture what it would take for a successful reduction of moral terms to occur. Smith goes on noting that given the Ramsey style sentence for moral terms above and supposing that the free variable x was the one given the role that the term right plays in this explicit statement of our moral concepts

"…you can define the property of being right as follows:

the property of being right *is* the x such that ∃x∃y∃z…M[xyz…] & (x*) (y*) (z*)…M[x*y*z*…] iff (x=x*,y=y*, z=z* …)" (TMP, p. 46)

And now we can give a fuller version of what Smith takes definitional naturalism in the case of ethics to be, the version required for him to formulate the permutation problem objection to definitional naturalism. Smith thinks the Ramsey sentence approach allows us to formulate what the definitional naturalist supposes is needed to reduce moral properties to natural properties in the form the following argument

"Conceptual claim: the property of being right *is* the x such that ∃x∃y∃z…M[xyz…] & (x*) (y*) (z*)…M[x*y*z*…] iff (x=x*,y=y*, z=z* …)

Substantive claim: the x such that ∃x∃y∃z…M[xyz…] & (x*) (y*) (z*)…M[x*y*z*…] iff (x=x*,y=y*, z=z* …) *is* natural property F

Conclusion: the property of being right *is* natural property F" (TMP, pp. 46-47).

Above I used the qualification that 'if our concepts were structured the right way' we could proceed to a successful reduction using this Ramsey sentence approach. Definitional naturalism would then appear to be the position that the successful application of this Ramsey sentence approach is a condition for any set of concepts, including moral concepts, to be successfully reduced. If this approach doesn't work then we might well be in a position to show that

our moral concepts for example, can't be successfully reduced because there is something wrong with the concepts. Both these claims are probably too strong and Smith does not attribute them to anyone and Jackson in FMtE does not exactly make them either. So perhaps we should simply note that Smith immediately attacks the idea the moral concepts **are** structured in a fashion that allows the use of a definitional reductive analysis of the kind described prior to deploying the machinery of a Ramsey sentence approach. Definitional naturalists suppose that it is possible to give an explication of all of our moral concepts that is sufficiently dense and related in the right way to enough non-moral properties to ensure that a Ramsey sentence approach will succeed in supplying a unique non-moral referent for all moral property terms. The way Smith illustrates what he thinks is wrong with definitional naturalism is to argue that there is a failure of the unique realiser condition in the ethics case. It is meant to be just like the failure of the unique realiser condition in the colour case and this occurs in the colour case because a Ramsey sentence approach, by restricting itself to relational properties, looses the ability to differentiate between colours. This is, in the case of colours, better taken as a failure of the reductive theory than of our colour concepts runs the argument.[16]

So now we can sharpen up to some extent what Smith's views about reductive naturalism in the ethics case are. Firstly he argues that the two main contender types of reductive naturalism (definitional naturalism and metaphysical naturalism) have fatal flaws when applied to the ethics case. He proposes that there is an alternative, circularity tolerant, and defensible reductive naturalism for the ethics case. Smith argues that it is significantly

[16] Versions of this claim are common ground between Smith, Jackson, and Lewis (importantly Lewis since Smith relies on Lewis 1972 for a justification of his distinctive approach in the case of colours as a defensibly reductive approach)

similar to how a reduction of colour terms must be effected. Terms for colours and colour experiences are causally interdefined and apparently ineliminably so.[17] Rather than being pathological, Smith argues that this circular definition:

> Provides useful information for a two stage reductive argument (that, given his arguments about the permutation problem in the colour case, is not available directly to a reductive analysis). (See, e.g., TMP, p. 53.)
>
> Appropriately reflects the epistemology of colour concept acquisition and colour term mastery, which depends on being 'hooked up' directly to colour experiences. (See, e.g., TMP, p. 52.)
>
> And that can be squared with a broader physicalist account of mental states including colour sensations or experiences (the details of which are provided by Lewis 1972). (See, e.g., TMP, pp. 205-206, note 9.)

Smith argues that the approach he illustrates with the his discussion of the colour case serves to show how a similar approach can be adopted in the case of 'right' standing in a circular definition with 'rational' (where 'right' and 'rational' are both relevantly normative) in the context of a none the less defensible reductive naturalism about ethical properties.

---

[17] This is also relatively uncontentious and as we have indicated above and will expand on in latter chapters is a view that can be explained and defended by supposing that what it takes to have colour concepts and effectively use colour terms depends on facts of causal interconnection between mental states, dispositions to make various interferences and the direct wiring of term use to these internal states just to name a few. The basic idea seems both simple and plausible. What is known a priori about the relationships between colour attributions mediated by our use of colour terms in various ways and colour experiences is limited, but not pathologically so, because of how the colour terms and colour attributions happen to be hooked up in the right way to the relevant mental states, amongst other things.

Much of the success of Smith's reductive programme, as expressed in TMP, will explicitly depend on the success of the reduction of colours and colour experiences to physical states, under the assumption that the concepts of colours and colour experiences are ineliminably, and tightly circularly interdefined.[18] So Smith, Lewis, and Jackson on colours will take up various amount of space in this thesis.

Smith's pattern of argument about reductive naturalism in ethics is to define and argue against definitional naturalism, define and argue against metaphysical naturalism, and defend the idea the his anti-Humean rationalist reduction of right to natural properties of acts can ultimately be squared with naturalism when you turn to considering the analysis and reduction of rationality. Smith's attack on definitional naturalism provides the buttress that keeps the analysis and reduction of right from falling into a complex network of analysis and reduction with rationality. My arguments in this thesis consequently start by considering definitional naturalism and then turning to metaphysical naturalism, amassing in the course of these arguments a cluster of problems for the claim that Smith's view in TMP being distinctive, reductive, and defensibly naturalistic. I show that Smith's arguments against definitional naturalism fail. The permutation problem as illustrated by the colour case and the argument from analogy between the

---

[18] We can use Lewis 1972, p. 257 footnote 15, to show the kind of interdefinitions Smith, Lewis, and Jackson have in mind for the colour case:

> "… we should be able to define 'sensation of red' roughly as 'that state apt for being brought about by the presence of something red (before one's open eyes, in good light, etc.)'" And
>
> "… we should be able to define 'red' roughly as 'that property of things apt for bringing about the sensation of red'"

colour case and the ethics case are respectively a flawed illustration and a flawed argument for the same sort of reason. That reason is just that colour and ethics cases are too different given how Smith describes them for the colour case to perform, as Smith needs it to. Smith's final argument – the inductive argument to a permutation problem for definitional naturalism - goes from a weak too an even weaker argument when we realise that Smith and Jackson have importantly different views about the state of current folk morality. This difference prevents the state of folk morality to date playing the evidential role in Smith's inductive argument to a permutation problem since Jackson provides an alternative explanation for and characterisation of that state. In chapters 3 and 4 I refine what is at stake in a permutation problem for definitional naturalism and consider whether Smith's meta-ethical theory has the resources to allow a rehabilitation of the permutation problem objection to definitional naturalism for ethics. As it turns out unless Smith adopts a modified metaphysical naturalist version of his anti-Humean rationalist meta-ethics he can't rehabilitate that part of the permutation problem that actually matters for his meta-ethical theory.

I argue that Smith's arguments against metaphysical naturalism (Hare's cannibals and missionaries argument and the analogy to twin earth water/water* cases) fail because they don't correctly factor reference fixing descriptions into a natural kinds treatment of ethical terms. When this is done using Smith's own meta-ethics as the source of a reference fixing description, Smith's objection to metaphysical naturalism can't be made.[19]

[19] There are many interesting and subtle objections to Smith's anti-Humean rationalism and Smith has published replies to many of them. These include objections to the conceptual links between beliefs and desires, the folk idea that morality involves the exchange of reasons, the ability to provide non-reductive analysis, the source of arguments for cognitivism, whether or not Smith's theory involves objective normative or moral reasons, is reductive, is prone to error theory, … the list goes on: See Bigelow and Smith: 1997; and Smith, 1995, 1996a, 1996b,

1.2 Jackson

Though Frank Jackson's position, particularly about moral functionalism (as it is found in FMtE), is opposed to Smith's views about defensible naturalisms very little about Jackson's theory will play a direct role in my evaluation and criticism of Smith in TMP. Despite this Jackson is a proponent of the kind of analytic reductive methodology that Smith takes himself to be rejecting in TMP. This thesis argues that Smith's meta-ethical theory and its reductive naturalism are not actually distinctively different from Jackson's moral functionalism. On first gloss, Jackson's moral functionalism looks like it just is an example definitional naturalism as Smith defines it. Jackson's views about reductive metaphysics and what he calls the location problem, which make explicit analysis and the application of a Ramsey sentence approach central and indispensible for evaluating what it would take for the world to be the way our ethical concepts suppose it to be, lends credence to this gloss. However, as I will show, this appearance is not perfectly accurate. This is because Jackson's general position on natural kinds and cases like the term 'water' effectively preserve the relevance of a prioristic reductive approaches in considering how to understand the function of natural kind terms and the sentences using them. This will mean that avoiding definitional naturalism, by adopting a metaphysical naturalist version of Smith's anti-Humean rationalist meta-ethics will not suffice to avoid Jackson's moral functionalist framework. At least that is the case if we accept Jackson's arguments that

---

1997, 1998b, 1999, 2002. None of these objections focus on the particulars of broad naturalist reduction in ethics or the details of the role of desiderative unity in Smith's theory, or its potential application as a reference fixing description for a natural kinds treatment of ethical terms, and so they will prove distal to this thesis.

metaphysical and logical necessity are not distinct[20] and that logical necessity and two-dimensional modal semantics are sufficient provide a comprehensive understanding of natural kind terms and the sentences using them.[21] This is contentious in a wide variety of ways. However that makes no difference to me. I am trying to evaluate whether or not Smith's non-analytic reductive naturalism is defensible and distinct from definitional and metaphysical naturalism. I will argue that if Smith is to avoid collapsing into a form of definitional naturalism he must adopt a form of metaphysical naturalism. By Jackson's lights, and arguably but not particularly interestingly here by Smith's lights too[22], a metaphysical naturalist version of Smith's anti-Humean rationalism is compatible with moral functionalism. Still all I argue is that Smith should adopt metaphysical naturalism. It is open for him to take it up in a way that is not friendly to the Jackson treatment of natural kinds and a posteriori necessary identifications.

1.2.1 Serious metaphysics and the location problem

In FMtE Jackson's project is to rehabilitate what he call's serious metaphysics and the role a prioristic considerations play in evaluating the implications of serious metaphysics. His project is contentious and Jackson devotes much to defending it against numerous objections but since I am only interested in the putative contrast between Smith and Jackson's meta-ethics I will not consider these sorts of issues.

---

[20] See FMtE: pp. 67-86, 144-150 and (Jackson: 2010)

[21] Of course the univocally logical nature of necessity and the A and C-intension distinction deployed in two dimensional modal semantics serve to do a lot more than this. But my focus on Smith will not require that these details be given consideration.

[22] Arguably but not necessarily; we will return briefly to this issue in the last chapter of the thesis.

A serious metaphysics is simply any view that supposes a relatively parsimonious bundle of properties, relations, and entities will suffice to give a complete picture of the way the world is, or at least could be. Creating and defending such a view is hard but probably not impossible. Though metaphysics is about the way the world is we will join Jackson in talking about theories about the way the world is. Theories are sentences, usually long, that exhaustively or exhaustively enough represent a way a world can be. Adopting this sentential representational focus is contentious. Jackson argues for doing so in a variety of locations in FMtE. He points out that we use conventions of representation in languages that use discriminations between possibilities to convey information about how we take the world to be and use them ubiquitously. So far as the concepts we have explicitly or implicitly commit us to taking things to be one way or another we can follow Jackson in using the assertoric functions of representational things like sentences to capture and evaluate these commitments. This idea is explicitly defended in FMtE when Jackson provides arguments for understanding natural kind terms and twin earth cases, in their counterfactual format, as being about the modal properties of sentences containing natural kind terms used to describe possibilities. This view is defended more generally in "Why we need A-Intensions", Jackson, 2004, where he extends this approach explicitly to non-counterfactual variants of twin earth possible cases and indeed to all[23] of those representations that serve to convey information – a job that the representational contents of things like sentences do and, he

[23] When discussing the issue of hyperintentionality in his 2004 paper Jackson notes that 'meaning', though an essential ratchet in fixing the representational contents (conceived a divisions among possibilia) for sentences and the like, and 'representational contents' are not the same. Meanings over run representational contents (conceived as divisions among possibilia). He uses an example of different conventions of coordinate designation used to inscribe a circle in the same location to both show how this may be and how it does not adversely affect his view.

argues, a job that amounts to trading in differentiations between possibilities. The basic idea then for a serious metaphysics and a location problem is that, so far as we wish to consider maters can communicate about, we are considering the matter of how claims about one way the world is or could be can make true or false other claims about the way the world is or could be. The details are complex and hotly debated but the view espoused supposes that ways worlds can be should be understood in terms of the properties and relations and entities distributed in them and these can be represented in large sentences whose truth conditions can be cashed out in terms of which possible worlds these sentences are true in. The grain of possibilities you use to do this can vary, though Jackson goes a long way using the relatively crude graining describing representational content thus:

> "For our purposes, what matters is that the right notion of content for understanding how language conveys putative information about how things are is the division among possible worlds one …" (Jackson, 2004, p. 259).

That is that representational contents divides world wise, the group of whole worlds in which the contents are the case distinguished from the rest of the possible worlds in which the contents are not the case.

The Ramsey sentence approach uses the assertoric and property ascription features of theories to use term and property pairs to differentiate theories. An ethical theory will use ethical terms that putatively refer to ethical

properties[24]. A complete (or near enough) descriptive theory will not use any normative terms and will only refer to non-ethical descriptive properties.[25] Evaluating whether an ethical theory can be reduced to a non-ethical theory (and so effectively if ethical properties can be reduced to non-ethical properties) amounts to evaluating whether a purely descriptive theory quantifying over only non-ethical properties can make true an ethical theory.[26] And the Ramsey sentence approach is designed to make that evaluation.

Distinguishing the properties we wish to discuss from how we represent them and then ensuring a non-trivial covariation between the two is an important issue. But Jackson, and Smith for that matter, thinks the job can be done since they both think that the Ramsey sentence approach is useful. A Ramsey sentence approach relates the way one statement (called a theory), given with the use of characteristic and interesting terms – like a statement about the way we take the world to be morally made in terms of moral properties – can be made true by another statement (also a theory) also given using characteristic and interesting terms – like a statement about the way we take the world to be physically made in terms of physical properties. Part of what supports trust in the reliable covariation in term reduction or theory embedding of the kind delineated by the Ramsey sentence approach and

---

[24] This is only 'putative' since a central motivation behind both Jackson and Smith's meta-ethical theories are that they must be reductive. That is ethical terms if they refer at all must refer to properties that are not essentially ethical or normative in nature.

[25] This is why the discussion of reductive issues is often couched in terms of 'terms' – ethical terms, natural kind terms, physical terms. But it is important to notice this is merely convenient shorthand. For Jackson and Smith networks of interdefined terms are the units of investigation when reduction is being evaluated. And these networks can be effectively rendered as sentences.

[26] I use the term non-ethical here. And earlier I used the term non-normative. What is important is that Jackson supposes that we can make sense of a conservative approach to any normative notion associated with any and all ethical concepts and terms. Descriptive terminology is what is left when all of those things are stripped out.

property reduction is just that the notion of properties Jackson is using - which is given in terms of divisions between possibilia (either between possible worlds or if needed in more sophisticated manners[27]). Possibilia are used for formulate theories of the way the world is or could be and to identify and differentiate properties (for one idea of what a property is) but the ontology of properties and possibilia are something Jackson is neutral about. Jackson argues that the ontology of possibility doesn't impact on how he supposes we can use possibilia to formulate the information conveying job description for representational contents.[28]

I will not be defending Jackson's position but I have noted that there is an extended description and defence of the explicitly sentential or representational focus to be found in FMtE and Jackson 2004. It is worth adding that the idea of conventionally encoded information exchange cashed out in terms of divisions among possibilia is **not** a description of the semantics of natural languages. It is a description of a job that natural languages and other representational systems appear to do – that job is to enable the exchange of information. For the same reason it would be a mistake to take Jackson's view about representational contents and other

---

[27] The more sophisticated options crop up when you are considering more sophisticated distinctions. In Jackson 2004 this comes up in discussing what he there calls egocentric content and centred possible worlds. The idea is that possibilia in general, whatever they turn out to be, can be used in a variety of ways to embody, underpin, instantiate or whatever either properties or the truth of property attributions – which is good enough for most of the purposes we have in mind in this thesis.

[28] For a statement of the ontological neutrality for which Jackson is aiming, consider the following:

> "As far as I can see, it does not matter for what follows precisely what ontological view among the at all plausible ones is taken of possible worlds in the sense of complete ways things might be: perhaps they are concrete entities of the same ontological type as our world, as David Lewis holds; perhaps, with the exception of our world, they are abstract entities, as Robert Stalnaker holds; …" (FMtE pp. 10-11)

matters as prescriptive. Neutrality about the ontology of possibilia is just one way the position enunciated in FMtE is not prescriptive. Jackson's theory of representational contents and its use in serious metaphysics and location problems is kind of '*what and one way how*' project. The 'what' is just that we conventionally convey information in representational activities – this is ubiquitous. The 'one way how' part is just the idea that you can use possibilia to track features of this representational economy in a way that allows you to benchmark what an adequate theory of this sort of activity has to have an account of and simultaneously is a one way to start making such an adequate theory.

1.2.2 Jackson on colour

Though I will not use Jackson's position on the reduction of colours to physical properties I will describe it here and contrast it with Smith's description of the dispositional, circular, summary style, analysis and non-analytic narrow reductive argument reducing colours to physical properties.

Jackson calls his view of how to reduce colours the 'primary quality view of colour'. It is reductive and causal and uses the following clause as a schema[29] for the essential causal roll colours play:

> "O is red at *t* iff there is a property *P* of O at *t* that typically interacts with normal human perceivers in normal circumstances to make something that has it look red in the right way for that experience to count as the presentation of *P* in that object…" (FMtE, p. 97)

[29] The claim that what follows are schema is qualified below.

The same sort of schema is used for the other colours. Jackson says prior to presenting this schema the following:

> "First, the primary quality account should regard attributions of colour as *relativized to a kind of creature and a circumstance of viewing*. The primary quality account is the result of combining a causal theory of colour – the view that the colours are the properties that stand in the right causal connections to our colour experiences – with empirical information about what causes colour experiences. And a causal theory of colour takes as fundamental: colour for a kind of creature in a circumstance." (FMtE p. 95)

This is similar to Smith's dispositional analysis of colour. The causal role of colours in both is essential and in both is given relative to 'looks red' mental states. And just like the strategy that Smith adopts from Lewis 1972 Jackson notes of his own that

> "It refers to colour experiences under their colour-experience names, it says nothing illuminating about how to understand colour experience. Once upon a time I was convinced that any adequate account of colour experiences required reference to *qualia* understood as properties over and above those that appear in the physicalists' story about our world. Nowadays I am much more sympathetic to physicalism." (FMtE, p. 101)

The point of leaving colour experiences in play under their colour experience names is that it saves the set of schemas for colours of the kind Jackson

proposes from a permutation problem. So there is an important sense in which these 'schema' are not perfectly schematic for 'looks red' experiences just are not the same as 'looks yellow' experiences and so on. It is common ground that at least from a first person perspective these differences in colour experiences are evident. So the issue is not exactly that colour experiences are a great mystery but rather one about how we go about coordinating our colour talk between people who have colour experiences. The assumption that colour experiences are differentiable from each other and the same for all the members of a relevant population is used to differentiate the colour properties, identified by way of their causal role. With this in mind we can see that physicalist theories of mental states will ultimately have to give an account of colour experiences that has, as a condition of adequacy, the ability to retain the distinctions we assume exist between colour states, the similarities we assume exist between the mental states of at least conspecifics, and finally does this in a fashion that allows us to explain how permutation problems are avoided in coordinating our use of colour terms. For Smith this task is at least partially accomplished on his behalf by Lewis 1972, though the success is dependent on the view that a causal conception of mental states including experience states is sufficient. And just like Jackson, Lewis' theory of psychophysical identifications (in Lewis 1972) explicitly leaves the issue of what an adequate theory of colour experiences or sensations is to one side. Lewis ultimately offers a solution to the problem (or to more of the problem) in his 1997 paper 'Naming the Colours'. Lewis 1997 uses a posteriori and parochial facts about colour attributions to objects as defeasibly[30]

[30] The approach is quiet subtle but for my purposes we only have to notice that Lewis 1997 does some small violence to folk colour theory by using contingent variables like 'our letter box is red' to solve permutation problems. The approach is defeasible because others might use some other object in the same role which just is the role of 'being the coordinating instance of what it is we are talking about when talking about a colour, red say' for various

'definitional' of colours and combines this with the assumption (i.e. a posteriori true if true at all) that the physical properties being tracked between agents making colour attributions are the same and the assumption that the internal states that count as colour experiences are the same kinds of state for members of the relevant population (this last assumption is just the claim that a functionalist causal conception of mental states is correct and, coincidentally, true).[31]

Jackson's most obvious difference with Smith's narrow reduction of colours is that Jackson's a priori clause explicitly relativizes colours to kinds of creature. Jackson says

> "The relativity to kinds of creatures arises from the fact that which properties of the world around us stand in the right relations to certain experiences for those experiences to count as presentations of the properties is, in part, matter of how the creatures having the experiences are, just as which kinds of intruders a burglar alarm latches onto is in part a matter of how the alarm is made, and which weather conditions a barometer records is in part a matter of how the barometer is calibrated" (FMtE, p. 95)

This passage does not give a physicalist theory of mental states, in particular states of colour experience. But it does suggest that Jackson is in a position to adopt a solution like the one found in Lewis 1997. Buy using essentially

---

values of we- and these differences can be accommodated and renegotiated in Lewis' proposal.

[31] This is perforce brief and so skirts over a lot of detail. The assumption that there are the same kinds of states counting as experiential states importantly does **not** use colour-experience names ineliminably as was the case in the 1972 paper. But we will not consider further details about Lewis' position here since they will not directly impact on the thesis.

contingent facts about the kinds of creatures we are to 'name the colours'. Lewis essentially a posteriori facts about local practices of colour attributions and facts about property tracking in the context of colour experiencers being relevantly similar under the hood as it were. Jackson's notion of a kind of creature is not spelled out but the analogies to burglar alarms construction and barometer calibration invite the idea that a physicalist might be able to type colour experiences without in principle ineliminable reference to the colour-experience names. How this could be done is not clear but it could at least copy Lewis' strategy of using a posteriori facts about the kind of animal we are and the kind of properties visual apparatus, as matter of contingent fact, track to anchor naming the colours.[32] We are left with not much difference between Smith and Jackson with regard to the colour case. Perhaps the only thing that might do the trick is that it is plausible that Jackson would defend the idea that all of the a priori and a posteriori facts that are deployed in a physicalist theory of colour will figure in the final theory of colour and be subjected to the same kind of Ramsey sentence approach outlined above to work out the colour identifications. I suspect this is not that promising a distinction but we will return to the matter in the course of this thesis in any case.

1.2.3 Jackson on Ethics

In the case of ethics Jackson's serious metaphysics and location problem amounts to working out what it would take for an account of the way the

[32] This kind of a posteriori identification of colours or colour experiences will probably do some violence to parts of our folk theory of colours – probably the ones that appear to concern rigidifying colours. I have in mind here claims like 'The sky is blue but it didn't need to be' for example. But this is akin to the trade-off against folk theories of colour that Lewis makes to sustain materialism in 'Naming the colours'.

world is given in purely non-normative terms to make ethical claims true. Jackson supposes that the way to evaluate this question in general, and in the case of ethics, is to identify which non-ethical properties are the ethical properties (more generally to work out which of the available realiser properties if any are to be counted as the properties we are trying to reduce) and this involves the use of what we have called the Ramsey sentence approach. On page 140 of FMtE Jackson lays out the same description of a Ramsey sentence approach applied to a complete explication of our moral concepts as the one detailed above by Smith. This locates Jackson as being ultimately committed to what Smith appears to reject when he argues that definitional network analysis reduction is inadequate because it suffers from a permutation problem. Smith appears on the face of it to think that the Ramsey sentence approach and definitional naturalism co-vary – though we have already qualified this above. Something like the following might be more precise – definitional naturalism supposes that the materials for the application of a successful Ramsey sentence approach to naturalist reductions in ethics are available ultimately, from explicit network interdefinitions of ethical terms that are in principle a priori available, and are a best explication of implicit folk moral concepts. Unfortunately it does not seem likely that this is precise enough as a major topic in this thesis concerns the question of how Smith's idea that there are a priori facts about rationality that conceptually determine facts about which properties 'right' tracks does not count as precisely one of these kinds of network analyses and reductions itself. The best we can do at this point is note that the primary example of a non-definitional analysis and reduction, according to Smith, is what he calls the dispositional analyses of colour and the two premise arguments for the reductions of colours to physical properties. And there the concepts appear to rely on components that can only be rendered to refer determinately

enough by adding a posteriori information. Definitional naturalism is then not really about using the Ramsey sentence approach but rather about what kind of information a Ramsey sentence approach is applied to.

Independently of the issues of specifying definitional naturalism how Smith and Jackson take morality to actually be now is significantly different despite a great deal of agreement. Smith and Jackson agree on, amongst other things the following: They agree that meta-ethics is an analytic project that involves novel a priori or conceptual facts. They agree these facts are novel in part because we don't have to have an explicit grasp of our conceptual commitments. This allows for analytic facts to fail to be obvious, to effectively often appear sensibly questionable when in fact trying to doing so is topic changing. They agree that though making an implicit concept explicit is in principle an a priori matter it is often a difficult and protracted matter. They agree that working out whether or not the ethical concepts we have are actually true requires we find successful reductions of ethical properties into non-ethical properties. They agree that this project requires a correct explicit statement of our ethical concepts. At this point the disagreements begin.

Smith, in TMP, is effectively arguing that he is offering the correct explication our implicit folk morality when he formulates his anti-Humean rationalist meta-ethics. He supposes he has unearthed the details of the architecture of our moral concepts. This puts him at odds with Jackson since it supposes the completion of a project Jackson does not think is complete. Jackson distinguishes our implicit folk morality and current expression of fragments of it from what he calls 'mature folk morality'. Mature folk morality is just the best coherent explication of the largest consistent fragment of implicit folk morality. Jackson thinks that until something like a single mature folk

morality comes into play it is not entirely correct to talk about 'our' moral concepts. Thus he is supposing that the contention in moral debate, most plausibly in the reductive meta-ethics part of it, is really contention over what our moral concepts are. And at least part of the explanation of this is the idea that it is quiet possible that on *first* gloss fragments of folk morality can be taken to be inconsistent.[33] This difference has a variety of consequences. Importantly for now it means that Jackson's moral functionalism is not a reductive theory of our ethical concepts – because according to Jackson we have yet to complete negotiating what those concepts are. But moral functionalism, as a schema, does encode some suppositions about how the network of interdefined moral terms could be marshalled prior to using a Ramsey sentence approach to reducing moral properties to non-moral properties. Jackson is committed to reduction in the case of ethics.

Moral concepts are interdefined but Jackson thinks there are characteristic types of clauses or claims our moral theory will throw up.[34]

> "In the case of ethics, we have *folk morality*: the network of moral opinions, intuitions, principles and concepts whose mastery is part and

[33] Jackson qualifies his meta-ethics by acknowledging that it assumes that the end of moral negotiation will yield a single mature folk morality but that of course this might not be the case. If it is not then he thinks the schema of reductive analysis that he describes for ethics will be mature folk moralities relative. Though it does not feature in the thesis it is notable that Jackson takes no view in FMtE of whether fragmentation at the level of mature folk morality is a relativism that should be understood as an error theory. As chapter 6 of the thesis illustrates this sort of matter is an open question and depends on what mature folk morality itself states – which is something we currently don't know according to Jackson.

[34] We should remember that moral concepts are conceived by both Jackson and Smith as implicit things made explicit and, so far as these things involve commitments about ways the world is, they can be made *comprehensively* explicit in a sentence conjoining and expressing all of these commitments to all of the various ways the world is conceived to be morally. That is both Smith and Jackson take moral concepts to have representational components and it is these representational components that the sentences of moral theories express. Hence analysis of our moral concepts issues in the formulation of statements that are effectively folk moral theories.

> parcel of having a sense of what is right and wrong, and to being able to engage in meaningful debate about what ought to be done. … like folk psychology, it contains input clauses, internal role clauses, and output clauses. The input clauses of folk morality tell us what kinds of situations described in descriptive, non-moral terms warrant what kinds of description in ethical terms … The internal role clauses of folk morality articulate the internal connections between matters described in ethical, normative language … The output clauses of folk morality take us from ethical judgements to facts about motivation and thus behaviour…" (FMtE, pp.130-131).

Jackson thinks that moral functionalism is unlike psychological functionalism on two counts. Moral functionalism is not a causal theory. This difference forms an important part of our latter evaluations of Smith's meta-ethical theory and analogies between it and the case of reducing colours.[35] He also supposes that the clauses are more contentious than those of folk psychology – and indeed this is just where and why he distinguishes mature folk morality from folk morality as we discussed above. So Jackson does not effect a definitional reduction of ethical concepts. But he is explicitly committed to doing so when supplied with a mature folk morality with sufficient information. From Jackson's moral functionalist view point Smith's anti-Humean rationalist meta-ethics is a contender for mature folk morality. But the qualification 'sufficient information' is tricky. If a natural kinds treatment of ethical concepts is viable relative to our folk morality then Jackson's moral functionalism is designed to accommodate it. Smith's anti-

[35] Reducing colours to physical states necessarily implicates mental states and the contender solutions for solving permutation problems for colours involve using causal theories of mental state concepts in effecting a physicalist reduction of those mental states. This is discussed at length in chapter 3 and 4.

Humean rationalism can be read, from the moral functionalist point of view, as a candidate for mature folk morality. But it is also open to a definitional naturalist or metaphysical naturalist interpretation according to a moral functionalist view. Finally it is only relative to the assumption that a particular analysis of our moral concepts is correct and sufficient for a mature folk morality that we can, according to Jackson, even begin to evaluate its reducibility.[36]

We can see, from the character of the input, output and internal roll clauses that Jackson expects folk morality to provide a comprehensive distribution of ethical terms among circumstances and conditions that can be described in non-ethical terms combined with all of the internal structure of moral concepts (Jackson provides examples of all these kinds of clause but for our purposes a good example of an internal role clause would just be Smith's conceptual linking of the right making properties of acts to the overlapping desires of fully rational idealisations of a relevant population). These also combine with motivational and behavioural expectations that folk morality suppose covary with the distribution of moral facts. There is a very straightforward way in which these clauses are uncontentious. They are just replicating the plausible, and plausibly platitudinous, features of folk morality. What appears uncontentious is that folk morality expects there to be facts about, for example, which acts are right and that there are moral and behavioural/motivational implications from these sorts of fact. According to Jackson the details are contentious – and since these details are what constitute the network of relations that serve to fix the way things have to be to count as being a moral way at all they effectively constitute folk moral

[36] From Smith's point of view in TMP choosing between definitional or metaphysical naturalism is not much of a choice but it is none the less important that failure of definitional reduction is not necessarily failure of reduction for Jackson's moral functionalism.

concepts. So contention over these sorts of detail is contention over what counts as our moral concepts.

Jackson's down stream commitment to a comprehensive statement of our folk moral theory that is able to have a Ramsey sentence approach applied to it to locate uniquely how moral properties are realised by non-moral properties looks very much like the kind of view Smith calls definitional naturalism. However Jackson in FMtE pp.67-68 argues at length that cases like natural kind terms, which involve necessary a posteriori identities, don't require the use of a distinct kind of necessity from logical necessity. A posteriori necessary identity relations involve metaphysical necessity and Jackson essentially argues that metaphysical necessity is not distinct from logical necessity. This position allows him to deny that natural kind terms and the sentences using them require a characteristically different reductive treatment from that on offer from Ramsey sentence approaches like Jacksons moral functionalist one. This underpins Jackson's argument that, for example, if the reduction of mental states to physical states is successful then a complete statement of the physical way the world is will logically entail the psychological way the world is. So though a metaphysical naturalism for ethics is not the same as definitional naturalism for ethics, it will, according to Jackson (given his view on metaphysical necessity), fit within the moral functionalist framework.

It is interesting to note that since Smith depends on Lewis 1972 to square a narrow reduction of colours to physical properties with a broader physicalism about the colour experiences that differentiate the colour properties he is not opposed to the use of Ramsey sentence approaches when he rejects definitional naturalism since a Ramsey sentence approach is central

to the reductive argument given by Lewis in that paper. Nor, as we saw above, can Smith differ from Jackson because Smith assumes that the colour experience states are different and known to be by those who are in them and that this is simply assumed when effecting all the relevant reductions. Jackson makes just the same kind of assumption. Much of this thesis is devoted to examining what differentiates Smith's reductive naturalism from the ones he rejects. We here will simply grant that Smith's takes his view to be different and then notice that the features of the analysis and reductions of colours are not so obviously as illustrative of this difference as Smith appears to suggest in TMP.

It is necessary to leave out a wealth of detailed argument when discussing Jackson in this thesis because of constrains of space. What has been left out will not directly bear on the thesis. An example is the topic of what Jackson calls global supervenience theses. Jackson describes these theses in terms duplication relations between possible worlds, for example

> "Any world which is a minimal physical duplicate of our world is a psychological duplicate of our world" (FMtE, p.14) or
>
> "For all *w* and *w**, if w and w* are exactly alike descriptively then they are exactly alike ethically." (FMtE, p. 119)

Supervenience theses allow Jackson to discuss some general constraints on reductions prior to considering particular cases. It also allows him to give an account of the asymmetric dependency of what is reduced on what it is reduced too. But all this cashed out in terms of what makes what the case, or what theory can make which set of claims true without additional

quantification over properties to for example. And the ultimate story is about relationships between possibilia so it makes not practical difference to the features of Jackson and Smith's theories being considered here.[37]

1.3 Next

In the remainder of this thesis I will examine how and to what purpose Smith distinguishes his anti-Humean rationalist meta-ethics from definitional and metaphysical naturalist reductive strategies. In TMP Smith combines the illustrative role of his version of a dispositional analysis and reduction of colours and analogies to that case to outline a putatively alternative, defensible, reductive, naturalism in ethics with arguments against the two main contender naturalisms (definitional and metaphysical). That is, according to Smith, since both definitional naturalism and metaphysical naturalism in ethics turn out to be indefensible we have to opt for some alternative that does not have the flaws of these two options. Summary style analysis and two stage non-analytic reductive arguments, as illustrated in the dispositional analysis and reduction of colours, is offered as the alternative.

This thesis challenges the arguments against the alternatives that Smith offers, considers and ultimately in both cases rejects the idea that some version of Smith's arguments can be created using the resources available in TMP, and finally concludes that his anti-Humean naturalism faces choosing between failure as a form of naturalism, adopting definitional naturalism with the immediate burden of having to provide a theory of desiderative

[37] The idea that supervenience relationships can be described or used as Jackson uses them is contentious. Certainly it seems like Jackson has a very specific use of Global supervenience relations in mind.

unity (which is not available), or adopting a form of metaphysical naturalism. The last option, I argue, is the best.

The next chapter will look at the permutation problem arguments against definitional naturalism. Smith uses the permutation problem for the physical reduction of colours, given the tight (circular) interdefinition of colour concepts, to illustrate what a permutation problem is and how a definitional naturalism in ethics suffers from a similar problem. To show the latter, he uses an argument from analogy to the dispositional analysis of colour as well as an inductive argument. The chapter argues that both arguments fail.

Chapter 2 Smith and the permutation problem for definitional naturalism

In this chapter my aim is to examine Smith's arguments that definitional naturalism suffers from a permutation problem and to show that they don't work. He begins with another case, the case of colour, and proposes that we replace definitional reductive physicalism in the colour case with what he calls a dispositional analysis of our colour concepts and a two stage argument for the reduction of colours to physical properties[1]. His aim is to illustrate how the permutation problem for definitional physicalism in the colour case can be solved, and he argues that a somewhat similar strategy can be adopted in the ethics case, leading to a summary style network analysis and reduction – summary style naturalism for short.

In section one I will give Smith's descriptions of the colour case and how it is meant to illustrate a permutation problem. In section two I will argue that there are significant disanalogies between the colour case and the ethics case. In section three I argue that Smith's inductive argument to a permutation problem is weakened significantly when you realise that the evidence for the induction, the state of meta-ethics to date, is described and used differently by Jackson. I will conclude that Smith's arguments for the permutation problem fail. I will note that the feature of Smith's anti-Humean rationalist meta-ethical theory that a permutation problem argument supplies is a reason to keep the analysis and reduction of 'right' separate from the analysis and reduction of 'rational'. This separation or gap is what stops Smith's anti-

[1] I will call an argument that concludes with an identity between a property P_1 described in non-descriptive terms and a property P_2 described in descriptive terms (for example, a statement identifying the colour red (in objects) with the surface reflectance property α) a reductive argument.

Humean rationalist meta-ethical theory from collapsing directly into an example of a definitional naturalism. Given the loss of the permutation problem argument, the following chapters will consider whether or not some other reason for this separation can be found from within the resources of Smith's meta-ethics.

2.1 The colour case

We get from Smith that it is analytic (at least in the sense that there are platitudes to this effect) that

> " … the property of being red causes objects to look red to normal perceivers under standard conditions, and the property of being red is more similar to the property of being orange than it is to the property of being yellow, and so on. …. the property of being orange is more similar to the property of being yellow than it is to the property of being green and so on. … the property of being yellow is more similar to the property of being green than it is to the property of being blue, and so on." (Smith, TMP, p48-49.)

Each colour causes a perceptual event, an experience of the colour in normal perceivers under standard conditions and each stands in a network of degrees of similarity relations to other colours. Smith then, for the sake of illustration, assumes that this structure of causal and similarity relation claims exhausts the a priori information about colours. As he puts it

"Let's assume that there are no platitudes about the colours that entail any claims beyond these about the properties of being red, or orange, or yellow, or the rest. … with this assumption in place look at what happens if we construct network analyses of the various colour properties by simply conjoining all the platitudes about these and following the Ramsey-Carnap-Lewis-Jackson procedure. Simplifying somewhat, we find that:

the property of being red *is*	the x such that ∃y∃z … objects have x iff they look x to normal perceivers under standard conditions, and x is more similar to y than it is to z … & … (uniqueness conditions) …
the property of being orange *is*	the y such that ∃z∃v … objects have y iff they look y to normal perceivers under standard conditions, and y is more similar to z than it is to v … & … (uniqueness conditions) …

… and so on" (Smith, TMP, p 49.)[2]

Smith points out

"…now look at the network of relations specified by the definitions on the right hand side. In each case it is the *very same* network of

[2] When 'looks red' is replaced with 'looks x' we drop out information and that information just is whatever it is that differentiates the colour experience states from each other and identifies them as the particular ones they are. This is possibly puzzling since it appears that Smith is insisting that definitional physicalists do not or cannot pay attention to the role mental states or at least qualitative properties of mental states play in determining colours in objects. We will examine this in more detail in Chapter 4 when we discuss the significance of the Lewis 1972 paper to Smith's position.

> relations. And what this means is, in essence, that in our definitions we have lost the distinction between the properties of being red, being, orange, being yellow and the rest. The uniqueness requirement thus cannot be satisfied." (Smith, TMP, p 50.)

and again we have

> "In short, then, because the claim to uniqueness is false, we lose any reason to believe that our network analyses of colour concepts allow us to pick out a unique set of physical properties to identify with the colours. Moreover, the problem here lies not with the world – we have not just demonstrated that there are no colours! – rather the problem is with the network analyses themselves. Thoroughly explicit and reductive network analyses of our colour concepts lose *a priori* information about the *differences* between the colours. They are therefore defective, as analyses.
>
> Let's call this the 'permutation problem'." (Smith, TMP, p. 50)[3]

This illustrative case is perhaps uninteresting if for no other reason than that we can recreate a permutation problem for a physicalist reductions of colours without using the restriction on a priori information to resemblance relations between colours. It seems tempting to suppose that without this restriction we would not have a permutation problem for colours. As it happens this is

[3] The claim that a definitional analysis drops out a priori information is also a puzzle though perhaps the same one as the last. It might be that certain claims about the causal roles of colours, mental states and associated qualitative states are all a priori, but though it might be a priori that there are differences in qualitative states and the mental states associated with them that individuals have a priori access to, moving from this to any other more complex set of claims that include other perceivers with the same or near enough the same kind of mental states in systematically similar circumstances is not a priori. Again the mater receives attention in chapter 4.

arguable. More interesting is that the ethics case as Smith describes it involves no resemblance relations and on the face of it only two normative terms – 'right' and 'rational'. Whatever permutation problems arise in the ethics case they will not be like the one described by way of restricting your attention to resemblance relations among colours.

Finally we should be very cautious about the use of this particular form of the illustration. By restricting the a priori facts to the resemblance relations between the colours we generate a permutation problem. But that restriction renders the illustration useless as a diagnostic tool. So when Smith tells us that

> "Thoroughly explicit and reductive network analyses of our colour concepts lose *a priori* information about the *differences* between the colours…" (TMP p. 50)

and that we can't use the Ramsey sentence approach to an analysis restricted to the relational properties between the colours to work out what a priori information about colour differences is lost we should be unsurprised. We have not established with this form of illustration of a permutation problem for the colour case that a Ramsey sentence approach must always fail. Nor should we be surprised that a Ramsey sentence approach is not able to retrieve the supressed a priori distinction between the colours – it is a tool for generating reductive identifications of properties that requires the supply of a theory, a priori or otherwise. Definitional or analytic reductive physicalism, if there is any such theory, is not the same thing as using the Ramsey sentence approach as a tool for effecting reductive identifications given a theory. What then is going on in the dispositional analysis and reduction of colours such

that we should, with Smith, *resist* the idea that a network definitional reduction of colours to physical properties will work?

2.1.2 The dispositional analysis and reduction of colours

The information that is lost in the definitional or analytic reduction of colours to physical properties isn't that colours stand by definition in particular causal relations to experiences, nor is it that experiences are defined in part by this causal relationship. Though the illustration might appear to suggest as much, actually that information is included. What is dropped is just whatever it is we use ourselves to name the colour-experiences as such. Smith supposes that we are such that we can know for ourselves and we assume for others that there are 'looks like' experiences that we simply do differentiate and reliably name as such and being able to do so is what enables us to track colours. Then non-reductive analysis, Smith's dispositional analysis, has the advantage of explicitly reflecting this fact

> "Colour terms thus seem to elude analysis in the thoroughly explicit and reductive style of network analyses. There should, I think, be no real surprise here. For we learn colour terms in part by being presented with paradigms of the various colours, paradigms which, for us, fit within a natural visual similarity space. In acquiring mastery of colour terms, we then acquire a disposition to judge visually presented cases of particular colours to be the particular colours that they are (Peacocke 1985). Having mastery of colour terms is thus, in part, a matter of having our use of colour terms directly 'hooked up' with the colours these colour terms pick out. For this reason, the

> conditions for mastery cannot be fully spelled out without stipulating that these direct links are in place. … Indeed the non-reductive character of the dispositional analysis may now seem to be a distinct advantage. For its non-reductive character can be seen to reflect the fact that in having mastery of colour terms, our use of those terms must be directly 'hooked up' with the colours that those terms pick out, something a thoroughly reductive analysis seems bound to ignore, or to capture only inappropriately, to its peril (Smith, 1986a)." (Smith, TMP, p 51-52.)

The idea that our usage of colour terms and possession of colour concepts is 'hooked up' with the colours those terms pick out is where the a priori information that analytic reduction misses is to be found. However, even this is a point that can be misleading. In the paragraph opening this passage I pointed out that it seems plausible that we have some first person way of distinguishing the colours, most plausibly because we distinguish somehow the colour experiences they are associated with. This is possibly a priori information that an analytic reduction of colours will have some difficulty accessing and using. But the assumption that this feature is common between users of colour terms and is how we prevent permutation problems is not obviously unavailable to a reductive analysis of colours. Also moving from first person, introspectively available and so a priori, information to publicly coordinated use of colour terms is not a trivial step.[4]

Smith presents his dispositional analysis and two stage reduction of colours as follows

[4] See Lewis: 1997, where the public coordination of colour term use is taken to be the feature that admits of permutation problems and that Lewis aims to provide a solution for.

"Conceptual claim: the property of being red *is* the property that causes objects to look red to normal perceivers under standard conditions

Substantive claim: the property that causes objects to look red to normal perceives under standard conditions *is* surface reflectance property α

Conclusion: the property of being red *is* surface reflectance property α"

This argument provides a narrow reduction of red to physical properties. It is narrow because the 'looks red' experience state is simply used unreduced. What is needed is an account of how the 'looks red' experience state can be reduced to a physical state itself (what Smith calls squaring the reduction of red with a broader physicalism), and counts it reasonably as a requirement for vindicating the apparent physicalism of the narrow reduction of red to the physical property α. The conceptual claim contains the summary style analysis of colour concepts. The idea is that this circular analysis captures, by way of summarising conceptual facts about the colour red and its relationships to experiencers, the colour-experience associated with red (the 'looks red' experience) and the impact of conditions of exposure on the relationship between instances of 'red' and 'looks red' experiences. The substantive claim is a posteriori.

Smith claims that

> "To have good reason to believe the premises of this two-stage argument we have to draw upon our prior understanding of the

> concept of being red, our prior beliefs about which objects would look red to normal perceivers under standard conditions. But, of course, that is neither here nor there given that our epistemic situation is one in which we do have such prior knowledge, and given that our interest in putting forward such an argument is squaring colour talk with physical talk." (TMP, p 53).

Here Smith is alluding to the idea that, at least individually, we are putatively hooked up to the 'looks red' experience state in the right way, and this state in turn is causally embedded in the right way for our colour attributions to objects work reliably. But the equivocation between individual epistemic states and the coordination of colour attributions occurs again rather than being resolved. That we severally have prior knowledge does not show that we collectively have it and prior knowledge is not necessarily a priori knowledge. It is clear that the summary style analysis makes an assumption about commonalities between people who can make *correct* colour attributions. But it is not entirely clear what this assumption amounts to because we have to make repeated use of the colour-experience name 'looks red' to describe what is going on in our assumptions, including the assumptions that our experience states are relevantly similar. 'Red' is defined causally in relation to a colour experience named as the particular 'looks red' colour experience. On the assumption that 'looks red' experiences and 'looks yellow' experiences and so on are all *assumed* to be uniquely specified – which is what continuing to use the colour-experience names 'looks red' and so on effectively does – then we can stabilise the causal definition of 'red' relative to these experience states by guaranteeing that the colours are uniquely realised. The advantage of the circular definition of 'red' in terms of a causal role relative to 'looks red' colour experiences is that this sort of definition

eliminates the permutation problem. It does replace it, however, with the altogether pressing issue of how assuming colour-experiences differentiated as such can be made consistent with a broader physicalism. It seems that the problem of what information is dropped by definitional reductions of colours involves distinctions between experience states that might plausibly be a priori for each individual but then the problem pops up again in the question of how these differences make it from facts about individuals to facts about **our** colours and colour experiences. We are forced to consider how Smith proposes to square the narrow reduction of colours to physical properties with a broader physicalism to find the answers to these questions.

In footnote 9 chapter 2 TMP Smith says

> "It might be thought that we don't yet have an argument that would allow us to square colour talk with a broader physicalism *per se*, as the argument just given has no bearing on whether a subject's experience of having something look red to her is itself a physical state. But the forgoing discussion [p 52-53] suggests an obvious strategy for squaring talk of colour experience with physical talk as well. The first step would be to construct an analysis of our concept of a colour experience. The second stage would be to show how these analyses allow us to identify colour experiences with, say, states of the brain. If, as seems plausible, our concept of colour experience is the concept of a state of a subject that, in conjunction with relevant desire, causally explains our bodily movements – for example, our picking out red objects from objects of other colours – then it should be clear enough how the attempt at vindication would go, and why it should be

deemed likely to be successful (compare Lewis, 1972)." (TMP, p 205-206).

In this footnote Smith gestures at a causal definition of mental states and then uses the causal role of picking out red objects to fix which causally defined state counts as the red experience. In effect he is repeating the strategy found in Lewis 1972 'Psychophysical and Theoretical Identities'. That strategy is couched in terms of the reduction of theoretical terms to 'old' terms. The idea is just that a Ramsey sentence approach will show you how to cash out the introduction of new terms in the format of a theory specifying the roles, relation, properties and whatever else of the things referred to with these new theoretical terms using pre-existing or 'old' terms. The reductive pattern and resultant identifications, of for example of the properties associated with theoretical terms, are general. Lewis explicitly restricts his account to those theories that are explicitly causal theories. The Ramsey-Lewis-Carnap method of the reduction of networks of interdefined theoretical terms is a model for how to reduce an interdefined network of mental state terms to physical state realisers. The key relevant passage is footnote 15. There Lewis expands on the observation that though mental state terms did not get introduced into our language as theoretical terms, in the sense his paper is describing them, it is none the less useful to treat them as if they had been since what the names of our mental states mean actually is what they would have meant if they had been introduced as theoretical terms. He says

"Part of my myth says that names of color-sensations were T-terms, introduced using names of colors as O-terms. If this is a good myth, we should be able to define 'sensation of red' roughly as 'that state apt for being brought about by the presence of something red (before one's

> open eyes, in good light, etc.)'. A second myth says that names of colors were T-terms introduced using names of color-sensations as O-terms. If this second myth is good, we should be able to define 'red' roughly as 'that property of things apt for bringing about the sensation of red'. The two myths could not both be true, for which came first: names of color-senstations or of colors? But they could both be good. We could have a circle in which colors are correctly defined in terms of sensations and sensations are correctly defined in terms of colors. We could not discover the meanings of both of the names of colors and the names of colour-sensations just by looking at the circle of correct definitions, but so what?" (Lewis, 1972, p. 257)

So how does this give us a reduction of colour sensations to physical states? It does so by assuming that all of the members of a relevant population ('humans' for example as we saw Jackson suggest in his discussion of colours) have the same psychological architecture that, amongst other things, is sufficiently conceptually described with a causal theory of mental states and associated properties and that they, relative to colour sensations, track the same properties in objects. With this kind of assumption in play you can use the reductive identification of red to anchor the identification of red sensations or vice versa. The trade-off, or at least a price, associated with this procedure is that it is not clear that you can ever show that the assumption of homogeneity between perceivers can ever be discharged. This is at least in part due to how the undischarged interdefinition has to simply assume that the colour-experience terms refer uniquely and appropriately homogenously to supply a useful causal conception of colours and so allow an empirical reduction of colours to physical properties. At least, that is, in the case of the narrow reduction of colours.

For Lewis the meaning of theoretical terms is given by their functional causal roles. When you use the distinction between colour-sensations to give the meanings of colour terms you are thus giving a causal role specification of what the it takes for a property to be one colour or another. You assume that the colour sensations are sufficient for this task but you cannot give them a 'meaning' while using them in this role. From the point of view of a causal theory of mental states the obverse strategy is just as good. You can assume differences between colours sufficient to uniquely specify colour sensations and thereby have a causal relationship to colours giving the meaning of colour-sensation names. But then you cannot give the colour terms a meaning while using them in this role. What the Lewis 1972 paper argues is that this approach necessitates reductive identifications and is a viable form of physicalism about mental states. What it leaves out of the picture is the complete theory of mental state differentiation. One way to get a handle on what is left out is that the theory in Lewis 1972 does not explain how the causal facts about the internal states of members of a population make it into the role of moderating public use of colour names. This is the feature of the problem that Lewis considers and takes up in his 1997 'Naming the Colours'[5]. As we discussed above Jackson notes that a theory of colour experiences is still owed after using the assumed differences between colour experiences to avoid a permutation problem for his own primary qualities theory of colour.

[5] We discuss some of the details of this paper in chapter 4. Lewis sacrifices the contingency of local paradigmatic colour attributions to particular objects to anchor the assumption that we are tracking the same property with our colour talk in a way that simultaneously avoids a permutation problem and allows the coordination of conventions of colour naming. The local or parochial definition of colours relative to local particular things is defeasible but none the less somewhat at odds with some of our folk theory of colour.

So we are left with what appears to be the following: Smith acknowledges that the narrow reduction of colours to physical properties requires an argument to show how colour experiences (the 'looks red' state for example) can be reduced to physical states. He refers us to Lewis 1972 who shows how a circular definition of colours and colours sensations (colour experiences) can permit a physicalist reductive identification of mental states with physical states so long as you accept that to do so you must accept (in Smith's case) the meanings of the names of colour sensations cannot be given. When Smith sketches how a reductive causal account of **red sensations** can be given he does so relative to which mental states and behaviours the **colour red** is typically causally associated with. When Smith does this, so far as he is using (Lewis: 1972), he is in effect dropping the dispositional analysis of colours and simply assuming that the colour properties are stipulated as the particular colours they are. In effect Smith's squaring argument for the physical reduction of colours appears to be only partial. Smith does not think this is objectionable and the Lewis paper goes some way to showing why this is.[6] But then it becomes unclear that an analytic reductive physicalism that does not suffer from a permutation problem is not possible. And this is just because the Lewis 1972 paper openly provides only a partial account of how reductive physicalism in the case of colours can be squared with a broader physicalism. This leaves open the possibility that a *complete* reductive theory of mental states could identify all the mental states in a reductive definitional network analysis.

[6] The argumentative strategy is just to show that the inter definition of colours and colour sensations block an account of the 'meaning' of both simultaneously but does not consequently block a physicalist reduction. But as I have been noting the cost is that the reductive physicalism is in one sense or another partial.

2.2 Is the colour case a good illustration of a permutation problem?

When we restrict ourselves to the relations of resemblance between colours a permutation problem seems easy enough to illustrate. But the restriction is questionable. The permutation problem for colours is more about the prospects of analytic reductions using more realistic candidates for the analysis of our colour concepts. Smith's dispositional analysis of colours and separated two premise argument for a physical reduction and Jackson's primary qualities theory of colours are just such theories. And both are designed with a mechanism for avoiding a permutation problem built in. And that mechanism looks like it is the same mechanism – using an assumption that the colour experience states are differentiable and identifiable in just the way required to avoid a permutation problem in the case of colours.[7] And both seem to suppose that there is further work required to square the reduction of colours with a broader reductive position about mental states. The only problem with this colour case now is how to use it to illustrate that definitional network reductions in the case of colours is prone to a permutation problem. At least one possibility is that an account of mental states that assumes a causal conception of them is adequate will have to use a posteriori features of the world, agent, populations or a mixture of these sorts of facts to successfully use folk concepts of colour and mental states in the course of a reductive account of them. And Jackson's only statement in FMtE relevant to this matter is that he used to think the relevant differentiations had to be made using non-physically specified ineliminably qualitative properties and in FMtE he has changed his opinion to one more

[7] And this is not so tricky really. It is just the assumption that 'looks red' experiences are correctly named and not the same as the other colour experiences and so on. That is, it is the assumption that there are unique realisers for colour experiences and that which is which is known.

sympathetic to physicalism. This is open ended. However Jackson does relativize colours to kinds of creature. Which kinds of creature we humans are is plausibly a contingent matter so perhaps Jackson's ultimate solution to the question of how to reduce colours will be one that makes use of a posteriori components in a central conceptual role. Then his position will be like Lewis' in Lewis: 1997 and so not an example of definitional physicalism for colours.

All in all, the illustrative role of the colour case when talking of a permutation problem is questionable and unclear. When the case is clarified the causes of the permutation problem become clearer but the claim that definitional network analysis reductions of colours to physical properties are particularly prone to them becomes rather unclear. We will simply go along with Smith for the sake of argument. We can agree that the reduction of colours does give an example of a permutation problem. We can also agree that if there is a permutation problem then it has to be removed and that in the colour case at least the strategies for removing it are not obviously part of a definitional reductive physicalism. This is because the candidate strategies all involve essential a posteriori information to remove the permutation problem. If the reduction of ethics to natural properties displays a permutation problem like that found in the colour case then at least definitional naturalism would have difficulties removing it.

2.3 Flaws with the analogy between the colour case and the ethics case

Smith makes frequent and heavy use of the colour case in the course of formulating a wide variety of components of his thesis. Though he points out

that reduction of colours restricted to resemblance relations was created merely for illustrative purposes he does make an argument from analogy between the unrestricted colour case and the ethics case to the conclusion that there is a permutation problem for definitional naturalism in ethics.

Smith claims

> "…the permutation problem arises for two related reasons. It arises because, first, we acquire mastery of colour terms inter alia by being presented with paradigms of the colours and by having our use of particular colour terms directly 'hooked up' with the particular colours these terms pick out, and because, second, as a consequence, the platitudes surrounding our use of colour terms therefore form an extremely tight-knit and interconnected group. The permutation problem arises because our colour concepts are not defined in terms of enough in the way of relations between colours and things that are not themselves colours…"(Smith 1994, p55)

On page 55 of TMP Smith precedes his inductive argument for a permutation problem for definitional naturalism in ethics with an analogy between the colour case and the ethics case. The analogy is that both sets of concepts are learnt by way of presentation of paradigms. He says

> "Just as we learn what the colours are by being presented with paradigmatic instances of the colours, we learn what a good argument is by being presented with paradigmatic cases of good arguments, we learn what rightness is by being presented with paradigmatic cases of

right actions, we learn what wrongness is by being presented with paradigmatic cases of wrong actions, and so on." (Smith, TMP, p 55).

Paradigmatic concept acquisition does not **entail** a permutation problem of whatever sort. It does not imply that a definitional analysis and reduction is impossible. Smith does notice that paradigmatic concept acquisition does imply that such concepts are interdefined. He says

> "The platitudes surrounding our use of normative terms generally, and thus moral terms as well, therefore form an extremely tight-knit and interconnected group. Such terms are largely interdefined. Perhaps the most striking way of bringing this out, in the case of moral terms, is by focusing on the various platitudes about procedure: that is, the various descriptions of the ways in which we justify our moral beliefs, what Rawls calls the method of 'reflective equilibrium'. For it is hard to believe that, once all normative terms are stripped out of these platitudes, there will be any determinate content left to them at all. And the loss of such content is just what makes for a permutation problem." (Smith, TMP, p 55).

Finally Smith says of normative concepts and reasons that he thinks they, like moral concepts are vulnerable to a permutation problem because

> "…what the discussion of colour concepts shows is that permutation problems arise when a set of concepts, acquired *inter alia* via the presentation of paradigms, is therefore largely interdefined. Permutation problems arise when there are very few concepts outside the circle of concepts to be defined playing a significant role in the

> platitudes we use to state an explicit definition of those inside the circle. And, of course, this is precisely what we find with our normative concepts; they are indeed *largely* interdefined. Very little outside the sphere of the normative is required to define the normative. And again, as with our colour concepts, this is because we learn our normative concepts by being presented with paradigms – paradigms of good arguments, of what it is for one proposition to support another, and so on – from which we learn to generalize" (TMP, pp.163-164)[8]

So the permutation problem arises because of the direct hooking up of colour concepts to colours (roughly) and because of a too tight interdefinition of various colour terms. These facts seem to be implicated in how it is that we learn colour term use and colour concepts more generally by way of exposure to paradigmatic cases. Ethical terms and concepts are also typically learnt by way of paradigmatic examples. This does not entail a permutation problem but does tend to covary with the tight interdefinition of terms. Smith thinks an example of this too tight definition is the way the 'platitudes of procedure' refer to the Rawlsian reflective equilibrium for justifying moral beliefs. Removing all the normative terms describing a Rawlsian reflective equilibration will leave insufficient information to elicit a definitional reduction. The Rawlsian turn in this argument is somewhat misleading, I think. The only type of reason canvassed in the colour case for too tight an interdefinition of colour terms is the dispositional analysis of colours – an ineliminably circular definition of colour terms in terms of colour experiences. The Rawlsian reflective equilibrium will lead to a permutation

[8] Given the way normative concepts determine the reference of moral ones, it follows that for Smith's anti-Humean rationalist meta-ethics this argument does not give a case for two permutation problems, one for ethical concepts and one for normative concepts.

problem if there is an ineliminable interdefinition of the ethical terms involved. Though it seems that Smith believes as much he makes no argument for it here. Moreover as we see in later chapters of this thesis, particularly chapter 3, the role a Rawlsian reflective equilibrium plays in explaining or determining the normative components of Smith's ethical theory is severely limited. So if any argument at all is indicated by the last quoted passage then it seems it is really an argument from analogy, perhaps an argument along the following lines: In the colour case there is a permutation problem. It arises because of the 'direct hook up' feature of colour concepts and the too tight interdefinition of colour terms or concepts – the dispositional analysis of colour shows this too tight interdefinition is manifested as the ineliminably circular nature of a priori available colour definitions. These facts co-vary to some extent with the paradigmatic learning of concepts. Ethical concepts are learnt by paradigmatic examples and appear to show in places a tight interdefinition of normative terms. This is enough similarity to make supposing a permutation problem in the ethics case persuasive.

I have two disanalogy objections to this argument. First the colour case and the ethics case differ in the modality of the substantive premise of their respective two premise reductive arguments for their narrow reductions. Second the colour case and the ethics case differ in that there are squaring arguments in the colour case and none in the ethics case, and the squaring arguments in the colour case appear characteristically a posteriori (it is unclear if the like is possible in the ethics case given the first disanalogy).

2.3.1 First disanalogy between colours and ethical terms

This is simple enough. In '*The Moral Problem*' pg. 185 we have:

> "Conceptual claim: Rightness in circumstances C is the feature we would want acts to have in C if we were fully rational, where these wants have the appropriate content
>
> Substantive claim: Fness is the feature we would want acts to have in C if we were fully rational, and Fness is a feature of the appropriate kind
>
> Conclusion: Rightness in C is Fness"

Smith thinks that the substantive claims in the case of ethics are all a priori. And we have seen that he thinks that the substantive claims in the reduction of colours to physical properties are a posteriori. In TMP pp. 190-193 Smith notes of a pair of arguments about colour and right actions respectively that rely on this disanalogy[9] and he says the following

> "There is, of course, a point of disanalogy between the two arguments just given, and it is worth while making this explicit.

[9] The arguments are not the general reductive ones here but the properties they have rely on the properties of these more general reductive arguments. In the colour argument we go from x has surface reflectance property α to x is red with a mediating premise that α is the property that actually causes objects to look red typically. In the ethics argument we go from giving to famine relief in a circumstance C to giving to famine relief in circumstance C is the right thing to do in C. Here the mediating premise is that giving to famine relief in C is the feature we would want act to have in C if we were fully rational and that want is the appropriate substantive kind. Both mediating premises are expressions of the general reductive arguments for colour and right making properties of acts respectively.

> Whereas, in the moral case, the second [mediating[10]] premise of the argument is not just necessary, but also knowable *a priori*, the second premise of the argument in the colour case, thought necessary, is itself only knowable *a posteriori*. However this feature of disanalogy should not hide the more striking features of analogy already mentioned. For in neither case is the second [mediating] premise a matter of definitional equivalence."
> (TMP, p. 192)

It might seem unfair to use an argument created after Smith supposes he has established that there is a permutation problem in the ethics case. I think not. The anti-Humean rationalist ethics uses facts about rationality to determine facts about right action because analysis of right action indicates, according to Smith, that we should. Smith also argues at length that the relevant facts about rationality – its analysis and the particulars of its dependence on the feature of desiderative unity and how this will determine in particular the desires that fix right action – are a priori facts. Later in this thesis I will argue that Smith should abandon this position but until he does it presents a problem. If all the facts about rationality down to the particulars of its effects on the contents of desires are knowable a priori then we should wonder why these facts are not part of the definitional network analysis and reduction of right. They are conceptually connected, the precise nature of the connection is explicitly known in principle a priori, as are all of the details about how rationality determines the contents of ideal desires and so determines which features of acts make acts right, relative to circumstances. The question is what else would be required for a definitional reduction of ethical properties to non-ethical properties?

[10] Ibid

2.3.2 Second disanalogy between colours and ethical terms

This relevant disanalogy between the colour case and the ethics case arises in the differences between the squaring arguments for each case. Setting aside the issue of what it takes to be an example of a definitional naturalism in ethics we should note that Smith supposes that in both the colour case and the ethics case, because of the use of uneliminated interdefinitions of importantly similar terms[11], we have to vindicate their respective narrow reductions[12] by showing how they can fit with broader reductive theories. In the case of colours there is a controversial but supported view that causal conceptual theories of mental states are correct. There is a scientific theory providing plausible candidate realiser states for these causal theories and there are philosophical theories about how to wed the two. These theories are also in important ways essentially a posteriori theories and their details serve to explain both the paucity of a priori information about colours and colour experiences and also to resolve that paucity of information with a physicalist reduction of the relevant states (mental states in the colour case). The offered squaring argument is arguably partial (Lewis 1972) but none the less enough to show how reductive physicalism might proceed. Two important features of this background of theories is that they explain why ineliminable circularity of interdefinition of terms for colours and terms for colour experiences occurs and show to some extent how essentially a

[11] With colours they are causally conceived interdefinitions of colour terms with terms for colour experiences. With ethics they are the relevantly normative terms 'right' and 'rational'.
[12] The narrow reductions are, in the colour case, the reduction of colours to physical properties of objects and, in the ethics case, the reduction of right to natural properties of acts, relative to circumstances.

posteriori approaches can resolve the impediment to reduction that this interdefinition presents.

Smith's analysis and reduction of 'right' uses an uneliminated circular interdefinition of 'right' with 'rationality'. Smith's anti-Humean rationalist meta-ethics rests the normative burden on this feature of his theory. More precisely the normativity of 'right' rests on the desiderative unity component of rationality. But Smith has no explicit theory of desiderative unity. The details of this claim form much of the next chapter. However supposing it is true then this is a major problem for any analogy between the colour case and the ethics case. In the absence of a theory like the a posteriori kinds found in the colour case we have no explanation of why the circular interdefinition of right with rational is the case. Given the a priori nature the relevant facts about rationality according to Smith we also have reason to suppose that an explicit squaring theory in the ethics case, which I will argue in chapter three amounts to an explicit theory of desiderative unity, will not give a reason to block the analysis and reduction of rationality forming a part of the analysis and reduction of right.[13] So even if we grant that Smith could provide a theory of desiderative unity we know that according to him its effects at least are knowable a priori. But since the desiderative unity features[14] effects are

[13] Nolan 2015 argues that much of philosophical activity, including the introspective consideration of conceptual commitments (a putatively essential component of conceptual analysis for both Smith and Jackson), is actually a matter of a posteriori investigation. This view will not really help Smith's position here because the disanalogy I am pointing to in this argument would remain in place even under Nolan's proposed scheme of things. Whatever is a posteriori about the discoveries of the relevance of surface reflectance properties to colour attributions is very different from the reflective process Smith claims yields moral information. There is also a question of what the Nolan scheme would make of definitional naturalism since the a aprioristic features we have been using thus far will not necessarily be in play.

[14] This is discussed in exhaustive detail in the rest of the thesis but for now we simply need remember that the desires of the fully rational maximise desiderative unity and this feature is what explains and constitutes the convergence on the same hypothetical desires relative to

what make for the convergent desiderative rationality Smith's rationalist meta-ethics requires for there to be fact about right and they are knowable a priori it seems like they should still figure in a definitional network analysis reduction of 'right' to natural properties.

To conclude: The ethics case and the colour case, despite the similarities in concept acquisitions and term interdefinition, are sufficiently significantly different to make it implausible to extend the reasons for accepting a permutation problem in the colour case to the ethics case. There are two core problems. One is that the ethics case, according to Smith, is a prioristic in its particulars and it unclear that this is not relevant to there being a definitional network analysis reduction of ethical properties to natural properties. The other is that there is a paucity of relevant background theory to play the role of squaring a narrow reduction of right to natural properties of acts with a broader naturalism. In the colour case much if not all of the burden of explaining and managing ineliminably circular interdefinitions of colours with colour experiences is carried by the squaring argument for colour and the suite of background theories that argument uses.

2.4 The induction argument for a permutation problem for definitional naturalism

Smith has an inductive argument for the permutation problem in TMP. He supposes that if definitional naturalism were possible then the failure of definitional naturalism to date would be remarkable. Smith thinks we can

---

circumstances in a population of fully rational agents that has to be the case for there to be any right actions. At least if Smith's analysis of right is correct.

suppose that this remarkable failure is explained by there being a permutation problem for definitional naturalism. He says

> "Surely the most plausible explanation of these failures is that such analyses are impossible. And the vulnerability of network analyses to a permutation problem is just what is required to establish this conclusion." (Smith, TMP, p 56)

This argument is repeated in TMP chapter 5 when he evaluates his anti-Humean theory of normative reasons. This is relevant to the reduction of right because moral reasons are just a subset of normative reasons.[15] Smith says there that his second reason for supposing that there is a permutation problem for our normative concepts (an analogy to the colour case permutation problem is the first reason) is that

> "… it seems to me that we have other inductive reasons for thinking that network analyses of our normative concepts are vulnerable to a permutation problem as well. For it is a remarkable fact about the history of philosophy that the analyses of normative concepts in non-normative terms have been such spectacular failures. It seems that any such analysis is vulnerable to a 'So what?' objection (Johnston, 1989; Gibbard 1990; chapter 1). What is needed to explain this remarkable fact is some principled reason why normative concepts elude non-normative analysis. The obvious conjecture is that network analyses of

[15] This is again a point that will be considered and reconsidered in some detail in this thesis but for now the following will suffice: Smith's anti-Humean rationalism includes the idea that normative and so moral reasons concern the desires of fully rational versions of people. What makes the fully rational desires of fully rational versions of people normative reasons is partly due to their being relevantly the same in all fully rational versions of all relevant people.

> our normative concepts are vulnerable to a permutation problem. For this is precisely the sort of principled reason that is needed." (TMP, p. 164)

Smith concludes from this that the only analysis available for normative reasons is the non-reductive summary style analysis he offers himself. Smith, presumably by way of his discussion of the colour case, thinks he has shown that analyses need not be reductive and explicit. Setting aside the pressure that we can put on this last claim when we pay more attention to the squaring arguments for the reduction of colours, Smith's inductive argument faces a particular problem. Jackson, in FMtE, with his distinction between folk morality and mature folk morality provides an alternative principled reason to explain the remarkable fact of the failure of definitional naturalism (the project to provide the reductive and explicit analyses of our normative concepts that philosophy has spectacularly failed to produce to date). That is just that there are as yet no uncontentious enough common fund of moral concepts upon which to perform a naturalist reduction in the first place.

Smith's inductive argument for a permutation problem for a naturalist reduction of normative concepts[16] requires that the apparent failure of definitional naturalism is just as it appears. That is that the intractable debates **really are** debates about the adequacy of proposed definitional naturalisms and so the intractability requires some kind of explanation that is

[16] This is one of the rear occasions where the difference between Jackson's reductive descriptivism and Smith's reductive naturalism might matter. According to Jackson's approach in FMtE reductive descriptivism will strip out all related normative concepts when specifying the non-ethical descriptive language to which ethical terms must be reduced. Smith's naturalism might permit distinguishing between the permutation problem for normative concepts and the permutation problem for ethical concepts but as I have argued the role of rationality and its normative components in determining the reference of ethical terms like 'right' make this a very unlikely position for Smith.

not likely to be forthcoming from proponents of definitional naturalisms. However, the inexplicability of intractable debate over adequate naturalistic reductions of ethics is only the case under the assumption that the conceptual terrain of ethics is fixed. In Jackson's terms this assumption amounts to the view that an explicit mature folk morality has been achieved – this is something Jackson thinks has not happened yet. Jackson provides an alternative view of the matter. Rather than intractable and mistaken debate caused by an unrecognised permutation problem a defender of Jackson's position allows an alternative account of this remarkable history. Jackson supposes that a large portion of ethical debate is actually conceptual debate. That is what is being argued about is not simply what the best explication of folk morality is but rather what folk morality is to be constituted by conceptually[17]. Jackson describes the goal of such disputation as mature folk morality. As we have seen mature folk morality is what you get when you achieve consensus on which consistent fragment of our inchoate implicit folk morality captures the most of that folk morality in the best way. Such a conceptual debate is possible if you accept, as both Smith and Jackson do, that concepts can be complexly interdefined and implicit allowing for apparently novel a priori content. These debates will necessarily intersect with analysis since contention over the content of mature folk morality can't make concepts up without explicit reference to the inchoate mass of implicit presuppositions that constitute folk morality prior to explication and regimentation. And it will be frequently contentious and apparently insolubly so since finding the best explication of the largest fragment of folk morality with which to make mature folk morality will always or at least often leave out material (for

[17] Technically in Jackson's scheme of things arguing over what is the best explication of folk morality is to argue over what our common moral concepts are to be. The aim here is to point out that this implies the alternative explanation Smith thinks philosophical history demands.

whatever reason) and so be prone to misplaced accusations of inappropriate revision. This will be exaggerated by the inclusion of actual revisions that proposed explications will likely sometimes include in the interests of coherence, explanatory depth, or other virtues one might favour in an explicit complete theory of the concepts of ethics.

There is no straightforward characterisation either of, which considerations are relevant to such a contestation, or how such considerations would interact so I will not attempt to generate a discussion of them here. Given Jackson's views about the absence of a mature folk morality and what such a morality is we can diagnose the putatively intractable debate inspired by definitional naturalism as conceptual debate – a debate over the some of the contents of mature folk morality. It is important to note that there is a feature of these conceptual 'debates' that is easily overlooked. They are not necessarily debates at all. Contests, negotiations, proselytising, and the like at least but debates over determinate matters of fact? Only sometimes. Since different views of mature folk moralities contents will propose that some concepts are kept and, presumably, some at least pre-theoretic implicit concepts not be kept then it is hard to see how an alternative proposal for preservation and omission could be disputing a matter of fact. And it is not the case that we are, by observing or even endorsing such exchanges that we thereby observe or endorse relativism. Certainly not relativism in ethics – everyone, in this account of the matter, is trying to generate an ethics <u>for everyone</u>. The requirements for grounding in pre-theoretic implicit folk morality the various attempts at a best, most comprehensive, consistent, explicit mature folk morality provides enough for it to at least be initially plausible that the conceptual 'debates' are not mere simple popularity contests either. For all we know there might well be a single or near enough mature folk morality.

Even if these negotiations were to effectively collapse into 'popularity contests' they would not necessarily be criticisable for that reason. Supposing there might be a standard for 'best' here is not the same as being able to or needing to enunciate it. More over topic overlap between mature explicit folk morality and implicit folk morality is secured to a great extent by requiring contender explications retrieve as much as is consistently possible of implicit folk morality. So, after all of this, if what occurs becomes a simple popularity contest there is no obvious objection to it. Concepts are just what we make of them so to speak. However the possibility that recalcitrant holdouts will prefer some variant on mature folk morality that is not widely shared might be a problem. The idea of a best, consistent, largest and so on explicit fragment of implicit folk morality does suggest that there are some sort of common standards assumed to be in play to avoid the possibility of non-culpable but arbitrary recalcitrance about mature folk morality. Perhaps Jackson should count the lack of such standards combined with a failure to converge for the most part on one mature folk morality as one of the several ways we can have error theory. Rather than conceptual incoherence for all individual bundles of moral concepts we find that the requirements for objectivity are not met because the population as a whole is non-culpably conceptually incoherent. For this to occur it has to be true that implicit folk morality can support these internally consistent but incompatible explications which cannot be coherently entertained simultaneously. That suggests that this scenario is a way of discovering that implicit folk morality is ineliminably, irremediably incoherent itself. Surely this is possible and there should be acceptable ways of discovering this.

It is notable that Jackson is aware of the possibility of multiple, equally good folk moralities, and explicitly relativizes his argument about how to apply the reductive framework expressed his moral functionalism to various possible non-convergent mature folk moralities.[18] Jackson does not make any claims about how to read this condition – whether it is relativistic error theory or not for example. It is arguable that this is not mere coyness. What divergent mature folk moralities should be considered as will in part be determined by what these divergent views make of each of them counting, so far as they do, as equally good 'best explications' of fragments of implicit folk morality. Convergent mature folk morality in this light is one success condition for shared moral concepts. But divergent mature folk moralities import will entirely depend on its explanation.

The purpose of this prolonged discussion of the role that implicit folk morality and convergent or divergent mature folk moralities play is not to defend Jackson's approach but rather to show that he can offer a *principled* explanation[19] of the remarkable history of failure that Smith grounds his inductive argument to a permutation problem in definitional naturalism on. This effectively blocks the persuasive power of Smith's inductive argument such as it was. Smith's inductive argument for the claim that definitional naturalism is vulnerable to a permutation problem fails.

---

[18] Should the assumption of convergent mature folk morality prove false, Jackson has this to say:

> "The identifications of the ethical properties should all be read as accounts, not of rightness *simpliciter*, but of rightness for this, that or the other moral community, where what defines a moral community is that it is a group of people who would converge on a single folk morality starting from current folk morality."(FMtE, p. 137)

[19] The principle is not obviously restricted to moral functionalism since the role an uncontentiously complete explication of an implicit folk concept is part and parcel of the defence of novel analysis that both Smith and Jackson share. And this invites a discussion of contentious explication, which in the case of folk morality Jackson does.

2.5 Next

In this chapter we have shown that the colour case that Smith uses to illustrate the permutation problem is, when restricted to the resemblance relations between colours, not useful. And when unrestricted, as it is in the dispositional analysis and reduction of colours Smith uses, the difference between Smith's summary style analysis and reduction and its competitors becomes obscured if it is supposed to be any more than the idea that crucial components of the colour concepts are a posteriori. We have shown that the argument from analogy to the colour case for definitional naturalism having a permutation problem fails because of significant relevant disanalogies between the colour case and the ethics case. And finally we have shown that Smith's inductive argument to a permutation problem for analytic reductive naturalism for normative concepts, including normative and moral reasons, fails because the central datum of that argument is given an alternative principled explanation by Jackson in FMtE. Jackson's explanation of the remarkable history the failure of definitional naturalism leaves open the possibility that a definitional naturalism might yet succeed.

Chapter 3 examines the role of desiderative unity in Smith's rationalist meta-ethics and argues that it is the normative hub of Smith's theory. This allows a more precise examination what the purpose of a permutation problem argument is in the context of Smith's rationalist theory. The chapter will argue that Smith's analysis will only count as an alternative to definitional naturalism if he can provide a reason for keeping the analysis and narrow reduction of right to natural properties of acts separate from the analysis and

reduction of rationality and in particular desiderative unity. The chapter will conclude that the features of the analyses of 'right' and 'rational' that Smith's theory explicitly characterise (successfully)[20] fail to provide that required reason. Later, in chapter 4, we turn from the narrow reduction of right to Smith's squaring argument for this narrow reduction to see if this can justify the separation of the conceptually linked analyses of 'right' and 'rational' and their respective reductions to natural properties.

[20] The qualification 'successful' is necessary here because I argue that at least one feature of Smith's explicit characterization of the analysis and reduction of 'rational' fails. This unsuccessful feature is just the use Smith makes of Rawlsian reflective equilibration to characterize desiderative unity.

Chapter 3 The Importance of Desiderative Unity

I have argued that the permutation problem no longer provides a reason to avoid a definitional network analysis and reduction of the interdefined terms 'right' and 'rational'. So Smith will need other reasons for maintaining what I call the semantic gap between the analysis and reduction of 'right' and the analysis and reduction of 'rationality'.[1] The present chapter suggests that Smith's notion of desiderative unity might play a substantive role in such a reason, and examines this suggestion in detail.

I will begin by arguing that desiderative unity is central to Smith's anti-Humean rationalism. Once we have established the role that desiderative unity plays in Smith's theory I will consider Smith's attempt to understand the notion in terms of Rawls's notion of reflective equilibrium,[2] and will argue that this attempt fails: a Rawlsian account is incomplete and fails to provide Smith with a reason to avoid definitional naturalism. I conclude that Smith's theory remains in need of a reason for preserving the semantic gap.

---

[1] The semantic gap between 'right' and 'rational' is a name for the position that Smith used the permutation problem to justify. This is the position that the analysis and reduction of 'rational' does not figure directly in the analysis and reduction of 'right'. If it did then Smith's anti-Humean analysis of normative reasons – given in terms of the analysis of rationality – would form part of a definitional network analysis and reduction of right. The fact that all the links between right and rational and properties of acts are conceptual **and** a priori makes a definitional network analytic reduction a plausible default position.

[2] See Brandt 1979 and Norman 1996 for discussion of the nature of Rawlsian reflective equilibria both applied to moral beliefs and applied to beliefs more broadly. Smith 2010 is a more recent example of his use of Rawlsian reflective equilibria. In that paper Smith is arguing for the possibility of objective moral reasons ; he claims that, under the assumption there are objective moral reasons, the Rawlsian method of reflective equilibrium is a good way to find out what our rational desires are. This use of Rawls in Smith 2010 adds nothing to what we find in TMP.

Section 3.1 The role of 'rationality' in 'right'

To understand the role of desiderative unity in rationality we have to understand the role rationality plays in Smith's account of right. We can give various formulations of the reduction of 'right' to what a fully rational version of ourselves would desire on our behalf given our actual circumstances – the reduction of 'right' to whatever it is that the fully rational versions of ourselves would advise we pursue given our actual circumstances. The final one Smith offers is that

> "Rightness in circumstance C is the feature we would want acts to have in C if we were fully rational, where these wants have the appropriate content" (TMP, pg. 185).

The rightness in a circumstance C is given by Smith's reduction of right to a natural property of acts. This is how Smith actually presents the reduction:

> "The analysis tells us that the rightness of acts in certain circumstances C – using our earlier terminology, let's call this the 'evaluated possible world' – is that feature that we would want acts to have in C if we were fully rational, where these wants have the appropriate content – and, again, using our earlier terminology, let's call this world, the world in which we are fully rational, the 'evaluating possible world'. Now though, for reasons already given, this does not itself constitute a naturalistic definition of rightness – though it is merely a non-reductive, summary style analysis (chapter 5) – it does provide us with the materials to construct a two-stage argument of the following kind.

> Conceptual claim: Rightness in circumstances C is the feature we would want acts to have in C if we were fully rational, where these wants have the appropriate content.
>
> Substantive claim: Fness is the feature we would want acts to have in C if we were fully rational, and Fness is a feature of the appropriate kind
>
> Conclusion: Rightness in C is Fness"
> (TMP, p185) [3] [4]

So fully rational desires at the very least pick out what is right. Thus far we have been using Smith's 'we' in the above quotes unselfconsciously. But it is meant to play a role. For it is only true that that there are rational and moral normative reasons if all fully rational idealisations of everyone (or near enough) converge on the same subset of hypothetical circumstance relative desires about what properties agents acts in those circumstances possess.

Recall that I call the reduction of 'right' to natural properties of actions the *narrow* reduction of ethical properties. It is narrow because the reductive argument made above has to be squared with a broader naturalism by giving

---

[3] In this quote Smith calls the circumstances where we are trying to make moral attributions 'the evaluated possible world'. What this amounts to bears on defending Smith's theory in general but it does not impact significantly on the goal of comparing Smith to Jackson. For what its worth it seems that specifying circumstances is an exercise in specifying morally relevant features of the world relative to the particulars of the agents in it and this is a task all moral theories are committed to having a view on. It is not a particular weakness of Smith's theory if it turns out that circumstance specification is tricky.

[4] Desires with the 'appropriate content' are just those desires we, as we are (not-fully rational), recognise as concerning moral matters. This precludes too much novelty in moral matters but whether or not this invites severe criticism of Smith analysis and reduction of right to the objects of desires of fully rational version of us is not at issue here.

a naturalistic account of full rationality. This parallels the reduction of colour to surface reflectance properties that Smith uses to illustrate non-definitional reductions. In the colour case the reduction of colours to physical properties in objects is only vindicated if it can be squared with a broader physicalism by giving a physicalist account of colour experiences. We will return the 'squaring' issue in the next chapter, but here simply mark this feature because of its relevance to later parts of our discussion.

At least on first blush the reduction of 'rational' is not, for Smith, part of the narrow reduction of 'right' to natural properties.[5] I say 'on first blush' just because though Smith's reduction of 'right' aims only to avail itself of the desires of fully rational versions of us (desires that are indexed to relevant variations in the circumstances of action) it seems to me that you can't get a grip on what these desires might be like, even in general terms, without an account of rational desire formation. Of course Smith might object to this claim by citing the paradigmatic learning of normative and ethical concepts. But though this claim might wash for the acquisition of current folk views of moral matters and perhaps serve to explain an intuitive grasp of the appeal of current folk views it does not seem like the kind of epistemic restriction one should make mandatory. The fact that moral concepts are acquired by way of paradigm based learning is contingent means we can't appeal to this fact to block the requirement for an account of rational desire formation. Moreover since Smith is optimistic about the progressive nature of the moral endeavour it seems like some part of the meta-ethical terrain has to be given the job of explaining how this progress is made, even if it does not do so by way of explicit definitions. And since progress implies that received wisdom can be

[5] It certainly looks as if this is the intention given the regular use Smith makes of parallels between the ethics case and the colour case and the examples he gives of how to find the property that makes an act right for an agent given a circumstance.

false it seems that paradigmatic learning, received wisdom about morality, and the acquisition of moral facts – especially new ones – must part company even for Smith.

So 'right' contains 'rationality' unanalysed, and 'rationality' contains (amongst other things) 'desiderative unity'. Yet it seems to me unclear how the reduction of 'right' relative to circumstances is to proceed without an account of 'rationality', and in turn any relevant unanalysed normative terms contained in the account of rationality. To be clearer, Smith does not reject the role of some kind of analysis of 'rationality' in the analysis and reduction of 'right'. But he must reject the availability of an analytic reductive analysis of 'rational' in the a priori understanding of 'right' or fail to have a view distinct from Frank Jackson's.[6] But unless Smith supposes an infinite regress of normative terms in definitional relationships, a view that is both implausible and for a naturalist arguably more trouble than it's worth, normative interdefinitions with informative summary style analyses that permit a naturalistic reduction have to terminate somewhere. And when they do the reductive status of the relevant normative term crops up, if for no other reason than because an account of the nature of that term and its referents is required for one to be able to square Smith's non-Humean rationalist meta-ethics with a broader naturalism. If a naturalist reduction cannot be effected on all the normative terms that crop up in the course of the summary style analysis and reduction of 'right' and all the normative terms nested therein then Smith's position will fail to be a naturalist one. As I will argue below the key nested term is 'desiderative unity'.

[6] That is, Frank Jackson's moral functionalism and particularly the definitional network analysis and reductive character of the moral functionalist framework Jackson provides in FMtE.

It appears that Smith suggests that the analysis and reduction of desiderative unity is not part of the analysis and reduction of 'rationality' because it can be reduced by way of a two premise reductive argument like the one used for 'right' and presumably 'rationality' (See TMP pg. 186). This is how Smith puts the matter:

> "Of course, the psychology of a fully ration creature is an idealized psychology, but such an idealization requires nothing non-natural for its realisation. Thus, if we wanted to, we could construct non-reductive analyses of the key normative concepts we use to characterize the normative features of such an idealized creature's psychology – the unity, the coherence, and the like, of its desires – and then use these analyses to construct two-stage arguments, much like that just given [for rationality], in order to identify these normative features of a fully rational creature's psychology with natural features of its psychology (for an analogy, see note 9 to chapter 2). Coherence and unity, though not naturalistically definable are therefore themselves just natural features of a psychology." (TMP, p 186)

So Smith is clear that his naturalist reduction of right to the natural properties is not vindicated until a complete account of all relevantly normative terms is given. He also seems to assume that reapplying his summary style analysis and reduction method to these terms will provide the required vindication. But the summary style analysis and reduction method assumes the semantic gap between the analysis and reduction of 'right' and 'rational', and thus effectively between 'right' and all the subsequent normative terms it depends on in Smith's theory. And this assumption requires a reason that was lost with the loss of the permutation problem argument against the definitional

alternative. Our point here is just that rationality serves to govern the reduction of right and appears to do so by introducing normative terms like desiderative unity. So even though rationality and the means by which it effect a reduction of 'right' is known a priori we already know that Smith accepts that an reductivist account of rationality and desiderative unity in turn are required.

3.1.1 Motivation, reasons, normative reasons, moral reasons – how rationality is used by Smith to link and distinguish these four things

We will Smith's taxonomy of motivations, reasons, normative reasons, and moral reasons. Smith makes the following distinctions in chapter 5 of TMP, particularly on page 166. This is the chapter where Smith provides his analysis of rationality.

Smith accepts a Humean theory of the distinctness of beliefs and desires, a Humean theory of motivation. But he rejects the Humean theory of normative reasons (roughly that there are no normative reasons relative to desires). These considerations are neither trivial nor uncontentious but the useful upshot for this thesis is just that motivations are instantiated, effective, desires. Normative reasons are by conceptual requirement, non-relative, and in Smiths analysis this means that they are the subset of fully rational desires that fully rational agents converge on, indexed by circumstances. Normative reasons are things everyone has reason to do in a given circumstance independently of our actual desires. Normative reasons concern the convergent overlapping desires of fully rational agents relative to a particular circumstance. According to Smith you have a normative reason to ϕ if it is

desirable to ϕ. It is desirable to ϕ for an agent S in a circumstance C if a fully rational version of S would desire that S ϕ in C. This is a normative reason because, by S's own lights, failure to desire ϕ in circumstance C is irrational.[7]

> "'Φ-ing in circumstances C is what we would desire that we do in c if we were fully rational" gives the content of the thought "ϕ-ing is desirable" (Smith, TMP, p 153)[8]

Smith argues (TMP, pp. 172-171) that normative reasons are by their nature non-relative. Analysis shows that normative reasons are concerned with the desirability of doing things. The desirability of doing a thing in a circumstance, by analysis, just is that a fully rational version of ourselves would desire we do that thing in the relevant circumstance. If there are non-relative normative reasons these claims amount to supposing that all fully rational versions of agents across all starting sets of desires will, relative to a circumstance, converge on the same desire regarding agents acts in those circumstances. Smith notes that this conception of normative reasons though conceptually required does not entail that any such reasons exist.

The rationally permissible could exceed the rationally required but that is of no matter here since we are considering the rationally required. A fully

---

[7] The satisfaction of the objectivity requirement on normative reasons by desiderative convergence of the appropriate sort in the appropriate idealized population is obscured in formulations concerning an individual agent and that agent's fully rational idealization. But it is a condition on their rationality that they belong to a population with the appropriate links to an ideal population with the appropriate psychological constitution. I will call this condition 'ideal desiderative convergence'.

[8] With the previous footnote in mind we should be able to see that we can use the terms 'fully rational' and 'desirable' in an agent relative way. If we do this we must be careful to avoid conflating this usage with the non-agent relative use – a use that for Smith only has significant content under the assumption of ideal desiderative convergence. The non-agent relative use is, confusingly perhaps, circumstance relative in a manner sensitive to the particulars of agents psychology when this is normatively relevant.

rational desire constitutes a normative reason when rationality mandates rather than permits it. And rationality mandates relative to a circumstance for all agents in that circumstance the same way according to Smith. So if there is a normative reason to act a certain way it is because all fully rational agents would desire as much given the circumstance. Rationally permitted but not required desires, so long as they manifest in a fully rational desiderative profile that idealise some particular agent, can be counted as reasons for that agent to do a thing.[9]

So in summary:

> Motivations are actual desires.
>
> Reasons are concerned with what a fully rational version of you would want you to do given your circumstances. Reasons that are not normative reasons concern those things particular to you in your circumstances that a fully rational version of you would want you to do. Such agent-specific reasons are not requirements on everyone. You may not be motivated to do what you have reason to do.
>
> Normative reasons are those things all rational agents would want you to do given your circumstances - normative reasons are convergent, non-relative, and their existence is not a matter of mere ruling by fiat. Smith notes that normative reasons may not exist actually. In fact it is a requirement that normative reasons might not exist since merely

[9] Smith canvasses ways to differentiate fully rational idealizations of agents and the ways in which you can use circumstances actual and hypothetical *relative to the fully rational agent* to distinguish where convergent desires are required and divergent desires are permitted at full rationality. But this detail is not relevant here so I will not further evaluate it.

> having a concept is not enough to suppose that anything conforms to that concept.

Normative and agent-specific reasons do not have to co-vary with an actual associated motivation. However so far as you are not motivated to act and want in concert with your reasons you are by your own lights irrational. Again Smith notes this defeasibility characteristic as advantage since the aim of his anti-Humean rationalism is to generate morality for actual agents, even when they are irrational. That the connection to motivation is analytic in some sense is also an advantage since this satisfies the practicality requirement of morality (that moral beliefs have at least potential motivational impact), a requirement Smith wishes to retain and reconcile with the objectivity requirement on moral judgements (effectively, considered beliefs about morality).[10] Moral reasons are that subset of normative reasons that concern moral matters (supposing that the latter does not exhaust the former).

Because agent-specific reasons and normative reasons are about desirability and don't entail possession of a motivation, Smith's account of reasons allows some degree of commitment to the reason by an agent's having a belief about the desirability of some thing. If that belief were true then a fully rational version of yourself would desire that thing and your failure to acquire the relevant desire would entail your own irrationality. Coming to believe that something is desirable, even when the belief is false, entails believing that you

[10] Agent-specific reasons and normative reasons are analytically connected to motivation in a fully rational agent simply because in a fully rational agent reasons are constituted partly by desires the fully rational agent has and motivations are desires. The connection between reasons and motivation in sub-fully rational agents is mediated by the fully rational idealization of these agents. Part of the account of right should include an explanation of why sub-fully rational agents should care that they are not motivated, as rationality requires they be. This issue is discussed further later in this chapter.

would acquire the relevant desire if you could and count yourself irrational if you failed. What you have reason to do by your own lights corresponds to facts about the desiderative profile of fully rational versions of yourself given your circumstance. What you suppose you have reason to do concerns what you believe about your reasons and these suppositions can be false, even if there are truths about your reasons and normative reasons. Talking about some degree of commitment to reasons in virtue of beliefs about desirability is deliberately coy because Smith is not grounding the relevance of rationality on non-derivative current motivations to be rational. This means that his theory will have offer some explanation of the relevance of the desires of fully rational idealisations to sub-fully rational agents. I argue, in section 3.3 that the feature of Smith's theory that could provide this explanation is desiderative unity.

So according to Smith: Desiderative rationality dictates the existence of agent-specific reasons, normative reasons, and moral reasons. According to Smith's anti-Humean meta-ethical theory all of these types of reasons have to be understood in terms of the desires of fully rational agents and for normative and moral reason they have to be understood non-relativistically. And again by analysis[11] the non-relativistic understanding of normative reasons and so of moral reasons requires the supposition of convergence in the relevant desires of the relevant population of fully rational agents. (TMP: pp.164-177)

So even though rationality is not analysed in Smith's narrow reduction its function and nature dictate the conceptual limits of 'right' and Smith

[11] Analysis is meant here in the ecumenical sense that permits Smith's avowed tolerance for circular definitions informatively guiding successful reductive theories. I admit however that I don't think the distinction is needed since Smith thinks the concepts of morality, rationality, reasons and desirability - when explicitly understood - dictate these requirements and relationships.

supposes we know a lot about these conceptual limits. We will now turn to what Smith takes full rationality to be in more detail and elaborate the central role of desiderative unity in determining the both the concept of full rationality and, latter, whether or not there are any normative or moral reasons.

3.2 Smith's account of full rationality and the central role desiderative unity plays in it.

Here we will consider both Smith's use of Bernard Williams' ideas of full rationality and his extension of the notion of correct deliberation to include desiderative coherence and unity. The purpose of these latter two notions is to provide a norm for the creation and extinction of non-derivative desires and, as desiderative unity is maximized, to provide a systematic justification of desires. Though Smith characterises desiderative unity in terms of Rawlsian reflective equilibration we will leave evaluation of this move until later in this chapter.

Smith offers Bernard Williams ideas about how, in the course of achieving full rationality, desires can be included among the objects of correct deliberations where rational beliefs and desires are the products of correct deliberation. In order to be fully rational, Williams thinks an agent must satisfy the following three conditions

> "(i) the agent must have no false beliefs
> (ii) the agent must have all relevant true beliefs
> (iii) the agent must deliberate correctly" (Smith, TMP, p 156)

It is the operation of correct deliberation that interests Smith because it is by means of correct deliberation that Williams, and he, suppose that norms for the modification of sets of desires can be worked out. The modification of sets of desires to generate fully rational desires is of some interest to Smith since it is the desires of the relevant fully rational population (and it's relevant overlaps) that Smith supposes constitutes normative reasons and a subset of normative reason, moral reasons. Normative reasons are just the convergent hypothetical desires of the fully rational relative to circumstances.

You might mistake the rationality gains of conforming your derivative desires to your means end beliefs and non-derivative desires for ends, as sufficient for normative reasons. I will call the rationality gain made conforming your derivative desires to those that a means end belief and non-derivative desire pair imply are good to desire 'instrumental rationality'[12]. But instrumental rationality is not sufficient for normative reasons according to Smith. TMP, chapter 5, question six, p 164-174, Smiths argument that normative reasons have to be non-relative and his discussion of how this is satisfied by the right kind of convergence in the circumstance relative desires of all fully rational agents shows this at some length.

Smith's analysis says that we have a normative reason to do something in a given circumstance if we would desire that we do that thing in that circumstance were we fully rational (Smith, TMP, p 181). Correct deliberation is how we come to know about the relevant determinants of our normative reasons, it is how we come to know about the desires of fully

[12] Many of the concepts traded in when describing Smith's rationalist theory even those bits that are not anti-Humean, as instrumental rationality is, are controversial. Luckily I am not defending rationalism in this thesis.

rational agents (relevant to us so far as they are fully rational versions of us). But Williams folds the capacity to work our or come to know what it is rational to desire with the capacity to modify your desires accordingly into the notion of 'correct deliberation'. Williams considers the mechanisms of correct deliberation and its effects on desires by giving instances of it. For example

> "Our desires and beliefs only generate new desires if we deliberate and do so correctly. Thus, for example, they generate new desires only if we reason in accordance with the means-ends principle, for only so does a desire for an end turn into a desires for the means" (Smith, TMP, p 157)

Smith notes that Williams proposes other mechanisms of correct deliberation coming to desire a thing might be a convenient, economical, pleasant, or way of satisfying an element in your set of desires. And these facts and their relevance (presumably) are controlled by other elements of your set of desires. Broader considerations of time ordering of the satisfaction of desires, relative weighting of desires where some of them are not mutually satisfiable, finding particular things to want that satisfy a general desire and so on. Williams also adds the exercise of the imagination to correct deliberation. Effectively the idea seems to be that imagination is a way of considering the impact of desire satisfaction without actually satisfying it first. Smith goes along with this so far as it goes

> "In general terms, Williams' conditions (I) through (iii) seem to me to constitute a fairly accurate spelling out of our idea of practical

rationality. I think that they need supplementation, however." (Smith, TMP, p 158)

Smith thinks akratic cases where one is addicted, subject to a compulsion, emotional disturbances and the like are not explicitly covered in the three conditions. He supposes their absence might be presupposed by correct deliberation. This is too swift, however. For according to this notion of correct deliberation we fail to correctly deliberate if we fail to adopt the relevant desires that someone who correctly deliberates would adopt. It seems to me that we might simply notice that the deliberative process of sub-fully rational agents like us have two dimensions of function, identifying the requirements of desiderative rationality and achieving the requirements of desiderative rationality. Certainly Smith distinguishes these two dimensions and needs to if he is to avoid necessarily connecting desires to beliefs – a goal he in fact has when try to resolve the inconsistent triad of claims that make up the moral problem.

More importantly, Smith also thinks that the Williams' account leaves out the main measure by which sets of desires are more or less rational. This is the extent to which they are systematically justifiable. Desires are systematically justifiable insofar as they, taken as sets, display more coherence and unity. But coherence, so far as it is distinct from unity appears to be simply consistency and so is just a matter of instrumental rationality. What Smith thinks is required for the provision of normative reasons is some feature of rationality that bears on non-derivative desires directly since this is what is needed to generate a circumstance relative desiderative convergence for ideal agents. This is because if non-derivative desires were not subject to rational criticism then only instrumental rationality would have a bearing on which

sets of desires are preferable and instrumental rationality leaves non-derivative desires unchanged. The feature most obviously in place to condition non-derivative desires is desiderative unity. Playing this role of accounting for the convergence of fully rational desires relative to circumstances makes desiderative unity the part of Smith's theory of rationality that matters to the determination of what 'right' refers to.

Section 3.3 How desiderative unity is the truth maker for moral claims.

This section has two tasks. The first is to clarify what desiderative unity is and how it operates to make claims about morality true or false, given Smith's conceptual theory. We have seen that according to Smith moral norms require rational norms and rational norms require that fully rational agents display a relevant convergence on desires relative to circumstances. We have discussed how Smith thinks that desiderative rationality is about in part about a justificatory structure between desires that is constituted by the maximisation of desiderative unity. In this section I will clarify features of desiderative unity since it's role in Smith's theory is arguably grounded in two different locations: the contents of desires (which are what display unity relations) and the presence or absence of desires as such (since facts about desiderative unity target rationality increasing desiderative change). Smith's theory requires that both features be included by the relevant moral or normative concepts and we can explain why this is by considering Smith's goal in TMP to resolve the inconsistent triad of claims[13]. The second task is to

[13] The goals I have in mind are to preserve cognitivism and the practicality of moral judgments. The third component – preserving the Humean ideas about the differences between beliefs and desires is achieved in part because Smith's rationalism involves an idealization, one we don't in virtue of only knowing about it instantiate.

show that actual desiderative unity is vital to the role that desiderative unity plays as a truth maker in Smith's ethical theory. Actual desiderative unity is just whatever piece of psychological machinery plays the desiderative unity role in actual agents. I will point out the implications Smith's conceptual theory has for the psychology of actual agents and in term how this psychology bears on the truth of Smith's anti-Humean rationalism. Actual desiderative unity, a piece of psychology in actual agents playing a desiderative unity role is essential for allowing the possibility of Smith's anti-Humean meta-ethics to be a comprehensive, accurate, or correct explication of our moral concepts while turning out to be actually false of us.

3.3.1 Clarifying desiderative unity – how it both encompasses properties of the contents of desires and constitutes part of a psychological mechanism bearing on desires as such

Desires are mental states, features of psychology. There is no need to give a full theory of desires here. We will, like Smith, accept the Humean psychology of belief and desires in terms of the relatively simple idea that beliefs map out the way we take the world to be and desires map out the way we want the world to be. Desires are essentially motivating. They have contents, at least in the sense that desiring a thing be a way involves the desire has some representational component that both the thing and the way we want that thing to be if we have the desire. This is a far from complete account of desires, or even a complete enough sketch of desires but it serves to make the distinctions we need in order to understand desiderative unity more clearly. Desires are not their contents. Sets of desires and the revision of sets of desires by the creation or destruction of desires in those sets is

something that occurs in a psychology and involves desires (not just their contents). Talking about revising your desire set to make it contain more of the desires of the fully rational version of yourself (on pain of irrationality if you are not able to do so) is talk about mechanisms of change in a psychology and focused on desires as such.

Talk about the justificatory and explanatory structure of sets of desires is not just about the desires as psychological states. Rather, as Smith indicates in his discussion of desiderative unity (TMP,155-161), the explanatory structure of sets of desires is a feature of the contents of desires. When Smith talks about how desiderative unity is increased in a set of desires when you add a more general desire that explains some more specific desires the general and specific nature of the desires are just the general an specific nature of the objects of the desires, represented in their contents (TMP 159).[14] Whatever else an explanatory structure among desires is it is a feature of relationships between their contents.

By pointing out that desiderative unity is concerned with the justificatory structure of desires (understood in part as an explanatory structure), Smith suggests that it is a property of the relations between the contents of a set of desires. He also suggests that it is maximised in the desire sets of fully rational agents and is increased in actual agents desire sets to the extent that they can come by true beliefs about the desires of the fully rational and can adjust their desires accordingly.[15]

---

[14] This is part of Smith's use of an analogy to Rawlsian reflective equilibration to account for what desiderative unity is – an account we will examine in more detail later in this chapter

[15] Though Smith's, and consequently my, discussion of desiderative unity is somewhat abstracted Smith does seem to suggest that desiderative unity is something that sub-rational agents can increase piecemeal or bit by bit by iterative rounds of deliberation.

Strictly speaking properties don't admit of increase or decrease, they come whole cloth, as it were. However, there are a number of different ways we can retrieve the idea in play here without doing violence to the notion of a property. One example is treating desiderative unity as properly a 'degree' where a set instantiates the property of 'desiderative unity to some degree x'. Then we can talk about a cluster of properties that are all desiderative unity to some degree and use these degrees to order them in some relevantly non-arbitrary way (perhaps via a partial ordering). You could suppose that desiderative unity is a scalar relating whole sets of desires in an ordering toward maximum desiderative unity. Finally you could suppose desiderative unity is a function from sets of desires to numbers or at least an ordering again where the ordering is non-arbitrary and admits of comparisons between desire sets that correspond to 'increasing desiderative unity'. The important point is that both the notions of a maximum desiderative unity and increases in desiderative unity have to be preserved in the appropriate more precise specification. Otherwise Smith and I should be indifferent so long as some more precise notion is available. I will where needed favour the cluster of properties approach and when talking about desiderative unity will ultimately be talking about clusters of desiderative unity properties that admit of degree.

So desiderative unity properties are properties of the relations between the contents of sets of desires and in this manner are properties of desire sets. It is a feature of desire sets that comes in degrees –fully rational desire sets maximise it. You can increase it to the extent that you can acquire one or another particular desire that a fully rational version of yourself would have relative to your actual circumstances. This mechanism of recognising increases in desiderative unity in a set of desires and responding by acquiring

those desires that yield the increase is at least a feature of the psychology of fully rational agents. And actual agents are more or less rational to the extent that they can likewise change the contents of their desire sets. This psychological mechanism of desire change can exist or not, be more or less reliable and vitally it must respond to increases in desiderative unity at least defeasibly. By this I mean we should allow that sub-fully rational agents could make mistakes about desiderative unity changes between desire sets in addition to being able to fail to respond to their beliefs about desiderative unity change by adopting the relevant desires themselves. We don't have to suppose the mechanism in anywise a perceptual one. Plausibly, given Smith theory, the psychology of desiderative change in rational agents is one that is responsive to beliefs about desiderative unity. Smith's interest in preserving the links between moral judgments and motivations seems like it would give a strong reason for taking this position.

So we can see that though desiderative unity can be regarded as a cluster of properties[16] of the contents of sets of desires it determines necessarily a feature of the psychology of rational agents – it partly determines their tendency to revise their desire sets in the direction of maximum desiderative unity. With this clarification in hand we can now turn to the impact of desiderative unity the truth of moral beliefs.

3.3.2 Desiderative unity and the truth of moral beliefs

[16] Such a cluster of properties way of talking about desiderative unity is designed to capture the idea that desire sets can be the same or different (to such-and-such a degree) with regard to desiderative unity, a cluster that takes in not just our actual selves but also more rational versions of ourselves, and all with the same gradable notion of unity in play. While no doubt more needs to be said to ground, and show the usefulness of, this way of talking, the above should suffice for the purpose of the thesis.

The first thing to note is that Smith's account tells us that a moral belief, about the rightness to act in a way given a circumstance, is true if a fully rational version of ourselves would desire that we act that way in that circumstance. But we also know that this is not enough to secure normative and then moral reasons and so to secure normative and moral truths. Normative and moral reasons are necessarily, conceptually, objective. And for Smith this amounts to supposing that all relevant fully rational agents would have the same desires for all agents to act a particular way given a particular circumstance. If they do not then acting that way in that circumstance is not morally or normatively required. But the desiderative profile of a fully rational agent is just one that maximises desiderative unity. In Smiths analysis the only feature that can yield comprehensive change in a particular direction for any fully rational agents set of desires is the increase in desiderative unity. So it is desiderative unity increase (to maximum) that will explain the convergence of the complex subset of circumstance relative desires between fully rational agents.[17] So desiderative unity being such that its maximisation brings about an overlap or convergence in the complex subset of circumstance relative desires between fully rational agents is what it takes for normative and moral truths to exist.[18]

[17] There might be a recherché possibility of a 'coincidental' overlap of the complex subset of circumstance relative desires between fully rational agents. However I don't think such a possibility can be raised in a non-question begging and clear way for anti-Humean rationalist theories like Smith's.

[18] Nothing is meant to turn on the use of the word 'complex' here. It is just worth remembering that the component of the desiderative profiles of individual rational agents that has to be the same between all relevant rational agents for there to be any normative and moral reasons is a set of desires that cover all the possible circumstances that agent (and the sub-rational agent it is an idealization of) might find themselves in.

Smith's conceptual theory, if true[19], implies that actual agents' psychologies instantiate a property, a desiderative unity property, the maximisation of which constitutes full desiderative rationality.[20] Smith himself notes that it is a substantial matter whether or not any agent is rational in the way Smith's conceptual theory requires. What this appears to amount is that it is contingent that there is a property or property cluster that deserves to be called desiderative unity. We should also allow that me might accept some near enough property of agents psychology to count as desiderative unity or 'desiderative unity^{ish}' and yet for that property of fail to lead to convergence of the right sort in the desires of the idealisations that 'desiderative unity^{ish}' permit. This is just a way for us to appear to be desideratively rational but turn out after all not to be. And as we have noted above it is an essential feature of a conceptual analysis of normative and moral reasons that possibility of this kind of failure to be true is preserved.

It appears that we could quibble with this last option since we might instead simply insist that any property or complex of properties does not deserve the name of 'desiderative unity' unless it generates the relevantly convergent desires in idealised agents generated by the maximisation of said property or

[19] To clarify a conceptual theory can be false of the concepts it is a theory of. However this can be confusing since a true theory of our moral concepts expresses the theory of the way the world is according to those concepts and that theory – the folk moral theory as it is oft times named – can be false. An analysis, according to Smith and Jackson, is a theory of our concepts. However when I am talking about the adequacy of an analysis as a theory of our concepts I will talk about the correctness of the analysis and when I wish to talk of the truth of the folk theory a correct analysis express I will talk about the truth of the analysis.

[20] We should keep in mind here the qualifications on properties and increase made above. We can read 'desiderative unity is a property …that increases' as short hand for more precise notions that preserve the ability to order sets of desires in terms of an increase in desiderative unity.

complex of properties.[21] But I think that insisting on this is ill advised because, as we will see below, there are independent reasons for Smith's account of desiderative unity to focus on and use a distinction between actual instantiated desiderative unity and the desiderative unity that kind of desiderative unity that idealises to a relevantly desideratively convergent fully rational population (see section 3.4). Talking about an actual desiderative unity that fails to generate the convergence in desires at full rationality that normative reasons and moral reasons require allows us to talk in some detail about how it is that Smith's conceptual theory could be true as a conceptual theory but false actually. And since preserving this possibility is a virtue of a rationalist theory it seems that the more you allow your rationalist theory to say about that possibility the better. Finally, for Smith a fully rational version of an agent can have desires that are not part of the convergent, circumstance relative, subset of their desires relevant to normative and moral reasons. For an agent, of whom this fully rational agent is an idealisation, non-convergent but ideal desires count as reasons for that agent in particular. There are at least two ways that a fully rational idealisation of a particular agent can have desires that are not shared by all other fully rational agents. The first is by way of rationally permissible but not required desires and the second is by way of variations in instrumental rationality. Allowing that the failure of morality can leave a relevantly similar, if not the same, notion of rationality in play is tricky. Smith's use of desiderative unity shows us how to understand this. Rationalist conditioning on desires that provides reasons for particular agents but not for all agents relative to particular circumstances might be the only kind of desiderative

[21] It is simpler to leave the complexities of precision out here. But again I think we can see that if we should need to then the comparisons between sets of desires, mediated by relevantly similar properties instantiated in the relations of the contents of their actual desires, and needed to cash out talk of increases in desiderative unity to a maximum should be easy enough to create.

rationality available. This option can be supplied by a 'desiderative unity' property cluster that does not engender the convergence needed for normative and moral reasons but still deserves the name since it's maximisation in idealisations for individual agents gives those agents reasons to act.

So we can dub the property I have in mind here 'actual desiderative unity'. It might exist. If it does it might be the kind of desiderative unity that generates the convergence on the same subset of circumstance relative desires between all rational agents and so make Smith's anti-Humean rationalism true. Or actual desiderative unity might only serve to rationally condition the desires of individual agents without generating convergence. Smith's conceptual theory is correct but actually false. Finally 'actual desiderative unity' might not exist at all. This is a condition where Smith's anti-Humean rationalism is false. It is not clear to me that we can decide that this circumstance shows Smith's anti-Humean rationalism is an incorrect explication of folk morality as well as false. Folk morality might just be a hopeless failure rather Smith's analysis be incorrect. One reason we should count the absence of any 'actual desiderative unity' as a failure of Smith's analysis is that Smith supposes that facts about fully rational desires are a priori available even if we don't actually desire them. The actual existence of a deserver of the name 'desiderative unity', that permits that Smith's anti-Humean rationalism turn out to be false while still being a correct explication of our folk moral concepts would at least show how we came to have and reasonably persuaded to Smith's conceptual theory. If nothing deserves the name 'actual desiderative unity' or something like this then the plausibility of Smith's conceptual theory in any regard would be mysterious. In effect I am suggesting that Smith's view should be that if we entertain anti-Humean

rationalist meta-ethics at all we should also bet that it can't turn out to be false because folk morality is a thorough going failure that doesn't even have residual evidentiary support.

Desiderative unity in whatever guise is central to Smith's anti-Humean rationalism since facts about it determine the truth or falsity of moral claims. We can even make a case for Smith's focusing on actual desiderative unity, where this kind of property both deserves the name and could be such as to make Smith's anti-Humean rationalist ethics true or false. We will now turn to another argument for this understanding of desiderative unity. Again we will be teasing out consequences of Smith's own theory.

Section 3.4 Reusing Smith's moral fetishism argument on desiderative unity.

In TMP pp. 71 to 75 Smith gives what he calls the fetishism objection to externalism. The objection, amongst other things, claims that if a thing is good in a moral sense then it should be valued for its own sake rather than valued only derivatively because that thing falls within the scope of a moral theory. This argument has implications other than those relevant to internalism versus externalism.

3.4.1 Smith's fetishism argument against externalism and its implications for property identification in the ethics case

Smith offers a fetishism argument against externalist non-cognitivist positions. I am interested in this argument but not for the purposes Smith

puts it to. I am indifferent to the issue of internalism versus externalism.[22] Rather I am interested in teasing out the implications of Smith's fetishism argument for the determination of which things count as the moral properties. In particular I think that you can not identify the moral properties with those properties that make actions right, in a narrow sense, even though Smiths reduction of right making properties of acts to those properties that fully rational versions of ourselves desire those acts have actually appears to do just this.[23] When you consider the properties that make acts right and suppose that such acts are all there are to (moral) goods you invite the possibility that counterfactually these properties are instantiated in the acts of agents who none the less cannot become more rational and potentially, though perhaps more controversially, may not be idealisable to our full rationality. Smith does not take this position up because he does not think that the reductive identification of right with the natural properties of acts is all there is to moral goods. However since relative to circumstances the properties that make acts right in those circumstance are identified with a natural property of those act it looks like Smith can't block the possibility that these properties occur without an agent either desiring them or being able to be idealised to desire them. That is Smith appears to have to entertain the possibility that moral properties can be instantiated in acts of agents who cannot nor, under any idealisation relative to themselves, could not come to desire that their acts instantiate these properties. Such a possible world

[22] Internalism and externalism dispute the source of moral motivation. Assuming cognitivism (the view that moral beliefs are truth apt and so really beliefs), internalists assert that there is a connection between forming moral beliefs and coming to be (at least defeasibly) motivated to act accordingly that is internal to the agent (as some would say, that the beliefs themselves are necessarily motivating). Externalists deny this claim and suppose that some factor external to rationality and the formation of moral beliefs is required for motivation to arise in the wake of moral belief formation.

[23] As we have noted above the 'narrow' reduction of right is effected with a two premise argument and gives circumstance relative identifications of right making properties of acts with natural properties of acts.

appears to be one in which you have (moral) goods that no one desires in the right way.

Smith's fetishism argument tries to show that it is essential that agents have the right attitude towards goods, moral or otherwise. They must love them for their own sake (de re) not because the fall under the terms of a moral or normative theory (de dicto). To suppose that the correct attitude towards goods is to value them because you believe a moral theory tells you to value them is to make of morality a fetish according to Smith. As far as it goes this argument seems very plausible. The requirement that you love the objective 'goods' for their own sake does not on the face of it bear on reference fixing on those properties of actual acts that fully rational versions of ourselves would prefer our acts have given the circumstances in which those acts occur. But Smith's account of the successful reduction right to a natural property in TMP is supposed to be result of the combined claims that it is definitional that what is right for agent S in circumstance C to do is what a fully rational version of S would desire they do in C and the non-definitional further claim that for S in C the property that a fully rational version of S would desire S acts have is Fness. From this we get 'Right for S in C is Fness' - an identity of right in C for S with F. And this identity claim appears to allow the counterfactual I have described above if you reference fix on Fness.[24]

Smith's internalism might be though to have no implications counterfactually. But if not then how could it both be that understanding the moral good

[24] Reference fixing in this context is meant to be just like the reference fixing found in the identity claims involving natural kind terms and their actual realisers. The issue raised by the narrow reduction of right to the natural properties of acts is just how to understand the resulting identification. If the natural properties can occur without the potential for fully rational appreciation of them then we have the possibility that shows a consequence of reference fixing on the actual realiser properties for 'right'.

requires you have the right attitude towards them and it be possible to reference fix on a property of acts external to the moral psychology the agents so acting? The idea here is that the identification of right with its narrow reductive natural property realiser is either only correct in the context of ideal rational agents and so not strictly speaking an identity relation at all or it is an identification and so allows for the possibility of right acts occurring without any agents so acting holding them in the required rational regard[25].

If Smith's fetishism argument shows that correct attitudes towards the moral goods is required of actual agents, then could Smith simply not reference fix on the properties that his analysis and reduction identify with right and remain indifferent to the counterfactual possibilities? It seems to me that he can't. We have seen the unpalatable result of reference fixing on the actual right making properties for acts. If you don't reference fix on the reductive element of an identity claim when evaluating the distribution of moral properties counterfactually then you seem vulnerable to having to allow that something like an indefinite description has all the information that is essential to being the moral property 'right'. But that is to say that the actual identity of moral properties with actual natural properties of acts is not relevant counterfactually. And of course it allows that the indefinite description is playing a meaning-giving role – that is that whatever gives you reason to favour properties actually as right makers is all there is to making a thing good or bad at all. The realiser properties themselves are inessential – that is they are not necessary.

[25] The details of types of identities in reductive contexts and the implications identity types have for the concepts governing them are complex and not needed here. Given a necessary identification the parameter Smith needs to fix is whether the necessity is a posteriori or not. You could suppose that the identity is not necessary but in the absence of a reason to prevent it this would lend itself to taking the description leading to a contingent identification as definitional.

There are at least two reasons for Smith to avoid such a claim.

The first is that adopting the view that indefinite description idea looks likely to lead to an immediate collapse of Smith's view into Jackson's moral functionalism - a definitional network analysis reduction. The collapse to Jackson would be got by supposing that you understand all there is to ethical properties by understanding the description that leads to actual right making properties being picked out. The identity of ethical properties becomes irrelevantly coincidental, and naturalism is secured if you can show that the mechanism for picking out right making properties is consistent with naturalism and picks out naturalistic properties actually.

This strategy seems to simply turn into a reason for adopting the reference fixing position on the actual identity claims that are the upshot of the narrow reduction of right to the natural properties of acts relative to circumstances that fully rational versions of agents in those circumstance desire those acts have. And that leaves us with the counterfactual that stands at odds with Smith's internalist principle that believing that some property is morally good requires the appropriate attitude towards that good.

The second reason for not adopting this strategy is that even if it does not have the features I have just described or one finds that these features are undisturbing we still should realise that the indexing of fully rational desires to circumstances is not restricted to all and only actual circumstances. On the one hand you most certainly can't restrict the circumstances relative to which fully rational versions of us have relevant desires to those circumstances that have happened to date. Smith, reasonably it seems, is optimistic about the

possibility of moral progress and a significant feature of moral change, progressive or not, is the arising of circumstances that we would find surprising and novel. The idea that you could restrict circumstances to all those that have existence actually (say in all past and future instances of human existence) seems empty just because we have no idea what the future circumstances are actually going to be. Finally is seems at least initially plausible that we and so fully rational versions of ourselves could have desires about mere possibilities just because the possibilities have moral implications. Smith does not offer an account of the circumstances that form a part of his analysis and reduction of right but these circumstances seem like the kind of thing that should remain relatively unrestricted. The range of moral interest across hypotheticals mitigates against identifying 'right' with any subset of natural property realisers of 'right' that come from the narrow reduction of right to properties of acts. This might show that the 'reference fixing on *actual* realiser properties' idea is imprecise given Smith's theory. However we can still ask the same question about the set of all the properties thrown up by the narrow reduction of right to properties of acts. Can any or all of these properties be instantiated in the acts of agents who cannot be idealised to full rationality? The point here is that Smith's theory has to take up a position on this possibility. I argue that the fetishism argument that Smith makes against externalism shows that he should find some way of denying this possibility is possible given his conceptual theory about the nature of 'right'.

It seems to me that the matter is being considered in an entirely wrongheaded manner when we decide to focus on the properties of acts that make them right. This focus is too narrow and drops out too much information and the considerations above seem explicable as a side effect of this. And Smith's

fetishism argument - that the properties that are morally good should be valued for their own sake and not because they fall within the scope of a moral theory on pain of making a fetish of that moral theory (a bad thing to do) - shows us just where the problem lies. Goods, moral or other wise, are properties or objects towards which an appropriate attitude must be had on pain of some sort of failure to understand them as goods at all. Smith wants to argue that the appropriate attitude towards a good (in the relevant sense of good) is just that you desire it or would if you were fully rational. That seems like the full measure for loving goods for their own sake rather than fetishistically via affection reserved for a moral theory. This desire is defeasible in Smith in the hopefully by now expected way. Knowing that a good is good, and knowing that this recognition should be accompanied by an appropriate desire does not entail that you have the relevant desire. It does however, for Smith, entail that if you don't have that desire then you suppose you are irrational by your own lights.

The responses of the agent who acts, or an idealisation of these responses combined with a disposition towards instantiating this ideal, play an important role in Smith's meta-ethical theory. Given that the desiderative psychology of fully rational agents is the primary determinant of what is actually normatively and morally a reason, given that fixing the desiderative profile of the relevant population of fully rational agents fixes without remainder what if anything is morally good (actually and counterfactually) then we have good reason to accept the view that the desiderative profile is wholly or significantly constitutive of normative and moral goods. That is the identification of 'right' with the right making natural properties of acts is necessarily incomplete and so we should not consider this putative reduction as anything other than a component of the reductive identification of 'right'

which must necessarily include an account of the desiderative structure and idealisablity of the agents concerned.

Since what fully rational agents want determines what is good for us in all circumstances then it seems to me that we should, if we wish to fix an element of the actual story to find the distribution of right making properties counterfactually, fix on that element that links us conceptually to the desiderative profile of fully rational populations. And Smith has told us what it is that we share with fully rational agents that links us to them in the right way. It is not the inclination to be as rational as possible. I think Smith implicitly gives us an idea of what that inclination consists in and it is to this that we should look when fixing our idea of moral properties. It is the maximisation of desiderative coherence and unity that fully rational agents achieve and to the extent that we emulate rational changes in our desires, particularly those that are non-derivative or have as their contents as ends as it were, we adopt changes that increase desiderative coherence and unity. And between coherence and unity I argue the desiderative unity does the brunt of the normative work in that increasing it, according to Smith, increases the justification of desires in part by marshalling them into structural relationships that embody or reflect explanatory relationships.

What ever the right story is about desiderative unity, what matters in this thesis is that it is a vitally important part of Smith's meta-ethics. Facts about this property, whatever it is, determine the desiderative profile of fully rational agents. Facts about it will determine whether or not the desiderative convergences necessary for normative and moral reasons occur. Facts about its increase determine for us actually at least the stepwise increase in our desiderative rationality. In so far as desiderative unity is a property of our

actual psychology and has the properties needed to generate fully rational desiderative convergence, it is the case that we, actually, by nature can be idealised to the fully rational ideal that Smith's meta-ethics requires. So since we and our idealisations embody roughly speaking the same desiderative unity properties and the nature of these properties will determine if there are any normative or moral reasons, then we should require that the reduction of right include these facts somehow.

Smith might reasonably have supposed that since his anti-Humean rationalism is about the desires of fully rational versions of ourselves and our actual commitments to realising that ideal rational desiderative and doxastic profile then his theory gets to generate the correct attitudes towards the objects of moral interest that the fetishism argument shows are important. After all, for agent S in circumstance C act ϕ-wise is right only if it possess the property Fness that a fully rational version of S would want acts for S in C to possess. Either S has the attitude of desiring F-ness or at least wanting to ϕ or counts themselves irrational by their own lights. So Smith appears to have just the defeasible connection to motivation that his rationalist theory needs. But I have argued that this is not enough to secure an appropriate distribution of right making properties and covarying attitudes counterfactually. I think we have a good reason, given the determinative role that fully rational desiderative psychology plays in picking out goods and the determinative role that properties of desiderative unity play in constituting a fully rational agent, to use desiderative unity properties to identify moral goods counterfactually.

We can reference fix on actual desiderative unity. At least we can reference fix on whatever it is that actually plays the desiderative unity role in our

psychology. Then, if there are moral or normative goods, the maximisation of actual desiderative unity will in actual and counterfactual agents pick out the properties that make acts right acts. Doing this has the benefit of knocking out the possibility of moral goods existing in a possible world where no agent does or could value them. This will be because having reference fixed on an element of desiderative psychology rather than something external to it we thereby secure conceptually the covariation of pro attitudes (at least defeasibly) with the properties that are valuable.

So it is not that Smith does not provide the material we need to understand what is going on in his theory. All the claims made so far on his behalf arise out of elaborating Smith's own claims. Rather it is just that the narrow reduction of right to the properties of acts appears immediately to require we given an account of the role of desiderative unity and agential psychology in the effecting the narrow reduction. Smith defers this discussion to the arena of squaring the narrow reduction of right to the natural properties of acts with a broader naturalism. That is where Smith discusses the analysis of rationality in terms of desiderative unity and the reduction of desiderative unity. But Smith's fetishism argument shows that this is not the right move. The identification of 'right' with any property has to also give an account of the necessary covariation of a defeasible rational desire for that property. And now we know that this account has to exclude of the possibility of the actual natural right making property of acts retaining the name 'right making' in the absence of the relevant defeasible ideal desire.

3.4.2 A simple argument to the same end

Smith assumes that if his conceptual theory of rationality is correct then where ever an act is right for an agent relative to a circumstance then they either want to act that way or are irrational by their own lights. This seems fair enough since his conceptual theory makes 'right' a matter of the desires of the fully rational version of agents relative to the circumstance that those agents find themselves in. Of course agents are not, by conceptual fiat (as Smith puts it), fully rational. To make agents fully rational by conceptual fiat is both question begging and probably false. But are agents rational at all? And why do we or should we care? Smith does not say much directly about this but he does say that for all he knows there are no normative or moral reasons, as his theory describes them. I have argued above that facts about the overlap or other wise of rational idealisations of us depend on the nature of desiderative unity and that the best way to understand this is that actual desiderative unity might not idealise in the manner required for normative and moral reasons (i.e. convergently). We should reference fix on desiderative unity to get the counterfactual cases right and ensure the covariation of 'right' makers and the correct attitudes towards them and to show how it is that Smith's conceptual theory could be substantially false. If we do reference fix on actual desiderative unity we have an additional benefit. Smith's conceptual theory, by reference fixing on actual desiderative unity, can explain why if the theory is true we care about the desires of fully rational versions of ourselves. The explanation is simple enough. We are constituted in just the same way as the fully rational agent in the important respect of desiderative unity. We instantiate desiderative unity and are responsive, in principle, to the demands for desiderative revision that desiderative unity makes. It is not that actual agents happen to like

rationality. Nor is it that all agents by nature must be idealisable to full rationality from their starting point. Rather it is that we happen to have a psychology that can display the desiderative convergence at full idealisation – we happen to be rationalisable contingently. To share the concept of rationality as Smith describes it the population of actual agents have to share a psychological constitution that at least in principle disposes them to a convergent desiderative rational ideal. And if things are not as Smith describes them then our constitution does not dispose us to a convergent desiderative rational ideal. This is what 'rational or irrational by our own lights' is about. Reference fixing on actual desiderative unity captures this feature of Smith's theory.

Section 3.5 Smith's Rawlsian account of desiderative unity and why it is inadequate

In sections 3.1 - 3.4 I have argued that desiderative unity is the primary component of rationality and so in turn the primary component of right. I have given two arguments for reference fixing on actual desiderative unity – the argument that it explains how Smith's conceptual theory is substantively defeasible and the argument that Smith's fetishism objection to externalism shows that appropriate attitudes must necessarily covary with the properties of the narrow reduction of right. Desiderative unity, whatever it is, is vital to Smith's anti-Humean rationalism.

Smith gives an account of desiderative unity (and coherence) by making an analogy to a Rawlsian reflective equilibration between moral beliefs. In this section we will examine this analogy. I will argue that it is not sufficient to

give an account of desiderative unity given the role desiderative unity plays in Smith's theory of determining the existence of normative and moral reasons and so determining if there are any moral properties at all.

I will present two objections to the adequacy of using Rawls to characterise desiderative unity. The first will turn on how a Rawlsian reflective equilibration among moral beliefs is nothing more than a contingently useful epistemic tool, if it is useful at all, because according to Smith's theory facts about the maximisation of desiderative unity and the desire sets this brings about in the fully rational are the truth makers for moral beliefs. The second will argue that an analogy to Rawlsian reflective equilibration applied to the contents of sets of desires fails. Smith sketches equilibration as the subsumption of more specific desiderative content under more general desiderative content[26] and this fails because it can't deal with competing general desires. Alternatively the subsumption sketch is bolstered with an unexplained notion of explanatory structure. At this juncture of Smith's theory an unexplained notion of explanatory structure is arguably fatal to his naturalist position.

Section 3.5.1 Smith's use of Rawlsian reflective equilibrium

Smith does not give a theory of desiderative unity except to suppose that it functions to generate desires whose contents are more general and that serve to 'explain and justify' desires whose contents are more particular. This is a straight analogue of the Rawlsian reflective equilibration for moral beliefs and Smith supposes that it can generate general non-derivative desires and

[26] TMP, pp. 159-160

extinguish more particular desires. But the driver of this process, so far as it is described at all, is the structuring of desires into explanatory clusters – where explanation relationship between desires appears to be effectively subsumption of particular desires (desires with more specific contents) under more general desires (desires with more general contents).

Smith does not think much useful determinative information is readily available about the desiderative profile of a fully rational agent. This amounts to not requiring a definitional reduction of desiderative unity. But even so that appears to leave us with only the structure of clustering desires whose contents are more specific under desires whose contents are more general. I call this the subsumption of specific desires under general desires. I don't suppose that Smith thinks that this is all there is to the relationship of desiderative explanation but I will argue that he can't leave this as the only thing we know about it explicitly. And the analogy to a Rawlsian reflective equilibration among moral beliefs where

> "… we might find that our specific value judgements would be more satisfyingly justified and explained by seeing them as all falling under a more general principle." (TMP, p 160)

does exactly that. Supposing explanation is only subsumption is something I am going to argue against. To be clear, again I don't suppose Smith thinks that explanatory structure among the desires of fully rational agents is merely subsumption. Rather I think Smith supposes that we can't analyse what this 'explanatory structure' really is and that this is acceptable nonetheless. It is this supposition that I aim to criticise. The supposition that a Rawlsian reflective equilibrium tracks this idealised structure is, I think, only plausible

to the extent that you suppose that there is an embedded and determinate feature to all agents that is sufficient to determine the nature of the idealised fully rational agent for all such agents. And we are primed to suppose this is not so very odd when we are encouraged by Smith to notice that in the case of the indubitably common place supposition of colours in the world we make the same sort assumption about background theory, one about the perceptual psychology of *all* agents possessed of the colour concepts and when Smith argues that the reduction in the colour case is much like reduction in the ethics case. But are ethical suppositions like perceptual attributions? Smith appears to say as much in places (see in particular TMP, pp. 191-192). We will consider the disanalogies between the ethics case and the colour case in detail again in the next chapter. Setting aside the claim that ethical deliberation is a species of perception there are too many relevant disanalogies between the colour case and the ethics case for them to be treated as relevantly alike without careful specification of the respects of similarity and our uses of them. So in the case of colour reduction the lack of an a priori and complete explicit physicalist theory of colour experiences is to a great extent rendered undisturbing by background theories. They are the folk theory of colours and colour perceptions which assumes inter-agential internal similarities and the adequacy of a causal characterisation of mental states and colour properties which are supported by a scientific neurophysiological theory that supplies very good physical candidates for realisers of the folk theory. But with the ethics case we don't have these sorts of background theories.

3.5.2 Truth maker objection to using Rawlsian reflective equilibration of moral beliefs to give an account of desiderative unity

Smith gives us the following:

> "For exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of desires, just as exhibiting unity is partially constitutive of having a systematically justified, and so rationally preferable, set of beliefs.
>
> The idea here is straightforwardly analogous to what Rawls has to say about the conditions under which we might come to think that we should acquire a new belief in a general principle given our stock of rather specific evaluative beliefs. For we might find that our specific value judgements would be more satisfyingly justified and explained by seeing them as all falling under a more general principle. The imaginary set of beliefs we get by adding the belief in the more general principle may more in the way of unity than our current stock of beliefs, just as our imaginary set of desires may exhibit more in the way of unity than our current set of desires." (TMP, p 159-160).

The objection to the analogy between increases in desiderative unity and a Rawlsian reflective equilibration among moral beliefs is simple enough. Smith's conceptual theory makes it that case that the maximisation of desiderative unity, whatever it is, is a property of the desire sets of fully rational idealisations of us. The normative and moral reasons are just a subset of these desires indexed to all the normatively and morally relevant circumstances agents of the actual and ideal populations could find themselves in. Normative and moral reasons conceptually require that

desiderative unity maximisation leads to a convergence or overlap in the desires fully rational idealised populations have concerning the actions agents relative to the circumstances they can find themselves in. Moral beliefs, even those found by a Rawlsian reflective equilibrium, are made true or false by facts about us and our relationships to this idealised population's desires and by facts about that idealised population's desires. In short facts about desiderative unity are the truth makers for moral beliefs. This means that moral beliefs produced by a Rawlsian reflective equilibrium are not made true *because* they are the product of the reflective equilibrium. Rather, so far as Rawlsian reflective equilibration among moral beliefs is useful at all, equilibration is an epistemic tool of only contingent value. The option of a defeasible conceptual connection between a Rawlsian reflective equilibration among moral beliefs and desiderative unity increase in the corresponding desire sets is cut off as well. This is because such a link would be inexplicable in Smith's conceptual theory. The defeasible conceptual links between moral judgements and motivations in Smith's theory are mediated by the 'rationality' of the agent making these judgements. This 'rationality' in actual agents has to be defeasible too and as we saw above it is the role of actual desiderative unity and its relationship to maximisation of (actual) desiderative unity and between this maximisation and the right kind of desiderative convergence that shows how this works. There is nothing to play role of maintaining and explaining how Smith's conceptual theory could be correct conceptually and false actually if any species of belief is a sufficient determinant conceptually of desiderative rationality. And this includes Rawlsian reflectively equilibrated moral beliefs.

If we made Rawlsian reflective equilibrated beliefs the conceptual determinant of desiderative unity increase between desire sets then we face

the question of what makes moral beliefs true. At least Smith faces this question because one of his primary goals in TMP is to preserve, in his theory, the status of moral judgements as a species of truth apt belief (i.e. the cognitivism). Quite obviously answering this question we would with the current proposal present something circular – ranging from moral beliefs through the desires of the fully rational where those very desires are determined by the very beliefs at question. What is objectionable, even from Smith's point of view, is not that we have a circularity of conceptual dependence here but rather that this circle is uninformative. There is no indication of where one is to look for a theory to cash out or explain how this proposal works – let alone enough information to guide a summary style two step reduction of the sort that Smith favours and needs here (for the naturalistic reduction of desiderative unity) to square the analysis and reduction of 'right' with a broader naturalism. The objection then is that trying to use Rawlsian equilibration among moral beliefs as a conceptual determinant would be *viciously* circular Rawlsian equilibration among moral beliefs as a conceptual determinant would be *viciously* circular[27] in Smith's anti-Humean rationalist ethics

[27] I will define vicious circularity by way contrast to the acceptably ineliminably circular interdefinitions that Smith thinks are found in what he calls the dispositional analysis of colour and the summary style anlaysis of right. A vicious circularity, for two interdefined terms for example, is one that is ineliminable, or at least one that is taken as uneliminated, but which is uninformative. To be uninformative is to fail to provide any useful information to an account of the reduction of one of its terms when used in a reductive argument - like the reductive arguments Smith gives for 'colour' to physical properties or for 'right' to properties of acts (roughly speaking). One way the presence of a viciously circular definition can be diagnosed in a reductive context is if no reductive account can be give one or other of the terms. So the reduction of red to a physical property using the circular definitions of red relative colour experiences and vice versa could turn out to be using a viciously circular interdefinition of red and red experiences if it turned out that no account of the reduction of colour experiences to physical properties could be had. The accounts do not have to be analytic reductions.

in Smith's anti-Humean rationalist ethics.[28]

This sort of failure matters here because we are considering ways to use Rawlsian reflective equilibration among moral beliefs to provide a background theory of desiderative unity – in particular of desiderative unity increase. We need this for two reasons. The first, which we will tackle again in the next chapter, is that we need to show how Smith's analysis and reduction of 'right', to the natural properties of acts that fully rational idealised populations desire relative to circumstances, can be squared with a broader naturalism. The centrality of desiderative unity and the constitutive role it plays in desiderative rationality and so, ultimately, the constitutive role it plays in 'right' shows that an account of desiderative unity is where we must go to find this background 'squaring' account. The second reason arises out of the arguments I make in section 3.4 - that Smith should differentiate actual and ideally convergent desiderative unity use the distinction in formulating his theory of 'right' and 'rational'. Smith's fetishism argument showed that the reduction of right requires the necessary covariation of the correct attitude of the actor with the natural properties of actions that constitute the narrow reduction of right. Including actual desiderative unity in the conditions for the narrow reduction provides both the correct attitude, it's necessary covariation with the natural properties of acts that make acts right acts, and provides both these in a manner that is defeasible in the two

[28] It is worth noting in passing trying to take this position puts pressure on the cognitivist character of Smith's theory. You might accept vicious (that is uninformative) circularity and accept that you stand in need of an explanation of how a Rawlsian reflective equilibration actually works, though I doubt Smith would. But if you did then you thereby accept putting the status of moral judgments as truth apt beliefs in jeopardy. This is because we have already abandoned, on this proposal, the best candidate for a truth apt treatment of moral beliefs and simultaneously have made them self-justifying in a manner not necessarily common among truth apt beliefs. Expressivism seems well placed to replace cognitivism at this juncture.

senses Smith thinks are required for his theory. First actual desiderative unity does not necessarily covary with the mechanism for desire revision in the direction of unity increase – so an agent can be irrational in this manner and so fail to acquire the motives that true moral beliefs require they acquire. Secondly using actual desiderative unity[29] in Smith's conceptual theory shows how it could be correct conceptually and substantively false. Collapsing desiderative unity into a conceptually mandated constitutional dependence on Rawlsian reflective equilibration among moral beliefs prevents desiderative unity, actual or otherwise, from playing these roles.

3.5.3 The incompleteness objection to using Rawlsian reflective equilibration among the contents of desire sets to give an account of desiderative unity

Smith, in the quote above, offers an analogy between Rawlsian reflective equilibration among beliefs and the justificatory structure between more general and more specific desires. I propose this is best understood as a claim about the relationships between the **contents** of sets of desires (if for no other reason than this is significantly similar between sets of moral beliefs and unified desire sets given Smith's theory). First I will consider reflective equilibration among the contents of desires as a minimally characterised subsumption relationship between desires with more general contents and desires with more specific contents and show that this minimal approach fails to distinguish rationally preferable desire sets as it should. So conceived this account is inadequate because it is incomplete.

[29] Actual desiderative unity is conceived as properties of the relationships between the contents of sets of desires here. This is the topic discussed in the next segment of section 5.

Then I will then consider a kind of enriched subsumption approach that includes the justificatory relationship among the contents of desire sets – that is the relationship of explanation of the contents of more specific desire by the contents of more general desires. This additional normative term either has to be given some explicit content of it fails to give any information about what desiderative unity is. There is no explicit content given for the notion of desiderative explanation so adding it into the Rawlsian account of reflective equilibration among the contents of desires adds no information to the inadequate minimal version of Rawlsian subsumption account of desiderative unity. This account is thus incomplete as well.

3.5.3.1 Reflective equilibration as subsumption

The Smith quote above shows how reflective equilibration among the contents of desires is supposed to work. Desires with more general contents are adopted if their content justifies the contents of more specific desires by explaining those more specific desires. This is raised as an analogy to Rawlsian reflective equilibration among moral beliefs. And beliefs are a domain where you might plausibly suppose explanation is understood to some extent. But since it is not clear what desiderative explanation is perhaps just the pattern of subsumption relationships between general and specific contents of desires in sets of desires is enough to sketch desiderative unity.[30]

[30] Whatever it is, this 'explanatory' relationship is **not** the relationship between means ends beliefs and non-derivative and derivative desires. This is simply because the specific desires that are 'explained' by more general desires are all non-derivative desires that exist prior to the general desires that this form of consideration (reflective equilibration among the contents of desires) would lead the rational to adopt. Means end reasoning moves a desiderative profile from the more general to the more specific roughly speaking, not the other way as we are imagining here.

There are possible revisions of sets of desires towards the supposedly converging desires of fully rational versions of ourselves that only desiderative unity is available to bring about. According to Smith the deliberation about unity increase is the main means to bring about non-derivative desires in the fully rational, though it is not the only means. Unique to desiderative unity and coherence however is that their increase either constitutes or is conceptually such as to necessarily covary with a justificatory structure in desires. The rational are able to respond to beliefs about desiderative unity increase by adopting the relevant desires. But what we know full desiderative rationality is according to Smith's theory is only that it involves the increase or maximisation of unity among desires.

So we have the recognition of unity increase and the subsumption account of desiderative unity to work with. We know that if this account is adequate and Smith's theory is actually true then unity increase will be the means by which competing moral explanations are settled. That is we can make sense of the idea of competing explanatory moral hypotheses. The subsumption model of desiderative unity would describe competing explanatory moral hypotheses as different general desires competing to subsume the same more specific desires.[31] Given this the following should be possible – desiderative unity increase could be the sole difference maker between these competing hypotheses.

But how is the mere subsumption of general desires under specific desires to go about explaining how competing general desires could be 'rationally' chosen between? It appears that this stripped down Rawlsian kind of story of

[31] This simplifies real competition between moral theories somewhat, but harmlessly so for our purposes here.

desire revision won't give an account of a flat conflict between different general desires over the subsumption of the same group of more particular desires. I am assuming that the conflict is resolvable in the direction of a more rational desire set. We don't need to know how desiderative unity retrieves this sort of change in sets of desires to generate more rational or more systematically justified sets of desires to know that it has to do just this sort of job sometimes. And I suppose that we do know that it does need to do just this sort of job sometimes because clashing moral theories appear sometimes to do equally well in other dimensions of rational evaluation. If we want to deny this appearance we still have to consider the *possibility* of such a balance and it seems radically implausible, to me at least, that we would want to say that it is *impossible* to have competing moral explanatory theories where we should suppose that there is no rational tiebreaker other than that provided by the subsumption of specific desires under more general ones in fully rational versions of ourselves.

Again, Smith does not suppose anything much about the explanatory structure of the desiderative profile of fully rational agents. However he does lean heavily on the use of both the analogy to the Rawlsian reflective equilibrium amongst moral beliefs and the supposed like equilibrium among the contents of desires. It is not that Smith thinks this anything like a complete explicit theory of desiderative unity is on offer making these sorts of observations about explanatory relations among desires. But then I am not arguing that he does. I am rather arguing that we can know immediately that the partial description of desiderative unity offered by analogy to a Rawlsian reflective equilibrium is inadequate. Moreover we can say that Smith's own theory offers a promissory note to the effect that an adequate story will be offered at some point and I argue that point is not deferrable in the manner

that Smith suggests we defer it.[32] More needs to be said about desiderative unity than that it is similar to the property described, however partially, by a Rawlsian reflective equilibrium among moral beliefs.

Another way of putting the same point again is just this. The possibility of conflicting general desires competing to subsume the same or near enough the same set of more specific desires shows us that the structure of subsumption is NOT enough to account for the idea of explanatory structure in the desiderative profile of fully, or more, rational agents. And just as it is in the colour case[33] some gesture has to be made at what a complete theory about moral explanation is going to look like. In the colour case a reductive theory of mind (one that I suppose is independently plausible) is offered. Smith does not have such a thing available in the case of ethics.[34]

The subsumption version of a Rawlsian account of desiderative unity is incomplete. It cannot deal with possible cases of general desires competing to subsume the same set of more specific desires correctly. It is at least possible

---

[32] Smith claims that we can simply assume that some story can be given of the reduction of normative terms that crop up in the analysis of 'right' and 'rational', much like the one given for the reduction of 'right' (TMP, p.186). Since I suppose that his assumption here depends on his analogy to the colour case and how in that case a narrow reduction of colour in objects can be reduced to physical properties of those objects can be squared with a broader physicalism we will leave this until the topic is more thoroughly discussed in the next chapter. I argue there that this analogy is flawed and Smith can only make use of it by altering his anti-Humean rationalism to allow desiderative unity to be treated like a natural kind term.

[33] The complete theory of colour according to Smith is the one where a narrow reduction of colours to surface reflectance properties is squared with a broader physicalism. Showing how colour experiences and, ultimately, all mental states can be reduced to physical states does this.

[34] But we do know some things about desiderative unity and fully rational desiderative profiles if Smith's rationalism were true. I will suggest here and latter that we can use this information to go part way towards fulfilling the promissory note Smith offers in TMP. That is the promise of squaring the reduction of right with a broader naturalism - which will require a theory of desiderative unity.

that where there is such competition one general desire is rationally preferable. But for Smith rational preference for a set of desires is governed by the desiderative unity differences between sets of desires, where the preferable set of desires displays the most unity.

Defending Smith you might want to claim that the condition I am imagining is impossible after all – that is that it is impossible for two general desires competing over the subsumption of the same set of more specific desires to be discriminated between by rationality.[35]

I argue this digging in strategy is no good for Smith since any merely theoretical commitment that induces error theory is to be avoided if at all possible. Given the role of desiderative unity and fully rational desires denying that it is possible for there to be two general desires competing to subsume the same set of more specific desires where general desire is rationally preferable to the other does not eliminate the possibility of two general desires competing to subsume the same set of more specific desires. Instead it forces the approach being considered here to treat two different desiderative profiles as equally good. This possibility is not necessarily bad for a minimalist subsumption account of desiderative unity but it does allow for a possibility of an error theory inducing relativism. Smith's objects error theory inducing relativism if it is brought about for the wrong reasons. This topic is the subject of the next chapter. I suggest that issuing a promissory note for a theory of desiderative unity and the associated property of explanatory structure in fully rational desires is preferable to the digging in strategy for Smith it looks like it induces relativism for the wrong reasons.

[35] Importantly, and with difficulty I shall argue, you could try to claim this and still think that Smith's anti-Humean rationalism works.

So could you say instead all there is to desiderative unity is the creation of a set of desires displaying the structure where the content of more general desires 'explain' the contents of more specific desires instead and not offer a theory of desiderative explanation? Here we are imagining that we don't know explicitly anything more about 'explanation' in this context than that it involves subsumption pattern. I am imagining the consideration under examination would be something like the evaluation one might make of competing incompatible moral theories over their account of the agreed upon common ground of more specific moral facts.[36] Now we can see, I hope, that the view I am trying to evaluate is one where competing general desires is an analogue of competing general moral theories and that where they are competing over the *same set of desires* we effectively stipulate out the appeal such general theories can make to the more specific moral facts that one might use to resolve this sort of dispute. This is just because the relevant moral facts are, according to Smith, the *desires of fully rational versions of ourselves*. With no differences in the relevant desires to appeal to we are left, or so goes the view, with either no competition (and so a disjoined moral theory) or competition of an unresolvable sort (from the rational point of view) and so an error theory. But for all this we must remember here that according to Smith in TMP all we know about 'explanation' in the context of desiderative unity is that it is reflected in the subsumption pattern between the more general contents of desires and the more specific contents of desires. We should also remember that according to Smith the truth makers for all moral beliefs are the overlapping desires in the sets of desires of fully rational agents. It is the properties of this population and its desiderative profiles

[36] 'Moral facts' as I use the phrase is not about moral judgments. Moral facts are the things that make moral judgments true or false.

and, presumably, of that set of desiderative unity properties that we, irrational as we are, share with them that make moral beliefs true. This is the source determining what desiderative unity and desiderative explanation amount to. Appeal to actual moral discussion does nothing to change this nor, until we show otherwise, does it ameliorate our current ignorance of what it is about desiderative unity (its explanatory role or whatever else it might be) that allows it to play the role it has too, again whatever that is, to make moral beliefs true or false. We can consider the abstracted features of the matter usefully then since even according to Smith those are the features we have ready to hand.

I think there are two reasons why the move of sort described above would not suit Smith's position:

The first is just that though there may be conceptual grounds for error theory – conceptual incoherence effectively – this way of getting it seems too easy. Also it is not the way Smith recommends we think about theoretical causes of error theory (we will see this in the next chapter in more detail). The way the story has developed here we get error theory out of restricting our account of desiderative unity to the subsumption structure discussed above. But finding that error theory is a ready upshot of such a restriction is a good reason to look for an alternative understanding of desiderative unity. So far all I am proposing is the following: desiderative unity is not only a matter of finding a structure of more general and more specific desires where the general desires 'explain' the more specific ones. Here the reason for accepting, at least provisionally, my proposal is that with it there is some chance of avoiding error theory under the conditions I am describing. Smith himself offers and

accepts just such considerations when he rejects the idea that ethical kinds can be treated as natural kinds. I take that discussion up in the next chapter.

The second reason is that unless there is more to be said about desiderative unity Smith appears to face the following difficulty. How is it that desiderative unity is an ***additional*** feature of correct deliberation on the ones offered by Williams? Smith quotes this from Williams.

> "…there are much wider possibilities for deliberation, such as: thinking how the satisfaction of elements in … [one's set of desires] … can be combined: e.g. by time-ordering; where there is some irresoluble conflict among the elements of … [one's set of desires] … considering which one attaches most weight to …; or, again, finding constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment. (1980:104)" (Smith, TMP, p 157)

My concern here is the last fragment of the quote where Williams' supposes correct deliberation involves finding amongst other things "constitutive solutions, such as deciding what would make for an entertaining evening, granted that one wants entertainment.". This looks very much like it is the finding of a desire with more specific contents to satisfy a desire with more general contents. If desiderative unity were just restricted to adding general desires whose contents explain more specific ones and removing more specific desires whose contents don't fit well with a more general desire then the only difference between it and Williams' "constitutive solutions" is the

pre-existence of the relevant general desire.[37] But this seems trivial. Finding that one is interested in particular entertainment options seems like a good way of finding out that one is interested in an entertaining evening. If this is right then Williams' correct deliberation already has a mechanism like the one that Smith is describing with desiderative unity. But that can't be right because for all that Williams' has to say finding 'constitutive solutions' looks just like a species of means ends reasoning covarying with desire revisions. There is a new direction of desire revision in desiderative unity, as it is described here, that is not found in means end reasoning. However in the absence of a notion like 'explanation' playing a very significant role the subsumption account of desiderative unity would add little more to Williams' account of full rationality. If the subsumption story were all we had for an account of the nature of desiderative unity among desires then, though we have added something to Williams' constitutive solutions we seem to have only added that it is sometimes the case that a specific desire implies a more general one. This is not strictly speaking a trivial matter. But it is the case that desiderative unity is, for Smith, intended to do more than supply the general desires that some specific ones may be 'constitutive solutions' for. Smith intends that more come from the notion of desiderative unity.

The use of the term 'explanation' in Smith's discussion of the relationships between desires in the desire sets of the fully ration is load bearing. Without significant discriminatory effect found somewhere in desiderative unity Smith's position is a non-starter. Smith can't dig in and hold that the

[37] Though I am using the word 'explain' here because Smith does I am also assuming that for the time being the only thing we know its use betokens is the presence of a structure of desiderative subsumption of desires with more specific contents under desires with more general contents. If we can't sustain this restriction here then I hope to thereby show that a proponent of Smith's position can't either.

subsumption version of the Rawlsian account of desiderative unity is complete (or near enough). To do so requires providing a treatment of general desires competing for the subsumption of more specific desires that forces error theory under conditions where error theory should not occur. Insisting on the subsumption version of the Rawlsian account of desiderative unity also undermines the role that Smith intends desiderative unity play in his theory – that is that it is distinctive and distinctively a normative way to sort sets of desires.

3.5.4 The objection to the explanatory version of the Rawlsian account of desiderative unity

The last proposal we will consider for the use of Rawlsian reflective equilibration is to supplement the subsumption version of Rawlsian reflective equilibration between the contents of desires with an explicit reliance on the explanatory relationship between desires that desiderative unity involves. We could accept that the notion of desiderative unity is not **just** that there are more general desires that are consistent with clusters of specific desires but that the relationship is that the general desires 'explain' specific desires. We then expect the explanatory relationship between general and specific desires to select amongst competing general desires those that succeed over those that don't at explaining the specific ones. If we accept these points then I think we have implicitly accepted that we suppose there is some substantial theory of desiderative unity. It is just that we don't know exactly what it is and we expect that it contain some information about what 'explanation' means in the context of desires. But we might think that this is an acceptable area in which to accept a promissory note rather than require a fuller account. We might think that this is just like the squaring of colour talk with a broader

physicalism, where we accepted that the analysis and reduction of colour did not require the explicit reduction of mental states to physical states. Instead we simply acknowledge that a physicalist reduction of mental states has to be supplied or the reduction of colours ultimately fails to be vindicated. But as we will see in the next chapter this is one the areas where the colour case and the ethics case are very different. There is no background theory of desiderative unity or explanation that can motivate accepting this promissory note.[38]

Put simply this strategy is incomplete because we know nothing explicit about desiderative explanation except that desiderative unity either constitutes it or is constituted by it. Either way Smith's theory is incomplete without some account of how this feature works. Reflective equilibration among contents of desires effectively presumes that we can recognise the explanatory or unity increase relationship between the contents of sets of desires upon reflection. This assumption is not an adequate account of desiderative unity; it is no account at all.

Since analogies to a Rawlsian reflective equilibration among moral beliefs fails because facts about desiderative unity are what makes these beliefs true any putative plausibility that it has in moral epistemology is not available when using it as an account of desiderative unity. We have no explicit or even potential account of either of desiderative unity or explanatory relations between the contents of desires as a yield from attempting to use a Rawlsian

[38] You might suppose that the broad Rawlsian reflective equilibrium among beliefs could be the beginnings of a general theory of explanation and so offer a location for the development of a notion of desiderative explanation. This faces the same problems as using the narrow Rawlsian reflective equilibration among moral beliefs as conceptually determinative of desiderative unity so I will set it aside here.

account of desiderative unity. The explanatory version of the Rawlsian account of desiderative unity fails outright.

3.5.5 Concluding remark

The analogy to Rawlsian reflective equilibration to suggest an actual or potential account of desiderative unity appears to be a comprehensive failure. Given Smith relatively simple brand of cognitivism this is perhaps not surprising. Smith's cognitivism forces facts about desiderative unity into the role of truth makers for moral judgements. This has the knock on effect of limiting the role that a Rawlsian reflective equilibration among moral beliefs can play to that of an only contingently valuable epistemic tool. As such its ability to retrieve the justificatory structure from the desiderative profiles of the fully rational is entirely derived from facts about that structure. A reflective equilibrium finds moral facts; it cannot for Smith make moral facts.

I am interested in this attempt at accounting of desiderative unity for two basic reasons. Given how desiderative unity is constitutively important to normative and moral reasons and so to moral properties we require at some point an account about how it works OR a plausible promise that some line of enquiry will throw up the required account eventually. More pressingly, for Smith's summary style analysis and reduction method, an account of desiderative unity is the only place to find a new reason to preserve the semantic gap between the analysis and reduction of right and the analysis and reduction of rationality. Without this semantic gap Smith's method collapses into a variant of Jackson's moral functionalism. And, perhaps more importantly, it collapses into an example of a definitional network analysis reduction version of naturalism.

The Rawlsian account of desiderative unity fails. Consequently it cannot provide a new reason for maintaining the semantic gap and avoiding a definitional network analysis reduction version of naturalism. The next chapter will return to the analogies and disanalogies between the colour case and the ethics case. I will argue that the way that the colour case goes about satisfying what I have been calling the 'squaring' requirement suggests an account of desiderative unity – treating it as akin to a natural kind – that does avoid the collapse into definitional naturalism.

3.6 Next

Desiderative unity is central to Smith anti-Humean rationalist meta-ethics. It's what differentiates his position from ones like Bernard Williams. It is the determinant of what makes one set of desires rationally preferable to another. It is the part of Smith's theory that has the job of giving an account of the possibility of the kind of convergent rationalism required for normative and moral reasons. In this chapter I argued Smith should further distinguish between instantiated or actual desiderative unity and how it idealises, and add this distinction to a convergent model of desiderative unity. This has the benefit of maintaining the covariance of the right attitude, or its possibility, with moral goods even in counterfactual cases. This addition also allows Smith to describe how his conceptual analysis can be correct and none the less false.

Desiderative unity is a plausible source for a reason to maintain the semantic gap between the analysis and reduction 'right' and 'rational'. But to evaluate

this we need an account of desiderative unity. Smith sketches a Rawlsian account but this fails comprehensively either because a Rawlsian reflective equilibration among moral beliefs is an epistemic tool contingently tracking facts about desiderative unity or, if used as an analogy, does not provide any useful information about desiderative unity or desiderative explanation.

We are left with no new reason to maintain the semantic gap and vivid reminders that we need to square the reduction of right to natural properties of acts with a broader naturalism – and this appears to require an account of desiderative unity.

Chapter 4 will consider how Smith argues that the reduction of right is squared with a broader naturalism and show that it is flawed. However, modifying Smith's theory to allow a natural kinds treatment of desiderative unity will provide a reason for the semantic gap and give a plausible reason to accept the current absence of a theory of desiderative unity and the promise that one could be found eventually.

Chapter 4 A dilemma for Smith's meta-ethical theory

In this chapter I argue that in the end what Smith needs to maintain the semantic gap is a natural kinds treatment of the term 'desiderative unity', something that he himself has argued against. This presents Smith with a dilemma. On the first horn, he stays with his disaffection for such a natural kinds treatment and accepts definitional naturalism. The first horn of the dilemma is mostly discussed in chapters 2 and 3. On the second horn, Smith avoids definitional naturalism but at the cost of adopting metaphysical naturalism. The second horn of the dilemma is discussed in this chapter.

Recapping chapter 2 and 3 we get the first horn of the dilemma:

Smith's narrow reduction of right conceptually connects right to what determines the narrow reduction (anti-Humean rationality, in particular the desiderative unity feature of it) of right and the particulars of this determination are known a priori – this is a strong reason to make the relationship between 'right', 'rational', and 'desiderative unity' explicitly definitional. Smith needs a reason to keep the semantic gap between the analysis and reduction of right and the analysis and reduction of rational or his anti-Humean rationalism collapses into a definitional network analysis and reduction. Smith's reason for keeping the semantic gap is his permutation problem argument against definitional naturalism. The permutation problem argument fails, however, and so Smith needs a new reason to keep the semantic gap should he wish to resist definitional naturalism.

We saw in chapter 3 that desiderative unity is central to Smith's theory, and this suggested that the nature of desiderative unity might provide a new reason for keeping the semantic gap. Smith himself provides a Rawlsian reflective equilibration account of desiderative unity, but I argued that any such Rawlsian account of desiderative unity is inadequate. Since Rawlsian accounts of desiderative unity are inadequate, Smith hasn't provided a new reason for keeping the semantic gap. So as it stands Smith's theory has no reason to insist on the semantic gap, and it again looks as if Smith's theory must collapse into definitional network analysis and reduction.

Separately, even if we grant Smith avoids collapse into definitional naturalism for argument's sake, Smith counts it a requirement on the narrow reduction of right to the natural properties of acts that it be squared with a broader naturalism by showing how all the relevant normative terms used in the narrow reduction are reducible to natural properties as well. Chapter 3 has shown that the role 'rationality' plays in determining the referents of 'right' in the narrow reduction of 'right' depend vitally on the function of desiderative unity in Smith's theory. Consequently, squaring the reduction of 'right' to natural properties of acts with a broader naturalism will require showing that all the elements of desiderative unity can be reduced to natural properties. As we have seen, Smith illustrates what squaring a narrow reduction with a broader reductive framework looks like by using the colour case. However, his own squaring argument for desiderative unity is quite unlike the one used in the colour case.

In section 4.1 I will examine Smith's squaring argument and show that it is flawed. Fixing the flaw will require stipulating that a squaring argument in

the case of ethics include an explicitly reductive theory of desiderative unity. We turn to the prospects for such a theory in section 4.2.

Section 4.2 will examine how a squaring argument works in the case of a narrow reduction of colours, since it is the illustrative case for Smith. We will find that the success of the squaring argument depends on a sophisticated set of explicit background theories that explain the narrow reduction of colours. There are several significant disanalogies between the colour case and the ethics case. The most important is that there are no similar background theories for desiderative unity and no reasonable expectation that any will be forthcoming. At this juncture accepting promissory notes looks bad.

Section 4.3 combines the considerations in the previous two sections to show that the failure to provide an explicit theory of desiderative unity, or the reasonable expectation that one could be found, leaves the first horn or the dilemma in place and increases the severity of the problem. Necessarily the absence of an explicit theory of desiderative unity will yield the collapse into a definitional network analysis and reduction of right. But it also threatens that Smith's metaethical theory could fail outright as a form of naturalism.

Section 4.4 proposes an alteration to Smith's theory that gives a reason to accept a promissory note for an explicit theory of desiderative unity and that provides the new reason to keep the semantic gap and so avoid the first horn of the dilemma. Smith should treat desiderative unity as a natural kind term. I show how this can be done and the advantages it brings. The disadvantage is that it requires altering Smith's theory in a direction he explicitly rejects when he argues against metaphysical-but-not-definitional naturalism. In short, it forces him to accept the second horn of the dilemma.

This establishes the dilemma that Smith faces. Either he keeps the a prioristic character of his anti-Humean rationalism and faces both a collapse into definitional network analysis and reduction and the possible failure of his metaethics as a reductive naturalism or he adopts a form of metaphysical-but-not-definitional naturalism. Either his theory fails or he has to alter it.

Section 4.1 Smith's squaring argument in the ethics case

Smith has his own squaring argument for the ethics case. He begins with a framework for what the actual narrow reduction of right looks like in the ethics case. I will reproduce it here just so that we can see both that Smith does suppose the narrow reduction of right is to natural properties of acts, if the actual world is naturalistic, and that he thinks this is not enough to secure naturalism. That is, Smith recognises that to square the narrow reduction of right to natural properties of acts you have to show that all of the relevant aspects of rationality ultimately reducible to natural properties.

So, as we have seen before, we have the narrow reduction of right given by the following argument:

> "Conceptual claim: Rightness in circumstances C is the feature we would want acts to have in C if we were fully rational, where these wants have the appropriate content.

Substantive claim: Fness is the feature we would want acts to have in C if we were fully rational, and Fness is a feature of the appropriate kind

Conclusion: Rightness in C is Fness" (Smith, 1994), p185.

Smith thinks that the narrow reduction of right is naturalistic in two ways. The first is given as follows:

> "And this argument [see above] is, in turn, broadly naturalistic in two respects. First, it is naturalistic in so far as the features that we would want our acts to have under conditions of full rationality, the feature that we would want acts to instantiate in the evaluated possible world [our actual one], are themselves all natural features whenever the evaluated world is itself naturalistic. Our non-reductive, summary style definition of rightness, in conjunction a substantive claim of the kind described, thus allow us to identify rightness with a natural feature of acts in a naturalistic worlds like the actual world: for example, in this case, Fness." (TMP, p 186)

I think this description of the narrow reduction of right is correct as far as it goes, but insufficient. I argued for this in chapter 3 by pointing out that the implication of Smith's fetishism argument against externalism is that Smith must explicitly restrict the narrow reduction to prevent Fness deserving the name right in circumstances C where no actor does or could instantiate the kind of desiderative unity that constitutes convergent desiderative rationality when it is maximised. We might suppose that the qualification "… were we fully rational…" implies that Fness is only rightness in C if Smith's anti-

Humean rationalism were, as he puts it, substantively true (TMP, p 173). But again I have argued that maintaining the possibility that Smith's metaethical theory is the correct explication of our normative concepts while simultaneous false is best served by distinguishing actual desiderative unity from ideal (i.e. convergent) desiderative unity. It is not that the qualification is false but that it is insufficiently precise.

This is effectively covered in chapter 3, but it is worthwhile reminding us that this idea of a narrow naturalist reduction of right to natural properties of acts is entirely dependent on a squaring argument regarding naturalising rationality. This iteration of Smith's summary style method for rationality seems to be part of what he suggests solves the problem of squaring for the reduction of right with a broader naturalism. In TMP, p. 186 Smith says

> "Thus, if we wanted to, we could construct non-reductive analyses of the key normative concepts we use to characterize the normative features of such and idealised creature – the unity, the coherence, and the like of its desires – and then use these analyses to construct two-stage arguments, much like that just given [see above quote the reductive argument of right in C to Fness for us], in order to identify these normative features of a fully rational creature's psychology with natural features of its psychology (for an analogy, see note 9 to chapter 2 [the use of the (Lewis: 1972) strategy in the colour case]). (TMP, p. 186).

So 'rationality' is defined in terms of desiderative coherence and unity, with this summary style analysis being used to inform the reduction of rationality (effectively, to the belief desire psychology of a fully rational creature)

without offering a definition of either the notions of coherence or unity. Smith then says in the very next sentence

> "Coherence and unity, though not naturalistically definable are therefore themselves just natural features of psychology. The evaluating possible world is therefore naturalistic in the relevant respect as well." (TMP, p. 186)

Whatever else these passages imply they do indicate that Smith requires that somehow it be shown that desiderative unity in its idealised form (maximum desiderative unity) can be reduced to a natural property. And again the particulars of the colour case will play a role since they model how non-analytic reductions are done while using ineliminably circular definitions. Smith requires an argument about naturalising desiderative unity in particular since, as I have argued in chapter 3, it is the central normative hub of Smith's rationalism. And it is important to remember that we know, by analysis, why this is the case. Desiderative unity has to condition desiderative rationality in the right way and has to be present as a property of actual psychological states (more particularly as a property of the relations between the contents of desires in a set of desires) for the narrow reduction to occur at all.

The second way that Smith thinks the anti-Humean rationalist meta-ethics in play in reduction of right naturalistic is given in the following passage:

> "And second, even though the analysis is not itself naturalistic – even though it defines rightness in terms of full rationality where this may not itself be definable in naturalistic terms – fully rational creatures in

> the evaluating possible world are themselves naturalistically realized. For a fully rational creature is simply someone with a certain psychology and, as you will recall, a natural feature is simply a feature that figures in one of the natural or social sciences, *including psychology* (chapter 2). Of course, the psychology of a fully rational creature is an idealized psychology, but such an idealization requires nothing non-natural for its realization. Thus, if we wanted to, we could construct non-reductive analyses [summary style network analyses] of the key normative concepts we use to characterize the normative features of such an idealized creature's psychology – the unity, coherence, and the like of its desires – and then use these analyses to construct two-stage arguments, much like that just given, in order to identify these normative features of a fully rational creature's psychology with natural features of its psychology (for an analogy, see note 9 to chapter 2) [it is an analogy to the case of colour terms and squaring colour talk with physicalism]. Coherence and unity, though not naturalistically definable are therefore themselves just natural features of psychology. The evaluating possible world is therefore naturalistic in the relevant respect as well." (Smith, 1994), p 187.

This passage gives Smith's squaring argument[96] for the narrow reduction of right. The argument starts with the claim that psychological properties are, by definition natural properties. Idealised psychologies don't require the addition of non-natural components to be realised ideally so the naturalistic nature of actual psychological states persists for idealised fully rational psychologies. The second phase of Smith's argument makes an analogy to

[96] So the squaring argument purports to show that the reduction of right can be squared with a broader naturalism using the (presumed) fact the reduction of 'rational' can be squared with a broader naturalism.

the squaring argument in the colour case – an analogy that will be comprehensively discussed below. This analogy and its merits are of concern here because I think that there is a fundamentally fatal objection to the first part of Smith's squaring argument. My objection I think also conditions how we should understand the analogy that Smith makes to the colour case in this passage.

The objection I have to the squaring argument presented in this passage is that it permits the possibility of primitively normative natural desiderative unity properties. Desiderative unity properties get to be 'natural' simply because desiderative unity is a feature of the relations between the contents of the desires of a set of desires and so desiderative unity properties are just features of a psychology and thus appears definitionally natural according to Smith (see TMP, p. 17, where he defines natural states of affairs as states of affairs which are the subject matter of the natural or social sciences[97] - including psychology). By 'primitively normative' I mean that desiderative unity by its nature just is such that its maximisation engenders the convergent desiderative rationality whose existence is required by normative and moral reasons. Smith arguably does not intend this possibility in the passage above since he appears to suggest that a two-stage reductive argument for 'rational' can be given and squared with naturalism in the same way that squaring is done in the colour case. That is by showing how

[97] A naturalist of this stripe might argue against the move from states of affairs to naturalistic features (Smith's does this) or the move from naturalistic features to natural properties (I do this). But Smith uses (Lewis: 1972) to provide an analogy here for the reduction desiderative unity. And in that paper Lewis uses Ramsey-Lewis-Carnap sentences, which without much fuss can be applied using terms that name or refer to properties I will simple ignore such argumentative inclinations. We will simple note that making the points I wish to make using property talk is just one way to go about it. Also Smith reduces right to natural properties so at least Smith should not demur.

desiderative unity can be reduced to a natural property. But this possibility is not explicitly blocked.

Why should Smith block this possibility? Initially we might think that primitively normative natural properties are not 'really' naturalistic, but I think this is too quick. We can't object to the squaring of the narrow reduction with a broader naturalism using a primitively normative natural property (desiderative unity) on the basis that this is circular. That is question begging.[98] Also a primitively normative natural desiderative unity has *versions* of at least three virtues an anti-Humean rationalist naturalism should have. I discuss these the three virtues when arguing that Smith's theory should distinguish and use 'actual desiderative unity' and 'ideal desiderative unity' in chapter 3. Such a theory can be false (its central concepts not realised, say), motivationally defeasible if true, and provide a constitutional connection between actual and fully rational agents that explains the relevance of fully rational desiderative profiles.

If it is true that our concepts of normativity permit or even require a primitively normative natural property we can none the less fail to instantiate to any degree this property. The conception of desiderative unity as a primitively normative natural property can be correct and substantively fail to exist as Smith puts the matter. If the primitively normative natural property version of Smith's meta-ethics we are imagining here is substantively true it remains the case that knowing moral facts only entails a *defeasible* connection to actual or operant motivations. Distinguishing

[98] It is question begging because the dispositional analysis and reduction of colours in objects to physical properties and the necessary reduction of the unanalysed term 'colour experiences' to talk of physical states is possible as we discuss in more detail in section 4.2. Circular definitions do not preclude the possibility of usefulness with regard to and consistency with a broader reduction.

contents of desires and the relations between them from the desires themselves and from mechanisms for desire change we can see that desiderative unity as a primitively normative and natural property of the relations of the contents of desires in a set of desires can entail which desire set changes will increase and even maximise desiderative unity without entailing anything about the mechanisms for desire change itself. Knowing what is rational and being rational are thus distinguished in this version Smith's anti-Humean rationalism even if desiderative unity is a primitively normative natural property. Finally I argued in chapter 3 that the interest we actual and typically irrational agents have in the desiderative profiles of fully rational versions of ourselves cannot exist because we actually desire to be more or fully rational because we do not necessarily actually desire to be rational. This leaves the interest in need of explanation and I argued that the nuanced notion of desiderative unity (as a property of the relationship between the contents of desires in a set of desires that has differentiated actual and ideal instances) allows us to suppose that we and fully rational versions of ourselves can share the 'same property' or more accurately members or the same cluster of properties[99] that the ideal agents instance the maximal member of. Treating desiderative unity as primitively normative doesn't change this relationship between less than and fully rational agents. It will just mean we don't expect any further explanation of the desiderative unity cluster of properties ordering structure.

Allowing primitively normative natural properties of desiderative unity does not obviously fail as a form of naturalism since at least some of the

[99] We discussed the niceties of cashing out the notion of desiderative unity increase and maximisation in terms of properties, given that a single property does not of itself admit of increase. There is no need to summarise this discussion here as it has no impact on the argument.

characteristic virtues of naturalism persist even so. My objection to permitting this property is that a primitively normative natural property of desiderative unity is profoundly uninformative in ways that we should not permit. I will discuss this by way of 3 problems for this version of naturalising desiderative unity.

4.1.1 First problem – the puzzle of concept acquisition

The first way that this lack of information is vexing is that when our concept of 'right' and 'rational' depend on a primitively normative desiderative unity **and** the concept is not realised (when this version of Smith's theory is conceptually true but actually false as I have put it) we generate a puzzle about the source of our ethical concepts. Smith's analogy to the colour case brings out this puzzle. Smith thinks that ethical concepts are just like the colour concepts so far as they are directly hooked up to, presumably, the faculties of rationality that constitute them. And this in turn, just like in the colour case, explains why there is little a priori information about ethical concepts like 'right' and why we learn about the use of the concept by exposure to paradigmatic examples. The puzzle is that if our actual psychology does not instantiate a primitively normative natural desiderative unity then we have no explanation of the origin of our concepts. By allowing a primitively normative, and so unreduced, naturalism we have no way of explaining what it is about our actual psychology that mislead us into believing we instantiated primitively normative desiderative unity and why primitively normative desiderative unity formed part of our ethical concepts. This is because we have no theory of how non-normative features of psychology could be such that they even *resemble* a primitively normative

desiderative unity. This is just a consequence of making the normativity in play primitive. And since we don't instantiate the primitively normative property it is hard to see how it came to play the role it is supposed to according to Smith. After all our explicit concepts of 'right' are by hypothesis **not** hooked up to a primitively normative desiderative unity when this kind of error theory is the case.

We have to be careful not to suggest that this 'primitively normative' version of Smith's anti-Humean rationalism requires that the conceptual theory be substantively true if it is the correct explication of our actual concepts. This is an easy mistake to make but is still a mistake. However what this kind of theory can't do is *explain* why we have the concept right that we do if at the same time we think it could turn out not to be a concept that applies to us. It can't even explain **how** this possibility **could** be the case.

4.1.2 Second problem – vicious circularity[100]

As we noted circularity of definitions is not necessarily fatal to a reductive naturalism. The odd thing with the possibility of a primitively normative natural desiderative unity is that, though the first round of circular interdefinition in the narrow reduction of right may not be viciously circular, the definitional components of the analysis and reduction of 'rational' look like they **must** be viciously circular. Smith should reject viciously circular definitions. When Smith accepts the burden of squaring a reduction depending on a circular definition with broader reductions as he does in both

[100] I specify what it means for a definition to be viciously circular in this context in footnote 27, section 3.5.2.

the colour and ethics cases he effectively accepts that vicious circularity is bad. Since rationality definitionally depends on desiderative unity and desiderative unity is being considered here as a primitively normative natural property we have a definition that supplies no information. And since the particular operation of desiderative unity in Smith's theory determines whether rationality supplies normative and moral reasons then this vicious circularity looks likely to have a serious impact on the narrow reduction of right.[101]

4.1.3 Third problem – a primitively normative natural desiderative unity makes moral epistemology mysterious

As we have discussed in chapter 3 the Rawlsian reflective equilibrium among moral beliefs can play the role of a contingently useful moral epistemology in Smith's anti-Humean rationalist meta- ethical theory. Given that it does not have a conceptually mandated connection to the facts that make moral beliefs true we have gaps to plug. We don't know how or why a Rawlsian reflective equilibration among moral beliefs is likely to track the relevant facts. Given the relevant facts are about the relationships between the contents of desires in sets of desires and then the differences between sets of desires determined by this kind of intra-desiderative set content property we might tempted to

[101] The same point can be made without trouble in terms of summary style analyses. A summary style analysis has to be informative and part of that includes showing how a naturalist reduction of rationality is to be done. Desiderative unity does not have to be defined for this to be possible **but** some account of how desiderative unity is to be reduced to a natural property is required. The most plausible reading of this requirement is that desiderative unity terms have to be reduced to other, natural property terms, or the desiderative unity properties have to be reduced to other natural properties, even if they are psychological terms or properties. Taking desiderative unity to be a primitively normative natural property does not do any of this. And if the summary style analysis and reduction of 'rationality' can't be squared with a broader naturalism then neither can the summary style analysis and reduction of 'right'.

assume that the relevant facts are available to some feature of introspective inspection. Certainly the a priori nature of the substantive premises in the narrow reductions of right relative to agents and circumstances appears to suggest as much. But this assumption bears some scrutiny. Even if it is true that facts about fully rational desires are a priori, it doesn't follow that coming to know them is a matter of introspection.[102] This is also true of properties of our psychology, especially if we are inclined to think they are naturalistically realised. We might in a fit of charity pass over these concerns but Smith and fellow reductive naturalists cannot. This is because reductive naturalists characteristically think our moral concepts may not be true of anything, that is that nothing at all actually instantiates our moral concepts. But how do we come to know our moral concepts given Smith's tacit commitment to their deep embedding in our psychology? How can we describe a possibility where we have primitive normative naturalistic moral concepts and no primitive normative naturalistic moral properties for these concepts to refer to?

Such demands for clarification of epistemological issues do not, at the end of the day, mean we can count a current lack of that clarity against Smith's anti-Humean ethics. It is just one of the many 'to do's' that complicated theories throw up, and it is a common problem, not unique to Smith. What is unique to the primitively normative natural desiderative unity version of Smith's ethical theory is that it is not at all clear that any forth-coming epistemological clarification is going to be possible at all. The causal powers to condition

[102] I am assuming the antecedent of this conditional. I believe this is reasonable even though the epistemology of a priori truths is itself a significant and complex topic. If one wants more argument for the conditional itself, then it is important to keep in mind that conceptual facts, according to both Smith and Jackson, are both a priori and can be complex and novel. Given this, it seems to me unreasonable to assume that a claim's being a priori true entail introspective accessibility to its truth, and use this as a cornerstone of an epistemology for a priori truths.

desire sets in the direction of desiderative unity increase are not a necessary property of desiderative unity. Desiderative unity does not have an essentially causal impact on actual desiderative profiles so we can't assume that changes in desiderative profiles will provide information about changes in desiderative unity. Yet Smith's theory, if substantially true, requires that we be able to detect desiderative unity differences. But again it must be defeasible or we lose our ability to explain moral debate and progress by losing the ability to distinguish moral errors. We know that we suppose that there is an accessible moral epistemology and that it admits of error and correction. Smith argues that we have reason to be optimistic about the possibility of moral progress. But if desiderative unity is primitively normative and instantiated both actually and ideally and we know about it to some extent why do we ever make moral mistakes? If it is a property we effectively detect somehow then is it a perceptual faculty that does the trick of supplying moral beliefs and if not why not?

The problem unique to a primitively normative natural desiderative unity is that it is not clear that any theory of its epistemology is going to be possible. Even supposing that, because it's a primitive, the core of moral epistemology is detection of desiderative unity, it is still unclear why you would have any expectations about its epistemological availability in public domains like moral reasoning. This is because we have separated the normative theory of moral reasons from a causal theory of psychological change and motivations. And this separation is not coincidental. It is essential to Smith's anti-Humean rationalism. Though Smith is anti-Humean about normative reasons he is not anti-Humean about motivational psychology. The conceptual connections between beliefs and actual desires must necessarily remain defeasible of the distinction between beliefs and desires that Smith agrees should be preserved

will be lost. And that distinction is just that coming to form moral beliefs is not sufficient to entail that you come to have relevant desires. This looks like it means that beliefs about morality can't be substituted for properties of desires in this context either. We won't make progress on solving Smith's moral problem if we make the feature of psychology that is primitively normative and naturalistic beliefs about morality instead of relationships between desires actual and ideal.

In short, allowing a primitively normative desiderative unity will make moral epistemology insolubly mysterious.

4.1.4 Desiderative unity can't be a primitively normative natural in Smith's meta-ethics

The three problems outlined in 4.1.1-3 amount to an objection to opting for a primitively normative natural property view of desiderative unity. They are all consequences of how a primitively normative natural desiderative unity forces a critical lack of information about what makes norms the way they are. This lack of information appears insoluble. Any room for an extension of Smith's anti-Humean rationalism into an account of the epistemology of ethics looks like it will be made intractable by allowing a primitively normative natural desiderative unity. On a related front concept acquisition of a primitively normative desiderative unity when that primitive is not essentially causally locatable in the dynamics of desiderative and doxastic psychology makes explaining how we acquire and use that concept inexplicable. There is a plausible tendency to think that our concepts of psychological states causal concepts. Smith's anti-Humean rationalism can

both accept and even use this, so long as the normative component of psychology does not generate necessary connections between beliefs and desires. But the problem is that primitively normative natural desiderative unity seems to block any further theory of it linking desiderative unity into a causally characterised psychology. It looks as if a primitive desiderative normative property should be causally inert or at least opaque to belief, or the Humean restriction that Smith endorses will be violated. Smith shares the assumption that much if not all of psychology is essentially causal in nature when he discusses the squaring of the narrow reduction of colour with a broader physicalism.[103]

The problem this poses for Smith is that epistemic mysteriousness is Smith's main objection to non-naturalism. So even if you allow a primitively normative natural property to play the role of a naturalistic vindication of the narrow reduction of right to natural properties the resulting victory for meta-ethical naturalism is pyrrhic at best.

We can conclude then that Smith's squaring argument fails. He needs to stipulate that the possibility of a primitively normative natural desiderative unity is ruled out and that implies that being a property of a psychology is not enough to secure the squaring of the reduction of rationality and so of right with a broader naturalism. The required stipulation to block primitively normative natural properties is a requirement that desiderative unity must be realised by properties of the states of psychology that can be described in

[103] Psychological states as such are essentially causal – or at least so Smith allows in his discussion of the colour case. Desiderative unity is a property of the relations between the contents of a class of psychological states. But this does not entail that these properties are themselves essentially causal. However I am not requiring that in my argument. I am requiring rather that primitive desiderative unity be able to be conceptually linked to accounts of psychological change that are themselves essentially causal and it is this sort of connection that primitively normative natural desiderative unity appears to block.

non-normative terms. This renders Smith's naturalism explicitly reductivist in just the way that Jackson's descriptivism is reductivist. We can now see that the observation that desiderative unity will, by definition, be a property of psychological states is no wise sufficient to even give a reason to expect a squaring argument about desiderative unity can be made let alone count as having made one.

Section 4.2 Squaring in the colour case

Smith's naturalism requires a squaring argument, and we have seen that the one he has provided fails. I will now examine how the colour case squares the narrow reduction of colour to surface reflectance properties with a broader physicalism in the hope that we might find something useful for Smith's reductive naturalism in the ethics case by doing so.

Squaring the reduction of colours using a dispositional analysis of colour terms is done using a background theory of mental states and involves an explicit account of how the reduction of mental states can be secured without an analysis of 'colour experiences'. The squaring argument in the colour case explained how the circular definition of colours could be useful, innocuous, and consistent with a broader naturalism. The colour case used an explicitly stated background theory of our concepts of mental states to secure these results. The disanalogy between the colour case and the ethics case then is not just that the substantive reductive premise in the argument for the narrow reduction right to natural properties of acts is a priori. Rather it is that there is a comprehensive lack to the right kinds of background theories of desiderative unity.

4.2.1 The colour case and squaring a narrow reduction of colour with a broader physicalism

The narrow reduction of colours to physical properties of objects is given in the following argument according to Smith

"Conceptual claim:	the property of being red *is* the property that causes objects to look red to normal perceivers under standard conditions
Substantive claim:	the property that causes objects to look red to normal perceives under standard conditions *is* surface reflectance property α
Conclusion:	the property of being red *is* surface reflectance property α"

Smith adds that

> "To have good reason to believe the premises of this two-stage argument we have to draw upon our prior understanding of the concept of being red, our prior beliefs about which objects would look red to normal per perceivers under standard conditions. But, of course, that is neither here nor there given that our epistemic situation is one in which we do have such prior knowledge, and given that our interest in putting forward such an argument is squaring colour talk with physical talk." (Smith, TMP, p 53.).

Of the two premise reductive argument for a summary style network analysis and reduction of colour (in objects) terms Smith says in TMP, chapter 2, footnote 9 the following

> "It might be thought that we don't yet have an argument that would allow us to square colour talk with a broader physicalism *per se*, as the argument just given has no bearing on whether a subject's experience of having something look red to her is itself a physical state. But the foregoing discussion suggests an obvious strategy for squaring talk of colour experience with physical talk as well. The first step would be to construct an analysis of our concept of a colour experience. The second stage would be to show how these analyses allow us to identify colour experiences with, say, states of the brain. If, as seems plausible, our concept of colour experience is the concept of a state of a subject that, in conjunctions with a relevant desire, causally explains our bodily movements – for example, our picking out red objects from objects of other colours – then it should be clear enough how the attempt at vindication would go, and why it should be deemed likely to be successful (compare Lewis, 1972)." (TMP, pp. 205-6)

'Psychophysical and Theoretical Identifications' (Lewis, 1972) assumes a causal theory of mind while allowing an uneliminated circularity to guide the reduction of colours to physical properties and mental states to neurological states. But this assumption is controversial in the case of sensations since a non-causal notion of qualia competes for the role of constituting sensations. And it seems like sensations are at the heart of perceptual experiences. Lewis, Smith, and Jackson all allow that the issues raised by qualia for physicalism (causally conceived) can be solved – though only controversially.

What is important to notice here is that squaring colour talk with a broader physicalism is not simple or straightforward even if it is very plausible that there is a plenitude of resources with which to do the job.

Lewis (1972) gives a theory of theoretical term introduction in sciences that requires theoretical terms be given functional/causal roles that have to be uniquely realised for the theoretical terms to refer. Theoretical terms (T-terms) are distinguished from pre-existing old terms (O-terms) that the theoretical terms are defined in relation to. They are given causal roles couched in O terms and a Ramsey sentence with a unique realiser clause is provided to generate the reductive identifications that allow t-terms to be dispensed with. Theoretical terms are identified with the unique realiser of their causal roles and, according to Lewis, they must be so identified as a matter of deductive inference. The arguments for this claim are complex but we will assume with Smith that they are successful and useful in understanding the case of the reduction of colours.

Lewis (1972) supposes that we have an independent well-established theory of neurology that provides states with the causal roles sufficient to uniquely realise a folk theory of mental states. Though supposing mental states, and particularly colour attributions and colour sensation states in particular, are theoretical postulates is a myth (that is it is false) he thinks it is a good myth. A good myth is where what our mental state names actually mean is what they would have meant if the myth were true. This is a potted argument for

treating mental state terms as theoretical terms, even though they do not naturally occur as such.[104] Lewis adds in footnote 15, p. 257:

> "Two myths which cannot both be true together can nevertheless both be good together. Part of my myth says that names of color-sensations were T-terms, introduced using names of colors as O-terms. If this is a good myth, we should be able to define 'sensation of red' roughly as 'that state apt for being brought about by the presence of something red (before one's open eyes, in good light, etc.)'. A second myth says that names of colors were T-terms introduced using names of color-sensations as O-terms. If this second myth is good, we should be able to define 'red' roughly as 'that property of things apt for bringing about the sensation of red'. The two myths could not both be true, for which came first: names of color-sensations or of colors ? But they could both be good. We could have a circle in which colors are correctly defined in terms of sensations and sensations are correctly defined in terms of colors. We could not discover the meanings both of names of colors and of names of color-sensations just by looking at the circle of correct definitions, but so what?" (Lewis, 1972, pg 257)

We should understand this footnote in the context of Lewis' functional definition of mental states and the hypothesis that this entails the identification of mental states with unique realisers. He says

> "If the names of mental states are like theoretical terms, they name nothing unless the theory (the cluster of platitudes) is more or less

[104] Of course this is arguable but I am not going to argue for it. It is sufficient to observe that Smith explicitly refers to this work and Jackson's framework is at least compatible with much or even all of it.

> true. Hence it is analytic that either pain, etc., do not exist or most of our platitudes about them are true. If this seems analytic to you, you should accept the myth, and be prepared for psychophysical identifications." (Lewis, 1972, p. 257)

The cluster of platitudes provide a network of causal interdefinition and the idea that they are able to be understood as theoretical terms combined with Lewis's theory of theoretical term introduction means that if mental states are realised at all then, for the most part, they are all realised. And the consideration of the question of the realisation of mental state terms does not require that any single term in the network be defined and reduced successfully alone. The circle of definition of colour and colour-sensations, in this context capture a fragment of the broader causal theory of mind upon which the definitions depend for the supply of referents, and so ultimately depend upon for the supply of their 'meanings'.

So the idea that the component of a broader theory of mental states can independently give sufficient grounds to reduce that component to a realiser property, that component being the one that concerns colours in objects and the colour-sensations (Smith calls them colour experiences but both are the mental states necessarily associated with colours in objects), just misunderstands how interdefined theoretical terms sink or swim together.

We should notice that Lewis in footnote 15 above only rules out knowing the meanings of both colour terms and mental state terms from the correct circular definitions

> "We could not discover the meanings both of names of colors and of names of color-sensations just by looking at the circle of correct definitions …" (Lewis, 1972, p 257)

This fits with the narrow reduction a priori premise Smith gives above failing to analyse the mental state terms for colour experiences. So effectively we have the following. You can define colours causally relative to mental states, without offering a definition of mental states. This allows you to putatively entertain a contingent identification of colours with surface reflectance properties. But that identification will only actually come off with colours **if** you have a successful reductive theory of mental states. That theory is provided, according to Lewis, if you have a statement of all of the relevant causal platitudes about mental states and the causal theory of mental states is a complete theory of mental states. Additionally it must be the case that the terms that name these mental states in this causal theory of mind can be successfully subjected to manipulation into a Ramsey sentence with a uniqueness requirement for the realisation of mental states. Finally you find an independently identifiable set of phenomena with the right kind of causal powers to count as realisers for the mental state terms.

In Smith's own footnote about squaring the narrow reduction with a broader physicalism it is clear that he is using the same assumption of a causal theory of mind as Lewis. But Smith leaves the notion of colour in objects unreduced and unanalysed in his discussion of an essentially causal theory of colour experiences, and in fact uses them to talk about the successful causal characterisation of colour experiences. But this does not fit with understanding the a priori premise in the narrow reduction as supplying the meaning of the colour term in that premise. So even though we have a kind

of 'definition' we also know, according the theory in Lewis 1972 that it is not meaning-giving. What meaning is in this context is hard to work out but for Lewis the meaning postulates[105] of a group of theoretical terms that can be Ramsified appear to be the logical specifications that require a set of theoretical terms be uniquely realised and that if they are they are identified with what realises them and if they are not uniquely realised then they are all denotationless. So what is of import is that your theory is uniquely realised and where all your T-terms can be explicitly be defined with O-terms, at least taken as a network. But for mental state terms, including the colour experience terms, this can only be done for all the mental state terms taken together and has to be such that they can be eliminated in favour of causal profiles linking mental states to each other causally and to their causes and to what they cause. It is not clear why we can deem this likely to be successful, given Lewis 1972 as Smith claims.

Another important thing to notice is that the capacity for this approach to avoid a permutation problem depends on choosing one or other of a colour or colour sensation pair to simply be stipulated as whatever it is qua colour. That is, if you wish to show how colours like red are reductively identified with a physical property and you only have the dispositional analysis and Lewis: 1972 in play, then you will only avoid a permutation problem for this strategy by assuming that red experiences are just that and yellow experiences are just that and so on. You solve a permutation problem for colour reductions by stipulating the differences in colour experience. Lewis 1972 argues that having to make stipulations like this is not sufficient reason to block psychophysical reductive identifications. But it does mean you don't have an explicit explanation of how the permutation problem is avoided.

[105] Lewis, 1972, pg.254

4.2.2 Disanalogies between the colour case and the ethics case

There are two immediate differences to note.

As we have already seen in Smith's opinion the ethics case uses an identification of those properties of acts that a fully rational version of the agent acting would desire the act to have given the circumstances. This sits in the role of the substantive claim and unlike the one made for the narrow reduction of colours (which is a posteriori) the ethics substantive claim is a priori. This, I have argued, requires that Smith suppose the particular determination of desires by desiderative unity maximisation in fully rational versions of agents is known a priori.

The colour case has a squaring argument that is characteristically different form the squaring argument Smith offered for the ethics case. In fact I think unless something is changed in Smith's theory the squaring argument is characteristically different from anything Smith can offer for the ethics case.

The Lewis 1972 reduction of mental state terms to physical state realisers, despite in the case of colour terms and colour experience/sensation terms there being a circular interdefinition that is not discharged, only works at all because both generally for theoretical terms and specifically for mental state terms Lewis assumes the explicit theory implied by our mental state concepts is essentially causal. That is mental states are defined mutually in terms of the functional and causal interdependence, and by what things typically cause them and by what things, collectively, they typically cause. Notice this

is hardly an obviously successful strategy. But it is a viable option supported by philosophical theories about both term reduction and mental states and bolstered by a well-evidenced candidate realiser phenomenon given by the neurological sciences. Even if we left Smith's arguably inadequate squaring argument in play it does not have either feature. The properties that constitute desiderative unity, maximised or not, are not essentially causal. And even if rationality has a necessarily causal but defeasible component regarding the alteration of sets of desires in the direction of desiderative unity maximisation this does not really impact in a relevant way. The essential truth maker for Smith anti-Humean rationalism is a property that is **not** essentially causal as the theory stands. There is no obviously functional theory of desiderative unity on offer either. This is because though we know that the ethical concepts require convergence on a subset of desires in the population of fully rational agents we don't know anything else about desiderative unity or the desires of the fully rational.

The point is not that we lack facts to secure a reduction – that much is true of the physicalism about mental states outlined in Lewis 1972. Rather, it is that there is no theory about what sorts of facts we are to look for when securing a realiser for desiderative unity. With the colour terms we know that the mental states of colour experiences will be provided by a wholesale theory of mental states and that this theory is essentially causal. We know that treating the content of our mental state concepts can, according to Lewis's proposal be usefully treated as a sort of folk scientific theory introducing scientific terms, that these terms work like names, that this entails that all the terms of an interdefined network of theoretical terms (and so for this approach to mental state terms) have to be uniquely realised all together or all fail to be realised at all. We also know, given the Lewis proposal, that a successful realisation

of mental state terms treated as theoretical terms will result, necessarily, in an identity relation between mental states and their physical realisers.

Lewis adds things to this bundle that, coincidentally, serve to show the advantage that squaring of colour talk with a broader physicalism has over the same effort in the ethics case. Lewis points out that the plausibility of behaviourism is at least explicable if the mental state concepts are essentially causally defined. And he offers an argument showing that his proposal is consistent with the infallibility of knowledge of mental states.[106] For our purposes here it is a matter of indifference how defensible Lewis's views about psychophysical and theoretical identifications are. What matters is that there are pre-existing, independently motivated, background theories that serve to explain both how it is that there is a circular definition of colour and colour-sensations and show how it could be that this circularity need not discharged and yet a broadly physicalist reduction still be effected.

There are no similar supporting background theories in the ethics case. This is not the claim that there are no platitudes of and theories about the distribution of ethical facts. Rather it is specific to the narrow reduction of right to properties of acts that fully rational versions of the actor would desire those acts have given the circumstance of the acts. This narrow reduction makes use of the notion of fully rational desires, which have as their primary determinative component desiderative unity. A squaring argument for Smith's anti-Humean rationalism would have to have an independently supported background theory of desiderative rationality that explained both how it is that we come to have a circular definition of 'right' and 'rationality'

[106] This is at least a component of what is persuasive about the ideas surrounding qualia and their putative irreducibility. So Lewis's argument shows that his treatment of mental state terms is at least partly compatible with the concepts associated with qualitative experience.

and how it is that this circularity need not be discharged an yet a broadly naturalist reduction be still effected.[107] But desiderative unity is not defined and is proposed as Smith's particular contribution to explicating desiderative rationality. And though moral platitudes (like the objectivity platitudes and the practicality platitudes as Smith calls them) constrain desiderative unity they do not provide the background theory of either desiderative rationality or of desiderative unity itself that would be required if reductive squaring of ethics was relevantly like the reductive squaring of colours.

In 'Naming the Colours' (Lewis 1997), Lewis considers the issue of squaring colour talk with a broader physicalism in more detail. Unfortunately the examination suggests that the colour case and the ethics case are more disanalogous rather than less. Lewis deploys more background theories to further defend a reduction of colours to physical properties – this time concerning the nature of conventions and public knowledge. They are used to solve a variant of the underdetermination of reference for colour names even given a reductive realisation them physically. The takeaway message for us is that the physical reduction effected in Lewis: 1972 stipulated a solution to the permutation problem for colours and Lewis: 1997 is intended to make progress on a way to remove that stipulation.

Lewis argues that the indeterminacy of reference in the reduction of colours is severe, sufficiently so to render the circular definition effectively useless or not definition at all (Lewis, 1997, p335). He supposes that the only solution is

[107] This is an explicitly reductive naturalism of the sort argued for in section 4.1. Desiderative unity would have to be reductively identified with properties of the psychology of rational agents, or with something that had the right kind of impact on properties of the psychology of rational agents, that are not essentially characterised or defined normatively. This seems like a requirement in fact if the approach in Lewis's1972 really is going to inform the reduction ethical properties.

that the folk theory of colours has to have components distinguishing colours one from the other, and it is because the fragment of folk theory of colours and mental states considered when looking at colour and colour-sensation pairs is only a fragment that this problem arises. Lewis's proposed solution is to use particular and contingent identifications of colours with colours of salient instances (blue as the colour of the sky, say) to supply the differentiating information. The solution has to be refined in the face of the fact that different groups will use different objects to play the role of reference fixing paradigms of the colours. Lewis suppose that this can be solved if, as it happens, disparate colour definition communities agreed about which physical objects bear particular colour properties, without this agreement needing to be common knowledge.[108]

This widens the gap between colours and ethics. The plausibility of Lewis's proposal is in part dependent on a well-supported assumption that we in fact do have a common ground colour language somehow or other. Lewis stipulates that materialism is a non-negotiable component of an adequate theory of colour and that it must be commonsensical. The commonsensical component can be altered somewhat but not too much. Though it is not clear which parts of our commonsensical folk theory of colours is non-negotiable it is clear that some parts are and so Lewis concludes that

> "In other words, it is a Moorean fact that the folk psychophysics of colour is close to true."

[108] Lewis configures the issue of colour reference given rigidification on local examples of colours as a matter semantic coordintation between different colour language groups using different local colour examples. The merits of this proposal are beyond the scope of this thesis.

And

> "Yet it is a Moorean fact that there are colours rightly so-called. Deny it, and the most credible explanation of your denial is that you are in the grip of some philosophical (or scientific) error." (Lewis, 1997, p. 325)

So Lewis can make the following claim, given that the assumed fact that there are colours rightly so-called is a Moorean fact. If you assume materialism (as he does) then somehow colours unambiguously play a causal role in our psychological economy and we successfully communicate about this fact. Within these constraints Lewis has a theory that relies on using local examples of colours in reference fixing roles – there are two potential prices one of which can be paid that other of which must be avoided. The acceptable price is that it makes "unrigidified things harder to say" (Lewis, 1997, p. 341). The unacceptable potential price would be loss of co-reference to the same colours and successful communication between different communities of colour talk. These different communities are differentiated by which local particulars are given the role of differentiating colours from each other. Lewis's proposal solves both issues simultaneously requiring that colour talk be **taken** to refer rigidly and supposing plausibly that these colour definition communities agree, as it happens, on which things are coloured and which colours they are.[109]

[109] Lewis states that the intension of co-referring rigid terms across colour language groups is unsettled. They have more in common than the mere extensions of their colour terms but whether this is intension or not is unsettled. If nothing else this gives reason to hesitate interpreting the rigidification of reference involved here as indicative of colour terms being natural kind terms with a common A-intension.

I suppose that the Moorean fact feature of Lewis's materialism is to some extent shared by Smith since, as we saw above Smith claims

> "To have good reason to believe the premises of this two-stage argument we have to draw upon our prior understanding of the concept of being red, our prior beliefs about which objects would look red to normal per perceivers under standard conditions. But, of course, that is neither here nor there given that our epistemic situation is one in which we do have such prior knowledge." (TMP, p. 53)

To claim we have enough prior knowledge of colours to support the plausibility of the narrow reduction of colours to surface reflectance properties looks like it is the kind of claim we should believe if the status of folk psychophysics is as Lewis's describes it – more likely to be mostly true than any theory that claims otherwise.

This just is not the case for ethics. Smith provides a good argument for optimism about the truth of ethical claims in chapter 6 of TMP. But that is a far cry from the folk theory of ethics being more likely to be true than any theory that contradicts it.

And again the latter Lewis shows that the squaring of colour talk with a broader physicalism is far from simple or obvious and the actual squaring argument in the colour case depends on marshalling many philosophical and scientific theories that are prima facie relevant. No such pool of resources exists for the ethics case.

Section 4.3 Why Smith needs an explicit theory of desiderative unity – the upshot of section 4.1 and 4.2

Failure to provide a squaring argument is a failure to vindicate reduction of right to natural properties in the context of a broader naturalism. This is a standard for naturalism that Smith accepts.

The squaring argument in the colour case explained how the circular definition of colours could be useful, innocuous, and consistent with a broader physicalism. The colour case used an explicitly stated theory background theory of mind to secure these results. In fact it uses several sophisticated philosophical theories and at least one scientific theory and relies essentially on the controversial claim that our concepts of mental states are causal concepts (or near enough). Though Smith uses Lewis (1972) 'Psychophysical and Theoretical Identifications' it is interesting to note that Lewis (1997) 'Naming the Colours' makes matters worse rather than better when it comes to the disanalogies between squaring narrow reduction of colour with a broader physicalism and squaring the narrow reduction of right with a broader naturalism. The Lewis (1997) makes explicit an assumption about the colour case that is part of Smith's argument TMP. That is just that it is a Moorean fact that folk psychophysics is close to true. Some thing like this claim seems required to support Smith's assertion that we have the relevant priori knowledge required to accept the narrow reduction of colour argument he presents.

There is no analogy argument from the squaring of colours with a broader physicalism to a squaring argument of right with a broader naturalism. This is because there are no relevant theories of desiderative unity or desiderative

rationality **at all**. It is not that there are no theories playing the same roles as the ones used in the squaring of the colours case. Rather it is that there are none at all on offer. This is a significant disanalogy.

The squaring argument Smith offered for the ethics case in TMP did not explain why the circular definition of right could be useful or innocuous. In fact it failed to be a squaring argument at all. As it stands arguing that desiderative unity is a natural property because it is a property of a psychology and so by definition natural (or near enough) permits a primitively normative natural desiderative unity. This possibility is as bad as non-naturalism and would render moral epistemology insolubly mysterious. Explicitly excluding this possibility shows that to square the narrow reduction of right Smith needs an explicitly reductive theory of desiderative unity. Smith does not have such a theory.

Smith needs at least the reasonable promise of a reductive theory of desiderative unity. There are no useful analogies between the colour case and the ethics case, given Smith's descriptions of them, that should persuade us that there is some possibility of an explicit theory of desiderative unity. Without an explicit theory of desiderative unity Smith's meta-ethics must collapse into definitional naturalism and looks likely to fail as a naturalist theory all together. This is because either it remains definitional naturalist and adopts the primitively normative natural property position which is as bad as non-naturalism or, if it rejects primitively normative natural properties, Smith's theory has to accept that to pass the squaring test for naturalism an explicit and reductive theory of desiderative unity has to be provided and that there is no such theory provided. So either Smith's theory collapses into definitional naturalism and then from there into a primitively

normative naturalism (as bad as non-naturalism) or acknowledges that it has to provide an explicit theory of desiderative unity to pass the squaring requirement for naturalism and, in the absence of such a theory, accept that it fails as a naturalist theory.[110]

These considerations strengthen the first horn of the dilemma facing Smith from the claim that his account faces collapse into a definitional network analysis and reduction for ethical terms to the claim that his account not only faces collapse into a definitional network analysis **but also** looks likely to fail the demand for a squaring argument and so looks likely to fail as a naturalist theory outright.

Section 4.4 Altering Smith's theory to avoid the first horn of the dilemma leads to the second horn

Section 4.1 through 4.3 has shown that Smith needs to solve two problems. He needs a new reason for keeping the semantic gap and he needs a squaring argument or the reasonable promise of one.

A consequence of the argument against primitively normative but natural desiderative unity is that an explicit and explicitly reductive theory of desiderative unity or something very much like it is required to satisfy the squaring requirement on the narrow reduction of right to the natural properties of acts that fully rational versions of actors would desire those acts

have given their circumstances. We don't have such a theory ready to hand. It might seem unreasonable to require one given that ethical theory could be argued to be a work in progress. However I think that it is reasonable to at least have some idea as to how such a squaring argument is to be provided. Smith seems to agree, in the case of colour explicitly so. Lewis in Psychophysical and theoretical identifications notably does the latter not the former. He spells out how am essentially causal theory of mind could be treated as a term introducing scientific theory and how unique realisation of the new terms by way of what the old or other terms refer too is a requirement. Coincidentally the circular definition of colour and colour-sensations won't block this reductive identification of mental states with neurological states. Lewis's use of background theories secures the plausibility of the possibility of a reduction of mental states to physical states, which would in turn allow the narrow reduction of colours use of colour-sensation terms to be rendered harmless – also clearly only a partial account of how the permutation problem is solved.

Smith has two arguments squaring the narrow reduction of right with a broader naturalism. The first is the explicitly stated argument using the status of the normative components of desiderative rationality as psychological properties. This argument fails. The second is the analogy between the colour case and the ethics case. If the analogy held we might accept that some reductive theory of normativity squaring the narrow reduction of right could be found. But this analogy argument also fails. Finally, given that Smith supposes that all the relevant information is a priori knowable, it is unclear that we should accept the promise that a squaring theory of desiderative unity will be made explicit at some future date. Why can we not provide an a priori theory of desiderative unity if we can know

particular consequences of its maximisation on the desires of the fully rational?

This section will consider how treating desiderative unity like a natural kind term secures both requirements.

4.4.1 Desiderative unity as a natural kind term

Smith could treat desiderative unity as or as if it were a natural kind term.[111] To do this he would have to have something like a reference fixing description of desiderative unity and he would have to give actual desiderative an role in his anti-Humean rationalist meta-ethical theory.

I have argued already that there are several advantages to distinguishing actual desiderative unity from the kind of desiderative unity that would secure the substantive truth of Smith's rationalist theory. A reference fixing notion of desiderative unity allows us to satisfy the requirement of appropriately valuing moral goods. Treating 'desiderative unity' as a natural kind term would involve our reference fixing on actual desiderative unity rather than on the ideally convergent desiderative unity that morality requires to be substantively true. The relationship between an ideally convergent desiderative rationality and normative reasons can be captured by the right kind of reference fixing description.

[111] The 'as or as if' qualification is here to mark a disinterest in the metaphysics of natural kind properties, particularly the ideas around perfectly natural kind properties.

In this proposal the features of full rationality that suffice to secure objective normative and moral reasons can be incorporated into the reference fixing description for actual desiderative unity and thereby incorporate the conditions for the substantive truth of Smith's anti-Humean rationalism. The list of things we know about desiderative unity if it exists is as follows:

> Desiderative unity is a property or cluster properties of the relations between the contents of desires in a set of desires.
>
> It is a property or cluster of properties that admits of degree (there is an ordering)
>
> At least a partial ordering of desire sets relative to increases in desiderative unity is possible.
>
> There are idea sets of desires that maximise desiderative unity, and these are the desire sets of fully rational idealisations using actual agents as a starting point.
>
> Increases of desiderative unity can actually be detected between sets of desires at least for some sets of desires.
>
> An actual or ideal capacity for desiderative unity detection is sufficient to track changes in the unity of sets of desires up to those sets where unity is maximised.

Desiderative unity maximisation will make it the case that fully rational idealisations of actual people will overlap relative to the acts agents have normative reason to do given circumstances.

Desiderative unity and the rational preference for its increase covary.

The ability to modify desire sets towards desiderative unity increase and the ability to detect desiderative unity increase do not *necessarily* covary, except in fully rational idealisations.

This list is not meant to be comprehensive or sufficient. It is however a kind of minimum job description that we can work Smith's anti-Humean rationalism gives for desiderative unity. This is a good start on a reference fixing description for desiderative unity.

4.4.2 Two immediate benefits to treating ethical kinds as natural kinds

Treating desiderative unity as a natural kind does not give a background theory of desiderative unity. However, it achieves two things.

First benefit:

If desiderative unity is a natural kind term then it involves the a posteriori necessary identification of desiderative unity with whatever properties actually realise it. This feature is enough to block a definitional reduction of the network of terms involving the interdefinition of 'right' with 'rational' and 'rational' with 'desiderative unity'. The reduction is essentially a

posteriori and so no a priori definition can supply the relevant facts for its realisation. This feature, should we adopt a natural kinds treatment of desiderative unity, is consistent with the narrow reduction of right to the natural properties of acts being a matter of predominantly a priori investigation.

It is worth noting that Lewis's argument in the 'Psychophysical and Theoretical Identifications' (Lewis, 1972) is at least consistent with the contingent realisation of mental states and the rigidification of the reference of colour terms. So it would be consistent with the a posteriori necessary identification of colours. But this is not precisely required by Lewis (1972). In 'Naming the Colours' (1997) Lewis makes an explicit commitment to rigidifying on clusters of local particular instances of colours and relies on colour definitions that differ in terms of the particular instances of colours they use, as it happens, picking out the same colours to solve communication problems this rigidification brings about. But again this is not exactly the same as treating colour terms as natural kind terms and Lewis is quite clear that his theory does not require any view of the intentional overlap of colour terms between different definitional communities. The upshot for my purposes here is that though the colour case squaring argument provides an explanation of the innocuousness and usefulness of the circular interdefinition colours and colour-sensations it is not necessarily by way of treating colour terms as natural kind terms.

But then my goal here is not to make ethical properties more like colour properties. Rather it is to show how you might, in the absence of a permutation argument, find a reason to keep the semantic gap between the analysis and reduction of right and the analysis and reduction of rationality.

And this proposal does just that. For though what we know about right and rational and indeed the desires of rational agents is a priori in a sense. And we also know that the final step in the reduction of desiderative unity will not be a priori and so at least the ultimate squaring of the narrow reduction of right with a broader naturalism will not involve definitional network reductions.

Second benefit:

The second thing we achieve is that we have an explanation of both where we should look for a background theory of desiderative unity and why failure to have one ready to hand is not pathological for Smith's anti-Humean rationalist ethics. We have a reason for accepting the promise that the narrow reduction of right to the natural properties of the acts of agents relative to their circumstances could ultimately be squared with a broader reductive[112] naturalism.

The explanation is just that having a ready, even a priori grounded, grip on a natural kind does not entail knowledge of the nature of that natural kind. This is really a feature of natural kind terms – they act as useful topic coordinating placeholders while we await the results of further investigation.[113]

---

[112] The possibility of primitive normative natural properties invites this disambiguation. The naturalism I have in mind here is consequently much like Jackson's descriptivism at least with respect to the following – reduction is only vindicated of it realizes normative terms with non-normative properties.

[113] This is not a claim about the actual semantics and practice explicitly known about prior to theorising in ethics or morality. Rather this is a claim about a good or useful way to understand the function of a component of a theory of desiderative rationality and its role in a supposedly folk theory of ethics.

4.5 Gains and Costs: The Second horn of the dilemma

Gains:

So we have solutions to both of Smith's problems. We have found a reason to keep the semantic gap between 'right' and 'rational' in place. Because a natural kinds treatment of desiderative unity would involve the necessary a posteriori identification of desiderative unity with whatever realises it actually the reduction of right to a natural kind cannot ultimately be found by way of a definitional network analysis and reduction.

And though we don't yet have an explicit theory of desiderative unity, we have a reasonable account both of why we don't – we have a pre-reductive natural kind concept in play – and an account of where we should look to find the relevant theory (at least initially we would look for instances of prosperities in actual psychology that satisfy, or near enough, our reference fixing description of desiderative unity). Nothing in this prevents coming at the problem from the point of view of constructing more detailed theories of full rationality. Rather it sets a kind of empirical benchmark for these theories. They are only relevant to the question of desiderative rationality and morality so far as they bear on finding and evaluating properties that collectively satisfy the requirements on our list above – that satisfy our reference fixing description of desiderative unity.

Costs:

Reference fixing on actual desiderative unity will make evaluating the normative structure in other psychologies difficult. This is just because we are here entertaining restricting the reference of the term 'rational' explicitly to features of actual agents. Should actual agents fail to embody a desiderative unity of the converging type, for example, then there just is no desiderative rationality and there just is no morality. Trying to talk about the possibility of some creature similar to us but constructed as we hoped we had been might be to talk about how to be a moral creature. Or it might not. On a like note, if we happen to have the right kind of psychology to underpin a rationalist morality as Smith envisages it and we adopt the proposal I have made then we might find that pan-Human desiderative rationality does not extend to other kinds of agent. This might be a problem. But I think that these problems will only exist relative to some particular theory of desiderative unity and it is not clear whether the problems are there for all possible variants on an explicit theory of desiderative unity or whether they will remain unsolved by the explicit theory of the referent of desiderative unity.

4.5.1 The second horn of the dilemma

Much more pressing and determinate a cost is that the solution I propose constitutes the second horn of the dilemma for Smith. This is because the proposal is a brand of metaphysical-but-not-definitional naturalism. And Smith argues that metaphysical-but-not-definitional naturalism must be wrong because it entails the possibility relativism when it should not. As it stands adopting the proposal would count as an unacceptable modification of Smith's anti-Humean rationalism.

This, then, is the dilemma that Smith faces:

Either Smith accepts that:

1 Given the removal of the permutation problem, and the failure of his own sketches of a Rawlsian account of desiderative unity and a the argument that being a psychological property is not enough to square desiderative unity with a broader naturalism, his theory collapses into definitional network analysis naturalism (of which Jackson's meta-ethical moral functionalism is an example, or near enough for our purposes) and then either becomes a non-reductive naturalism as bad as non-naturalism or fails outright.

Or

2 He adopts a natural kinds treatment of desiderative unity, which modifies his theory in a way that gives a new reason to keep the semantic gap between the analysis and reduction of right and rational, and gives a reasonable promise of a squaring theory for the narrow reduction of right. This comes at the cost, according to Smith, of being too permissive of relativism and so being false.

4.6 Next

Accepting either the collapse to definitional network analysis naturalism and likely failure of his theory as a naturalist theory or treating desiderative unity as a natural kind and thereby rendering his theory inappropriately relativistic looks bad either way for Smith. The next chapter is devoted to showing that

Smith's objection to metaphysical-but-not-definitional naturalism is not compelling. A consequence is that we might perhaps be able to blunt the second horn of the dilemma, suggesting a way out for Smith's anti-Humean rationalist meta-ethical theory.

Chapter 5 Evaluating Smith's objections to natural kinds treatments of ethical properties

The last chapter concluded that Smith faced the choice between a collapse of his meta-ethical views into definitional naturalism (and the likelihood that his anti-Humean ethics failed to be reductively naturalistic) and adopting a natural kinds treatment of desiderative unity. Given that desiderative unity is the central normative component of his theory, the second choice is tantamount to adopting a natural kinds treatment of ethical terms. But Smith rejects this kind of approach, which he calls metaphysical-but-not-definitional naturalism[114].

In this chapter I will argue that Smith's objections metaphysical naturalism do not succeed and then argue that his rationalist ethics can nonetheless be rendered compatible with metaphysical naturalism, albeit at a cost he might refuse to meet.

Smith presents two arguments against metaphysical naturalism. The first is Richard Hare's cannibals and missionaries case (Hare 1952: pp. 148ff.). The second is an analogy argument from the natural kind term 'water' to a natural kinds treatment of 'right'. The conclusion of both arguments is that a natural kinds treatment of ethical terms will allow the possibility of

[114] In this chapter I am continuing to follow my usual convention of shortening 'metaphysical-but-not-definitional naturalism in the ethics case' to 'metaphysical naturalism'.

relativism when folk morality tells us it should not. Smith takes this to be a good reason to reject a natural kinds treatment of ethical terms.

I will argue that Smith's use of the Hare case perpetuates an error Hare is making about what a natural kind treatment of ethical terms would look like. Hare's error is to use a radically underdescribed reference fixing description for 'good'. Smith's second argument claims that in the water case we would be indifferent to an occurrence of different referents for the terms that play the water role between different, and presumably isolated, communities. He then infers that the term 'right' if treated like a natural kind term will share this feature of the term 'water'. My reply to this is two fold: on the one hand it is not at all clear that this is a true claim about 'water', and on the other, even if it were true of 'water' it is not clear that a natural kinds treatment needs to agree that it is true of 'right'.

5.1 Reply to Smith's first argument

I dispose of some preliminary considerations concerning natural kinds terms and then turn to Smith's argument about cannibals and missionaries.

5.1.1 Natural kind terms

My arguments in this chapter concern the role a reference fixing description plays in how a natural kind term functions. I will be arguing that a folk theory of the natural kind, prior to the empirical investigation establishing the nature of the actual stuff being referred to, is what a reference fixing description is composed of. So though we can summarise the reference fixing description for the natural kind term 'water' with the phrase 'the watery stuff of our acquaintance' what we do when we cash out the meaning of this phrase by expanding on what it takes to count as an example of 'watery stuff', is to express both the various familiar ways in which water occurs in our environment and is thereby a salient presence for us and the pre-theoretic (usually pre-scientific) views we have of the character of this stuff. In this fashion a reference fixing description is like the theoretical term defining statements we examined in chapter 4. But there are significant differences. With a natural kind term like 'water' there is a more or less explicit lessening of the importance of the pre-scientific folk theory of water when treating 'water' as a natural kind term. Though we must use a folk theory to coordinate the specification of samples of the watery stuff , by focusing on the nature of the stuff rather than the properties we associate with it pre-scientifically we can allow that significant portions of the pre-scientific folk theory of water can be false. Though a degree of failure to satisfy the requirements of a term introducing theory might be permitted and even managed using causal clauses just like the ones relevant to 'water', too much change and a theoretical term stops referring. Natural kind terms are generally much more tolerant of the fact that the referents of the terms may

well fail to satisfy a pre-scientific view of the properties of the referents. This is the point of our guarded way of describing the reference-fixing mechanism of the term 'water'. The latter allows us to discount elements of pre-scientific folk theories while still managing to secure coordinated reference. When we say 'water is H_2O' we aim, in part to be showing that a scientific theory about H_2O really is a theory about water – something we always were referring to and forming views about, perhaps false views, prior to the H_2O theory came about. This is not novel but it bears repeating.[115]

But despite this difference we can't be cavalier about the importance of the role that reference fixing descriptions play in how natural kind terms work. It would be extraordinary to think, for example, that H_2O never played the roles of watery stuff in our lives and still deserved the name 'water'.[116] This is particularly important for the kinds of arguments Smith is using against metaphysical naturalism. This is because the cases he finds objectionable only arise when you have divergence across the same possible world in referents for natural kind terms that have the same reference fixing description. In the cases that Smith discusses what coordinates relevant similarity between the terms that refer differently is the sameness in the role the normative terms are playing in both the language and lives of disparate imagined communities. A reference fixing description approach adds nothing contentious to this and aims to express just these role facts. So Smith

[115] So though this story accords with the idea that what causes the use of a natural kind term is what that term refers to (TMP, p. 32) it also qualifies it. If it is too informal a way to understand natural kind terms, especially the ones that refer to perfectly natural kinds and the like, then we can qualify what I have said as being a ready way for a metaphysical naturalist in ethics to go about treating ethical terms *like* natural kind terms.

[116] This would be extraordinary to the point of incoherence, in fact. Since Smith focuses on non-counterfactual cases, so do we. What is not extraordinary (as we know from the work of Putnam and Kripke), is that in certain counterfactual cases H_2O has no watery properties; but that is not relevant here.

can't very well object to the requirement that normative terms treated as natural kind terms **must** share the same reference fixing description.

Though Smith states that a natural kind term refers to whatever natural kind it is that causes the use of a the term he also explicitly adds the idea that the role a term like 'water' or 'right' plays in our language and in our lives is relevant to the way the term refers. This I take to be sufficient evidence that Smith's view of natural kind terms is amenable to supposing that the a priori knowable information associated with a natural kind term is accommodated by an appropriately rich reference fixing description that alludes to the roles played by the intended referent and the fact that we are confronted by — acquainted with — this referent, whether it be stuff like water or (if 'right' is treated as a natural kind term) a property like right.

5.1.2 Moral relativism

A clarification will prove useful. Even for Smith there are acceptable relativisms of a sort. That is just the relativism of indexing 'right act' to the circumstances of the agent acting. But this is not moral relativism since, by hypothesis, every agent in like circumstances will find the same actions right. Likewise we don't generate moral relativism for wholesale circumstantial variations applying to whole populations should a variation arise. Moral relativism of the kind that conflicts with what Smith calls the objectivity platitudes of folk morality is difficult to comprehensively define. The idea is

roughly that there are conditions where competing inconsistent moral norms will generate different moral evaluations of the same circumstances and there is no fact of the matter as to which of them is right or wrong. 'Morality' becomes a term that can only be applied to things relative to adopting one or other of these norms, where it is incoherent to adopt both. This idea of 'moral relativism' is not consistent with our conception of the term 'right' in so far as we suppose it is objective. Moral objectivity is roughly the idea that, for example, one circumstance engenders only one 'right' act (or one of several equally 'right' actions that are) and if there is disagreement and everyone is using the same concept of 'right' (the objective one) then someone is making a mistake of a factual nature. As we will discover below this description of moral relativism does not cope with at least one condition that Smith's anti-Humean rationalism permits.[117]

For convenience we can use the following definition of relativism

> *"Metaethical Moral Relativism* (*MMR*). The truth or falsity of moral judgments, or their justification, is not absolute or universal, but is relative to the traditions, convictions, or practices of a group of persons." (Gowans 2015)

[117] This is the condition of a moral requirement to, in a given circumstance, have acts instantiate one or other of a disjunction of properties while simultaneously being morally indifferent over which disjunct is instantiated. The tricky part of this possibility that the disjunction could either ramify into more widespread divergences in what is morally required, for example, or could evince moral indifference across possibilities that are incompatible in some manner. None the less, this is not necessarily moral relativism.

This definition is useful, though we will refine it as we go. It covers all the cases we will discuss in this chapter and contains the nub of Smith's concerns about relativism, which are just that conditions or theories that fail to enforce a common truth maker for putatively the same or similar enough moral assertions and thereby permit moral relativism contradict the objectivity platitudes of folk morality. Contradicting the 'objectivity platitudes' Smith thinks leads immediately to error theory.

5.1.3 Smith's first argument against metaphysical naturalism

The first is found on pages 33-35 of "The Moral Problem" and the second is found in the footnote 7 at the end of the segment of argument on page 205. The first version of Smith's argument rests on an argument made by Hare in "The Language of Morals" pp. 146-9. That argument, using the cannibals versus the missionaries' case, tries to show how treating ethical terms like natural kind terms will lead to relativism. Smith finds this objectionable. What is objectionable can't be relativism or relativism in the face of our objectivity platitudes alone since error theory due to relativism is a possibility that Smith should not close off if his theory is to remain substantively defeasible[118]. What is objectionable is relativism for inappropriate reasons, in this case (as Smith thinks) because of the application of a false philosophical theory.

[118] Smith's position on the defeasibility of his conception of ethics has been discussed in chapter 3 of this thesis. See Smith's discussion of how his theory of normative reasons (of which moral reasons are a subset) is non-relative TMP pp. 164 -174 (particularly pp. 173-174).

I will argue that much of the plausibility of variable referents in the cannibals versus the missionaries' case turns on restricting the reference fixing description severely – much more severely than is typical with actual natural kind terms. In effect, Hare's case mistakes how natural kind terms work and reusing it the way Smith does simply perpetuates this error.

The objection considers a missionary on a cannibal island

> "The vocabulary of his [the missionaries] grammar book gives him the equivalent, in the cannibals' language, of the English word' good'. Let us suppose that, by a strange coincidence, the word is 'good'. And let us suppose, also that it really is the equivalent – that it is, as the Oxford English Dictionary puts it, 'the most general adjective of commendation'." (TMP p.33).

We should note that we are engaged in a discussion about what it would take to properly use and understand ethical terms. So where Hare claims that the missionaries' and cannibals' terms 'good' really are equivalent we should instead take the case, for our and Smith's purposes, to be one of really sharing the same reference fixing description – i.e. 'whatever it is that we are

acquainted with that plays the role of the most general adjective of commendation' or something very like this.[119]

Hare goes on

> "…If the missionary has mastered his vocabulary, he can, so long as he uses the word evaluatively and not descriptively, communicate with them about morals quite happily. They know that when he uses the word he is commending the person or object that he applies it to. The only thing they find odd is that he applies it to such unexpected people, people who are meek and gentle and do not collect large quantities of scalps; whereas they themselves are accustomed to commend people who are bold and burly and collect more scalps than the average." (TMP, p. 33).

Rather than continuing with the Hare position we can continue with Smith's view of it.

> "In our terms, Hare's argument can be put like this: if the cannibals use their words 'good' and 'right' to refer to the causes of their uses of the words 'good' and 'right', and the missionaries use their words 'good' and 'right' to refer to the causes of their uses of the world 'good' and

[119] Note that with natural kind terms that have different referents, like Hillary Putnam's 'water' and 'water*', what makes these words in any way relevantly similar is the fact that they share the same reference fixing description.

> 'right', *and if no more can be said about the content of their respective judgements*, then a radical relativism is on the horizon. For we seem to have good reason to suppose that the causes of the 'cannibals' and the 'missionaries' uses of the words 'good' and 'right' are very different from each other. And in that case we cannot suppose that the cannibals and the missionaries disagree with each other about what is *really* good and right." (TMP p. 34).

So we have relativism because there is no fact of the matter about who is right about what 'right' and 'good' refers to under the current proposal. Different contexts of use yield different referents for the terms when they shouldn't. On the face of it the missionary and the cannibal both should want to say the other one is wrong about what, and who, are 'good' and the objectivity platitudes indicate that we (the folk of 'folk morality') think someone has to be wrong in this case.

Smith now claims that the problem for metaphysical naturalism is that it forces us to suppose that the folk theory is wrong about this case. We have to be careful here, for Smith says

> "They [naturalists] can advance a form of metaphysical-but not-definitional naturalism. But even if they do it looks like they will run into trouble. For they must choose a description to fix the reference of the moral terms. And in doing this they must make sure that moral claims do not turn out to have different contents in different contexts.

> And yet this seems inevitable if they simply say that, for example, the word 'right' is used to refer to the features of acts that is causally responsible for our uses of the term 'right'. For if the cause of A's and B's uses of the word 'right' are not the same, then contrary to the platitude that if A says 'x is right' and B says 'x is not right' then A and B disagree, A and B are not disagreeing. A's judgement that x is right has a different content from B's judgement that x is right."
> (Underlining added. TMP, p. 35).[120]

If we look at the first underlined part of the passage above Smith appears to be saying that the problem with metaphysical naturalism is that it fails to block moral claims having different contents in different contexts. This leads to relativism and relativism violates the objectivity platitudes of morality. But Smith's objection to metaphysical naturalism can't just be that it allows a possibility where relativism is the case. Despite everything, error theory is a possibility. And as we saw in our discussion of Jackson's idea of mature folk morality in chapter 2 it is possible that implicit folk morality — the source of the objectivity platitudes Smith uses in his argument — turns out to be incoherent, where one possible way for implicit folk morality to be incoherent is for it to permit moral relativism by requiring, and failing to achieve, moral objectivity. The possibility of relativism, then, is not the problem.

---

[120] In this passage the need to avoid different referents in different contexts is mentioned. To clarify: what is meant is that because of the objectivity platitudes any ethical theory should prevent different referents for ethical terms in contexts that are not *relevantly* different. The culturally endorsed traditional cannibalism provides a different context of use for the term 'good' from that of the missionary but these differences are supposed to *not* be such that the referent of the term 'good', as we and perhaps the missionary use it, changes reference. (What is of course permitted are circumstance-relative differences in what is or isn't good.) The case is meant to highlight the objectionable nature of the position by having the term 'good' refer, at the cannibals context of use, to a property most of us would refer to as 'bad'.)

Remember too that Smith's own explication of the concepts of folk morality, his anti-Humean rationalism, permits the content of moral judgements to vary from circumstance to circumstance. Clearly in the above passage he must mean that metaphysical naturalism permits different properties to be the referents for the moral terms applied under the *same* 'circumstances'. As we will see when we consider the matter of metaphysical naturalism, assuming Smith's theory is a correct analysis of our concepts, the case of different referents for the same circumstance can be allowed without inducing non-objective relativism.

So the last formulation of the problem of moral relativism is not really precise enough. More pressingly for this formulation, it is hard to see how a circumstance can remain the same and a context vary, given Smith's anti-Humean rationalism. Since folk morality is a good candidate for the reference fixing description of a metaphysical naturalism for ethics and Smith's theory is putatively an explication of folk morality, then the Hare case only shows the flaws of using reference fixing definitions lifted from dictionaries.

The second underlined piece of the above passage is important for a different reason. The apparent relativism objection from Smith "seems inevitable" only **if** "they [metaphysical naturalists] say that,…, the word 'right' is used to refer to what is causally responsible for our uses of the term 'right'". The objection is that if the circumstances of use for apparently the same ethical term (same because of a shared reference fixing description) vary in the right

way then metaphysical-but-not-definitional naturalists of the kind Smith is imaging must insist that relativism is true and they must do so **because their semantic theory forces them to do so**. This is fatal to their theory because no ethical considerations force relativism here[121] - in fact, according to Smith, the objectivity platitudes indicate that ethical considerations point to non-relativism in just such circumstances instead.

This *seeming* inevitability of a forced relativism for metaphysical naturalism and its dependence on the flat restriction of the reference of the term 'right' to its causes of use invite considering the question of how robust this 'seeming' is and whether or not the metaphysical naturalist really is restricted to saying nothing more than that 'right' refers to the causes of its uses.

The claim that the normative terms refer to the cause of their use, between cannibals and missionaries, is only acceptable if this condition captures a sufficient amount of the reference fixing description that allows you to identify cross community relevantly similar normative terms. But as we have noted an adequate reference fixing description is typically rich enough to pick up much or all of the pre-theoretic or pre-scientific folk theory of the kind in question and will be able to allow motivated variations away from that theory when a referent is offered that does not play all the causal roles of the folk theory. What we have discovered when we find that the Hare case conflicts with our folk morality is <u>not</u> that metaphysical naturalism is at fault. Rather we have discovered that Hare's case is using an inadequate reference fixing description.

[121] At least according to the objection being made against metaphysical naturalism. It is another matter whether or not this is true.

The effect of this is that even if Smith does not use Hare's inadequate reference fixing description he is making a mistake when he drops the reference fixing description for normative terms out of his evaluation of the idea of reference to the cause of a normative terms use. This is not just because the appropriately folk morality enriched[122] reference fixing description might always avoid all possible cases of divergent reference for normative terms with the same reference fixing description; after all it might not. Rather it is because the case is necessarily underdescribed without the role of a reference fixing description taken into account.

The conditions Smith offers in the above quote (for A and B to successfully use normative terms to refer differently) aren't necessarily flatly false. Either they are misleading because they underdescribe the conditions that need to be met for something to count as the cause of a normative terms use, or they are is false (since failing to provide some account of the restrictions reference fixing descriptions place on successful reference generally and in the ethics case simply gets metaphysical naturalism wrong).

5.1.4 The reply to the first argument

Hare's case of cannibal's and missionaries is a bad case. It does not illustrate the inadequacies of metaphysical naturalism; rather it illustrates what can go

[122] 'Enriched' is not a superlative here. Rather it is, I am arguing, a matter of adequacy that the reference fixing description for moral terms captures some goodly part of folk morality.

wrong if you fail to use an appropriate reference fixing description. Smith's objection to metaphysical naturalism just is that there are cases where metaphysical naturalism with force putative disputants into talking past one another when folk morality tells us that we would not accept such a result. But even Smith's discussion of the Hare case and this problem repeats Hare's error. Both fail to consider the effects of including some or all of folk morality in the reference fixing description of normative or ethical terms.

The upshot is that Smith's first argument does not show, as it purports to, that metaphysical naturalism is bound to violate folk moralities objectivity platitudes without a good reason to do so.

5.2 Reply to Smith's second argument

Smith's second argument is found in footnote 7 (p. 35, TMP) at the end of his discussion of metaphysical naturalism. It makes an analogy between the cases of 'water' and 'right'. The argument is significantly different from the one Smith uses in his discussion of Hare's cannibals and missionaries' case. The argument makes an analogy claim between twin earth type cases of water and similar cases for right. He points out that though we don't mind the breakdown in objectivity that a twin earth water* possibility would force on arguments about the distribution of potable stuff of the kind found in rivers and lakes, etc., folk morality tells us that we would mind the breakdown of objectivity that a twin earth right* case would force on

arguments about what we should do as a matter of morality. I think that it is not clear that Smith's claim about the 'water' case is true. Even if we accept Smith's view of the 'water' case, the claim of forced indifference due to topic change in the 'right' case is only plausible if you exclude folk morality from the reference fixing description for 'right'. And if you do this then you fail to object to metaphysical naturalism at all. The consequences of including folk morality into the reference fixing description of 'right' is that metaphysical naturalism is no longer **necessarily** (and so unmotivatedly) forced to treat a twin earth right* case as a moral relativism case. This is enough to block the argument in footnote 7. What follows is a detailed description of twin earth water/water* cases, Smith's argument, and my reply.

5.2.1 Smith's second argument

Smith begins

> "An analogy might be helpful. The view that 'right' refers to the cause of our uses of the word 'right' might usefully be compared to a related view about natural kind terms; indeed, one way of understanding it is as the view that 'right' is a natural kind term." (TMP, p. 205).

As we will see in the next quote the term 'water' refers to whatever natural kind causes its use. But 'right' is not taken to be referring to the natural kind that causes its use in the above passage.[123]

> "…let's assume, with the metaphysical-but-not-definitional naturalists, that we use the word 'right' to refer to the property of acts that is causally responsible for our uses of the word 'right." (TMP, p. 32).

This passage restricts the metaphysical naturalist to the view that causing the use of a normative term is sufficient to count as the referent for a normative term.[124] I prefer the version of metaphysical naturalism that treats ethical terms *like* natural kind terms. It leaves more options open. However, though we will explore some of these options in this chapter we will avoid taking a position on the theory of natural kinds as such and focus rather on the semantics of natural kind terms since this is where Smith thinks the problem lies. Also though some of our argument here will point out that if 'right' is given the appropriately rich reference fixing description then at least trivial multiple realisation cases like the Hare's missionaries and cannibals are avoided, we still have to consider the possibility of a non-trivial case, one

[123] This only matters if we care to consider the ontology of natural kinds or are insistent on a precise theory of natural kind terms. I am in these arguments indifferent to the ontology (reasonably I think), and will to adopt the of version of metaphysical naturalism Smith hints at in TMP, 205 footnote 7. That is I will restrict my attention to the proposal that treating ethical terms *like* natural kind terms is a good idea.

[124] In addition I prefer the view that natural kind terms like 'water' that suppose we should understand twin earth water cases as providing lessons about the counterfactual reference of water terms (Jackson's for example). But it is clear enough I think that whatever version of metaphysical naturalism we take up Smith's objection to it is given in terms of a possible Twin earth case – that is explicitly not to be understood counterfactually. And I will be going along with him in this regard, especially when we turn to the 'options' for possible twin earth cases.

where the relevant rich reference fixing description yields dual reference in the same possible world. Our argument against the Hare case effectively was that it needed to treat the term 'right' the **same** way the term 'water' is treated — the analogy from a water/water* to a right/right* case can be construed as complying with the requirement of an appropriately rich reference fixing description, at least in the water case.

This is how Smith describes the water/water* case:

> "Suppose our word 'water' refers to whatever natural kind is the cause of our use of the word the 'water'. Then, as Putnam (1981) famously points out, there may well be another community which uses a word, 'water*' say, a word which plays a role in their language just like the role 'water' plays in our language – they may use it to refer to the stuff that comes from river, lakes and streams, is good to drink, and so on – but whose reference differs from the reference of our word 'water'. For whereas the causal history of our word 'water', given that it is a natural kind term, ensures that it refers to H_2O, the causal history of their word 'water*', given that it is a natural kind term, ensures that it refers to XYZ. Thus, even though our words 'water' and 'water*' pay the same role – they are each used to refer to the natural kind that is found in rivers, lakes and streams, is good to drink, and so on – this will not by itself guarantee that we would be disagreeing if we said, of certain stuff, 'That stuff is water' and they said, of the same stuff, 'That stuff is not water*'." (TMP, p. 205).

This passage shows how natural kind terms are susceptible to changes in their content if the relevant circumstances change – or at least relevantly similar natural kind terms if not the same natural kind term.[125] We appear to have not one term here but rather two relevantly similar terms 'water' and 'water*'.[126] They are the same in that they share the same reference fixing description – that they refer to the stuff of our [the language user's] acquaintance, whatever it is, that is found in rivers, lakes and streams, is good to drink, and so on. They are different in that their different causal histories supply them with different referents and this changes the content of any statements using these terms. The key issue for the case of ethics is the dissolution of apparent disagreements about the distribution of stuff into error free talking past one another. With water and water* this is acceptable but in the case of 'right' this is not acceptable. As Smith puts it

> "Whereas the possibility of explaining such disagreements away is acceptable in the case of two communities who use natural kind terms – like 'water' and 'water*' – to plays the same role in their lives, the possibility of explaining such disagreements away is unacceptable in the case of two communities who use a word to play the same role in their lives as the word 'right' plays in our lives. Yet metaphysical-but-

[125] Term identification is not very important here I think. Certainly we can avoid any view on the mater generally so long as we keep track of which context and term pair we are talking about at least – and this is why in what follows both in this and the next chapter we will use water/water* or right/right* where need.

[126] Though this is not the correct description, or not a complete description, if we adopt Jackson's two dimensional modal semantics approach to natural kind terms. Then we could as well say we have one term that refers differently in different circumstances. Smith does not make it explicit which view he favours. It will appear latter in this section that I think Smith should adopt Jackson's view if he has not already, and that doing so shows that the reference fixing description plays a central role in the water/water* case.

not-definitional naturalism leaves open the possibility that we should explain such disagreements away." - Underlining added TMP, p. 205).

It is questionable that we would be indifferent to the existence of water and water*. Of particular interest to us, plausibly, would be whether H_2O and X_yZ would be interchangeable between populations and if not why not. If the substances were interchangeable for the purposes alluded to in the shared reference fixing descriptions then the difference between H_2O and X_yZ would be of some interest. It is not clear what we would make of the circumstance but at least two things seem evident. The water/water* case is not as simple as Smith is supposing, for if we really were indifferent to the difference between water and water* it would seem that we would need an explanation of why we would be indifferent. In at least one possibility, where H_2O and X_yZ are interchangeable for the purposes of drinking, etc., it is not at all clear that we would **not** accept the idea that water and water* were the same term with a previously undiscovered disjunctive referent. This would mean that using the semantics of natural kind terms does not necessarily force talking past one another in the case of water/water*.

Of course it could — perhaps more general considerations would motivate us to agree to a semantics of natural kind terms that insisted on no disjunctive referents or on the causal link to samples of stuff trumping the satisfiabillity of interests in potable thirst quenching liquids. But now we are in a position to understand that, if the semantics of natural kind terms did force a 'water' / 'water*' concept or term difference because of different referents, we would have to be sure that enough of what interests us about the potable stuff

around us that we call 'water' was built into the term's reference fixing description before we would agree that the water/water* case should be treated that way.[127] Implicit in Smith's claim that we would be indifferent to the occurrence of a water/water* possibility, then, is the idea that knowing the 'role played in our lives' by a substance / substances of this kind — a role exhaustively described in a common ground reference fixing description — tells us why we would be indifferent in a water/water* case. If two locally specific equally potable and similarly propertied and relatively distributed substances, both watery, existed, then an assumption of unique realisation fails. But so what? Nothing interesting about potables has changed. Unless we stipulate more details, then we can simply point out that Smith has passed over the reason we are indifferent in the water/water* case — the multitudinous roles played by watery stuff is played by both of X_YZ and H_2O and these roles exhaust our interest in substances to hand.

If this sounds implausible, then we can return to my prior point, that the water/water* case is underdescribed. It is underdescribed because if this 'interests' story sounds implausible then I suspect it is because there really are more general concerns that inform or dominate in the semantics of natural kinds — substance tracking or topic coordination across theory change are examples, and we know they are in play in the 'water/water*' case. And if the case is underdescribed then analogies made to it will be underdetermined

[127] I am suggesting, contentiously for water/water* cases, that organizing natural kind terms reference and meaning in a way that is too stringently independent of our folk interests can motivate the folk to abandon such a usage and this is a reason for a theory of the semantics of natural kind terms to either explicitly motivate deviations from folk interests of this sort or for a theory of the semantics of natural kind terms to be formulated in a more nuanced manner. But really we need not worry over much about water. The claims I am making here are ones Smith would agree with in the case of 'right'. But it should be allowed as common ground between Smith and metaphysical naturalists.

as well. The analogy argument from water/water* to right/right* appears to fail.

The underlined segments of the last passage from "The Moral Problem" highlight another important feature of Smith's objection that is not as obvious in the first version. This is just that it is not actual cannibals that are the real problem. It is the possibility of such that is the problem. Smith objects to metaphysical naturalism because it has to allow that a right/right* case (just like the water/water*) case is possible. And this possible case would count as a form of relativism. As it stands Smith appears to be saying that allowing the possibility of relativism is enough of a reason to reject metaphysical-but-not-definitional naturalism. This can't be right and, of course, it is not what Smith is objecting to.[128] Smith's objection is that metaphysical naturalism has to allow a possibility that folk morality would not allow and it has allowed it for the sake of a semantic theory rather than ethically relevant reasons. So much the worse for metaphysical naturalism if this is the case, according to Smith.

As it stands, it seems that we should agree with him. My objections above to the argument only show that a natural kinds treatment of ethical terms does not necessarily have to interpret the right/right* case as relativistic because the water/water* case need not have this upshot unless it is further described. We can suppose that with 'water' and 'water*' we could stipulate that there really is a failure to disagree about the distribution of types of potable

[128] As we have discussed above, there are at least two kinds of relativism that **must** remain open or Smith's meta-ethical theory would fail to be defeasible in the right way. It is not relativism that is the problem. It is the reason for it occurring in the right/right* case.

substances and that, since we are only interested in potable distribution (not types of potable distribution) we can indeed be indifferent about water/water* possibilities. And prima facie, just as Smith points out, the objectivity platitudes suggest that we simply are not indifferent in the case of right/right*. But is this an inappropriately theoretically motivated possibility?

We should notice two things. First, Smith appears to object to a possibility (produced by an inappropriate use of a merely semantic theory), but he doesn't directly consider what such a possibility would be like if we had rich reference fixing descriptions for ethical terms. Secondly, relativism is only entailed if the protagonists take themselves to be disagreeing when they are not. But were such an eventuality to arise we have two directions to go. Considering whether the imagined protagonists, after finding much in common via a rich reference fixing description, will continue to disagree is one. They could react to a right/right* case by accepting disjunctive reference for one common term 'right'. The other way to go is permitting that the imagined protagonists are talking past one another and accepting that the dispute is a false dispute where there should be a genuine dispute, with relativism being the result, but denying that the cause of this relativism is illegitimate.

There is an interesting feature to the right/right* case. When two communities find that they have different referents for the term 'right' while sharing the same reference fixing description they have thereby discovered that they have different terms because their terms have different referents.

But this is not how Smith reads the case and it is worth noting why. Smith thinks that rather than saying we have two words where we thought we had one we should think instead that we have a case of relativism. But for that to be right we have to accept that the reference fixing description for ethical terms is the ***only*** guide to what is going on with the term. One way of putting this is that the reference fixing description, in the case of ethical terms, is what determines your term count. Smith might say that prior to investigation the metaphysical naturalist has access only to the reference fixing description and counting two terms where you thought there was just one *is* relativism if that is all you can say about the different referents. This is just because you have no a priori reason to claim that one or the other of the two new terms (or two referents) is the one that both parties should adopt, and we know a priori from the objectivity platitudes that we can't adopt both. We know that there should be some mechanism for selecting between the two new terms (or the two referents) and metaphysical naturalism fails to provide one. Put this way, however, the objection is soluble. A metaphysical naturalist can use this possibility as a reason to revise the reference fixing description in order to secure a better fit with folk morality by using the latter as a source for that reference fixing description.

If the claim that the folk morality that is the source of the reference fixing description is correctly and fully explicated by that reference fixing description (aside from the objectivity platitudes), then we can effectively stipulate that a right/right* case just is one where folk morality turns out to explicitly accept disjoined reference. If this amounts to relativism, then folk theory would appear to have generated an error theory implicating relativism because folk theory supplies the reference fixing description in the first place.

Moreover folk morality explains why relativism is the accepted result in the right/right* case, and this leaves us with folk morality opposing its own objectivity platitudes (at least if you insist that these platitudes require unique referents). Given that Smith's objection to the metaphysical naturalists treatment of the right/right* case just was that relativism induced merely by a semantic theory should be rejected, we can offer a successful reply to this objection by insisting that folk morality has filled the role of supplying the relevantly rich reference fixing description for 'right' and 'right*'.

Not allowing metaphysical naturalism to use folk morality (as the source of a reference fixing descriptions) is simply to return to a variant of Hare's case, which should be resisted. When folk morality supplies the reference fixing description then, by hypothesis, folk morality finds nothing morally relevant with which differentiate the right and right* circumstances. If the right/right* case is possible then it is only possible if folk morality deems it so. If this is relativism then it is the error theory of a relativism that arises from the surprising discovery that folk morality is conceptually incoherent.

5.2.2 Reply to Smith's second argument

Given the above discussion, it is clear how we can respond to Smith. The possibility of two communities using the same reference fixing description to pick out different referents should be left open if you are treating ethical terms like natural kind terms. So just as there could be a water/water* case

there can be a right/right* case. Smith's objection is that metaphysical naturalism leaves this possibility open inappropriately. It must permit relativism when folk theory rejects it and it does so for reasons that are irrelevant to moral theory. Smith thinks these morally irrelevant reasons are to do with the semantics of natural kind terms, and implies that the natural kinds approach should therefore be dispensed with.

The first problem with Smith's objection is that treating ethical terms like natural kind terms does not necessarily force the metaphysical naturalist to accept different referents as equally good. That will entirely depend on the reference fixing description. As we saw in the Hare case of cannibals and missionaries a minimum requirement is that there be an appropriately rich reference fixing description, which there is not in the Hare case. Stipulating that there might be a possible right/right* case just like the water/water* possible case, with the same reference fixing description, as rich as you like, picking out different referents for different communities in the same possible world, avoids the inadequate reference fixing description reply.

However, there are two lines of reply left. The first is that claim that we are indifferent to the 'water' relativism forced by a twin earth water/water* case is not clearly true. What we would make of the water/water* case would depend a lot on the particular details of the water/water* possibility. Thus an analogy to the water/water* case for the right/right* case is of little use as the water/water* case is underdescribed.

The second line of reply is that the reference fixing description is plausibly just the explication of folk morality. Hare's cannibals and missionaries case made the mistake of not using a folk-morality-informed reference fixing description and this is why it allowed objectionable disjunctive reference for the term 'good' to occur. If it is possible for the best explicit statement of folk morality to yield different referents in different contexts[129] then though a metaphysical naturalist might be forced to accept that such a possibility is a form of moral relativism and so induces an error theory, they are not forced to accept that it is inappropriate relativism. For Smith describes inappropriate relativism as favouring, for merely theoretical reasons, a relativistic outcome in a right/right* case when folk moralities objectivity platitudes recommend rejecting relativism. Smith imputes that metaphysical naturalism has no recourse but to fly in the face of folk morality to its own detriment. But this is not at all clear, since metaphysical naturalism has no plausible source of a reference fixing description other than folk morality. If we accept that the right/right* case prompts close examination of the folk theory and reference fixing description and the only thing that results is that the same reference fixing description has different referents and we find no reason to accommodate this as a disjunctive reference case, then we have relativism. But it is relativism that folk morality can't show the error of by hypothesis. So we are left without grounds for objecting that this is inappropriate relativism caused by metaphysical naturalism. We have, instead, by hypothesis, entirely appropriate relativism an error theory probably due to a surprising conceptual incoherence in folk morality.

[129] Smith's anti-Humean rationalism is explicitly circumstance relative. Presumably then the problem with context relativism is just that no morally relevant difference differentiates them, for if they did then we would not have a right/right* case at all. So Smith's problem is with morally **unmotivated** relativism.

The best conclusion here is to simply reject Smith's argument against metaphysical naturalism. It depends too much on stipulations without **any** support from folk morality, and though some sort of possibility of a right/right* case is open, given a natural kinds treatment of ethical terms, what should be made of this is profoundly unclear. Relativism might indeed be possible. Alternatively, relativism of this kind might be impossible and a disjunctively referring objective ethics (in a sense at least) might be the upshot. Whatever the case, Smith's argument does not succeed.

5.3 Next

Perhaps a 'by hypothesis' argument isn't really good enough here. Metaphysical naturalism does look like it has to allow at least the bare possibility of a twin earth right/right* case, and the reply to Smith's objection to this possibility stipulates that either folk morality will show right/right* cases are impossible or that right/right* cases reveal folk morality is incoherent. It is hard to shake a degree of residual scepticism about the effectiveness of this reply to Smith's objection to metaphysical naturalism. We will try then to reformulate it with an example of a candidate explication of folk morality playing a role as a reference fixing description

In the next chapter we will use Smith's own explication of folk morality, and in particular features of desiderative unity, to recreate the possibility of a twin earth right/right* case and see if any of these cases are objectionable in the way Smith supposes metaphysical naturalism is objectionable.

Chapter 6 Using Smith's meta-ethical theory to evaluate 'Twin Earth' cases of right

In the last chapter I argued that both Smith's arguments against metaphysical naturalism in TMP had problems with the reference fixing descriptions used in his cases and so has problems with making his charge of inappropriate relativism against metaphysical naturalism stick. The first argument, using Hare's cannibals versus missionaries case, got metaphysical naturalism wrong by relying on a dictionary to provide an unusually and inappropriately austere reference fixing description. The second argument underdescribed the details of the water/water* case, making it useless for the purposes of analogy. Importantly it also failed to appreciate implications that arise given that the appropriate source of a reference fixing description is the associated folk theory. In the case of 'right' the reference fixing description that a natural kinds treatment of ethical terms should use is the best explication of implicit folk morality. It is unclear what this explication of folk morality is but usefully Smith has gone to some length to supply us with his version of it. In this chapter we will use the key features of Smith's anti-Humean meta-ethics in the role of reference fixing description, and evaluate the possible Twin Earth right/right* type cases we can make with it.

Before we begin on this task, it is important to remind ourselves that for Smith Twin Earth is not a counterfactual, other-worldly version of actual Earth. Earth and Twin Earth are supposed to occupy the same possible world, where the disparate referents of the same natural kind term are instantiated in separately locations, and being tracked by different communities who share this natural kind term. The terms have the same

reference fixing description, but in the different contexts of the different communities refer differently. The '*' in water* is just a context indicator.[130]

6.1 Setting up the case

Evaluating Smith's objection to metaphysical naturalism requires we consider its two parts. The first is that metaphysical naturalism will leave open possibilities that require us to interpret the case as an instance of moral relativism. The second is that when metaphysical naturalism does this it must contradict the folk moral objectivity platitudes inappropriately. As I argued in chapter 5 'inappropriate' is not idle in the previous sentence since leaving open possibilities for error theory, as we must, requires we allow for the possibility of relativism. And one way for relativism to be the case is that we find that folk morality is incoherent. So the possibility of flatly contradicting the objectivity platitudes is not the problem that Smith's objection raises. Rather Smith's objection is that metaphysical naturalism must allow the possibilities of contradicting the objectivity platitudes without support from folk morality; that is metaphysical naturalism contradicts the objectivity platitudes of morality without a reason for doing so that gets at least part of its warrant as a relevant reason from folk morality.

Setting up a Twin Earth case with two communities using the same reference fixing description grounded in Smith's explication will help us evaluate

[130] See TMP, p. 205, footnote 7. In this passage Smith doesn't specifically say that he is working with the same-world version of Twin Earth cases as opposed to the counterfactual version, but we can take it for granted that he is; the other version raises no interesting issues for the views we are discussing. Note that from here on I will use the terminology of '[possible] Twin Earth' to refer to this same-world possibility of two disparate locations in the same possible world; when using 'Earth[ers]' and 'twin Earth[ers]' I am talking about the two locations and their denizens.

Smith's objection. But even if we grant that Smith's anti-Humean rationalism is a correct explication of folk morality we know we don't have a vital component – a theory of what desiderative unity is (as opposed to what it must enable as far as the desiderative profiles of the relevant population of fully rational idealisations of agents are concerned). In what follows we will simply have to grant that what desiderative unity needs to enable it does enable. As we know what must be 'enabled' is idealised fully rational desiderative convergence. This adds a complication to setting up a Twin Earth case for metaphysical naturalism using Smith's explication of folk morality in the role of a reference fixing description. 'Right' at a possible world applies to what it does because of facts about another possible world. (In TMP, 151) Smith calls the possible world of agents being idealised the evaluated possible world and the possible world with fully rational idealisations of these agents in it the evaluating possible world. When applied to 'us' this makes the evaluated possible world the actual one, but for a Twin Earth case it is entirely open – Twin Earth is a possibility but we don't restrict it relative to the actual world. We might think that this should be a factor in how to understand Twin Earth cases. But I think we should suppress this complication. Smith, in his first version using Hare's story of missionaries and cannibals, seems to me to be indicating that what is really interesting is how metaphysical naturalism pans out when faced by psychological and historical diversity among people in the actual world. The point of a metaphysical naturalist using an explication of folk morality in the role of a reference fixing description is to preserve a kind of relevance to us for a Twin Earth case. This means we should resist fitting the Twin Earth case into a counterfactual format by bringing in the contrast between evaluated

and evaluating possible worlds. This contrast is not where Smith's objection lies.[131]

As I have argued at length in previous chapters, the pivotal piece of Smith's theory is desiderative unity. It is *in virtue* of this cluster of properties, whatever they are, that there are normative and moral reasons if there are any such things. And moral reasons pick out both the properties of acts that make them right and the properties of agents psychologies that make them correctly disposed towards these properties. Wherever there are right-making properties of acts then there have to be agents with a psychology which instance the desiderative unity features that lead to convergent rational idealisations of these agents. I think that, since we don't have a theory of desiderative unity that would allow us to be more precise we can simply stipulate that both communities in the Twin Earth case are such that they simply realise Smith's explication of folk morality. But they may or may not do this in a relevantly different way (we are trying to imagine that they offer different right-making properties for acts in the same circumstances so we can't insist before we get going that the communities simply are not otherwise different without thereby begging questions against both Smith and the metaphysical naturalist). I think the easiest way to keep track of these relevant variations is to simply focus on desiderative unity – its instantiations in the Twin Earth communities of our case and in its maximised form in the idealisations of the members of these communities. We will assume that each community taken alone idealises to a convergent

[131] If we decide that, none the less, we should be considering the counterfactual story, then I think we have to conclude that whatever else is going on we would have to defer understanding what Smith's objection amounts to until actual moral progress reaches fruition or terminates in error theory. That is question begging all around; it suggest the metaphysical naturalist can't be objected to but also can't be vindicated. In that case Smith can't reject a metaphysical naturalism but only because we have deferred considering the objection he has indefinitely. This seems a mistake.

fully rational idealisation where the desire sets of the convergent idealised population display the same maximum desiderative unity property. I prefer property talk here and so think of a maximised desiderative unity feature as a maximum desiderative unity property – the same property for all desiderative profiles in the desideratively convergent fully rational ideal population needed for right-making-relative-to-circumstances properties of acts. This forces me to talk about desiderative unity properties instantiated in sub-fully rational agents. Cashing out Smith's idea of increasing desiderative unity stepwise to a fully rational ideal maximisation, we can see these properties as forming clusters indexed to a maximum desiderative unity property. But we could replace property talk with something else if need be. We know what the desires of the full rational are meant to do and how they are related to our desires and reasoned desire changes if Smith's theory were true. We know how norms are constituted by desiderative unity properties of the relationships between the contents of desires in a set of desires (of an agent)[132] This is the basic structure we need in what follows.

As we have seen in previous chapters, for Smith, schematically,

> "our ϕ-ing in circumstances C is right if and only if we would desire that we ϕ in C if we were fully rational, *where ϕ-ing in C is an act of the*

[132] 'An agent' adds two kinds to thing into the mix: beliefs, and the capacity to change states of mind relative these beliefs, and the capacity to acquire true or evidentially warranted beliefs. None of these capacities have to be perfect in the manner they are assumed to be at the ideal of full rationality. But these features will not bear on the Twin Earth case so I will assume they have no impact. The reason for this is that we are not interested in the Twin Earth's population's abilities to become more like an ideal. We are interested in what 'right' is for those populations and whether the possible Twin Earth case allows an objection of inappropriate relativism against metaphysical naturalism using Smith's explication of folk morality as a reference fixing description.

appropriate substantive kind: that is, it is an act of the kind picked out in the platitudes of substance" (TMP, p. 184)[133]

And we have the by now familiar two stage reductive argument

"Conceptual claim: Rightness in circumstances C is the feature we would want acts to have in C if we were fully rational, where these wants has the appropriate content

Substantive claim: Fness is the feature we would want acts to have in C if we were fully rational, and Fness is a feature of the appropriate kind[134]

Conclusion: Rightness in C is Fness" (TMP, 185)

The 'we' in this argument necessarily refers to a population given that objective and so moral and normative reasons require desideratively convergent idealisations of that population. Plausibly the population Smith has in mind comprises actual agents, and at least humans. In a Twin Earth case the 'we' refers to the population of Twin Earth. Acts in a circumstance have a right-making property (Fness in the above argument) and it is the

[133] As we have seen in previous chapters the 'platitudes of substance' distinguish moral norms from the ones that correspond to normative reasons and agent-specific reasons, where agent-specific reasons are not normative reasons and where moral reasons are a subset of normative reasons. The requirement to distinguish normative from moral reasons is met using what I have called the recognisability constraint. This constraint is just that the relevant desires contents concern recognisably moral matters. I have argued this recognisability constraint is innocuous enough. Important as these distinctions are (especially between agent-specific reasons and moral reason [see TMP p 183: Smith's discussion of personal perfection) they will not matter to the Twin Earth case and Smith's objection to metaphysical naturalism.

[134] Ibid.

same property for all acts relative to a circumstance and for all the member of the 'we' population whose rightness standard is in play.[135]

If we are to set up a possible Twin Earth case (of right/right*), allowing the metaphysical naturalist to use Smith's anti-Humean rationalism as a reference fixing description, we have to describe the two populations in play (of earth and Twin Earth) in the right way. Both have to be equally good candidates to realise Smith's anti-Humean rationalism. The composition of the populations matters since it is relative to them that the relevant rationalist ideal is specified in Smith's explication of folk morality. But we will simply assume what we need to of them to allow 'rationality' as Smith's theory describes it to be applied to them. As I have discussed above, the important features that allow Smith's rationalist theory to be applied to a population all concern instanced desiderative unity properties and an ideal maximum desiderative unity property. We will assume that the Twin Earth populations' psychologies instantiate desiderative unity. We will assume that the members of each community can be idealised to fully desideratively rational versions of themselves (in the sense that it is true that such an idealisation is possible at least) and that members of a community idealise convergently. I will name the feature of a possible Twin Earth community's having the *same* maximum desiderative unity property (ideally) the community's maximum desiderative unity. The possible Twin Earth case involves two communities. Because I want to emphasise their similarity I will name them community 1 and community 1* (C1 and C1* for short or Earthers and twin Earthers*). Finally we have to stipulate that there is a circumstance

[135] Smith accepts that there is plenty of room for complications of relevant sorts, of varying degrees of complexitiy: For example TMP (pp. 154-155 – akarasia cases, pp. 193-194 Foot's rational villain and Harman's rational criminal), Smith: (1996a), (Bigelow and Smith: 1997) just to cite a few. But we will simply set these complications to one side since we should grant Smith as much as possible so we can evaluate his objection to metaphysical naturalism.

C relative to which acts are either right (for community 1) or right* (for community 1*). And as we saw above with Smith's reductive argument, right for agents in a circumstance C requires the acts in C all have a property[136] Fness where Fness is the object of the relevant convergent desires of a relevant ideal population relative to the acts of agents in circumstance C. For community 1 let us call the relevant right-making property *r-ness* and for community 2 let us call the relevant right-making property *r*-ness*.

6.1.1 Possible Twin Earth right/right cases using Smith's rationalism*

So possible Twin Earth, for metaphysical naturalists using Smith's explication of folk morality as a reference fixing description, should be like this:

> There are two communities that don't interact (community 1 and community 1*)
>
> Both populations ideally rationalise to convergent fully rational versions of their community members. Members of community 1 idealise convergently because the fully rational versions of the members of community 1 have desiderative profiles that instantiate the same maximum desiderative unity property – call it *Dmax*, for short. Community 1* is precisely similar in this regard and the idealisation of them involves a population that instantiate the same maximum desiderative unity* property – *D*max*.

[136] Smith and Jackson allow a very open-ended notion of properties (though neither discuss perfectly natural properties since they don't need them to achieve reductive naturalism or descriptive reductivism). It allows bundle of properties to be bundled into a new property, for example. But I will simplify out the tricky business of identifying properties and any complications by talking about one property making acts right (or right*).

> For acts relative to circumstance C that act is right if it instantiates r-ness (that is, satisfies 'right' as used by members of community 1). For acts relative to circumstance C that act is right* if it instantiates r*-ness (that is, satisfies 'right' or 'right*' as used by members of community 2).

It seems that for any of this to be at all relevant we have to assume that the right making properties r-ness and r*-ness are not the same. We will simply stipulate this. It might be somewhere between hard to impossible to imagine that it could be that two communities are as densely similar as Smith's rationalist meta-ethical theory would require them to be when you use it in the way I am using it here, as a reference fixing description, and disagree as proposed over the distribution of moral goods. I certainly have this intuition. However, we can't leave it at that without trying to formulate and evaluate possible Twin Earth right/right* case or else we end up begging the question in both directions. So, for now, we will assume that we can move on to the next step.

We have laid out the features of Smith's anti-Humean rationalism that we need in play to allow a metaphysical naturalist to use Smith's anti-Humean rationalism as a reference fixing description. The metaphysical naturalist I am considering here treats ethical terms like they were natural kind terms, like they were like the term 'water'. This allows the following problem to be disposed of prior to turning to the cases below. The problem is that, as I have argued in previous chapters, the normative component of Smith's theory –

the desiderative unity component – can't be defined in causal terms[137]. This means that a metaphysical naturalist looking to use Smith's theory as a reference fixing description will necessarily have to adopt a more sophisticated view about the relevance of 'cause' to the reference fixing mechanism. So though local community facts about sub-fully rational psychology are what plays the role of 'stuff' in the 'stuff we are acquainted with that '….'' it is not plausible that merely causing the use of the term 'right' or 'right*' is all it takes to count as the referent for the term right. The possible Twin Earth cases have to be understood in this way if we are not to beg the question against metaphysical naturalism.

6.2 Different options for the Twin Earth case

I am going to formulate the Twin Earth cases with more specific claims about the characteristics I argued are the relevant ones in the last section. The pattern of argument within each more specified possible Twin Earth case (call them possible Twin Earth case options) will to formulate Smith's objection to metaphysical naturalism given the option and consider replies from metaphysical naturalists. The replies will come in three flavours:

[137] We have made note of this issue earlier in the thesis. Here is a *very* cursory account of the reason. Giving a causal account of desiderative unity would require a causal characterisation of the contents of desires since desiderative unity is a property of the contents of desires. This account would have to describe whatever it is about desiderative unity that is normative and do so causally. But for fully rational agents this would amount to supposing that there was a state (a content state) that disposes the agent to acquire some particular desires. And then we have a problem – because nothing prevents this state being accessible to a fully rational agent and thus playing a role in the content of a belief and that would mean that a rational agent would simply by having a belief with the right contents come to instantiate a motivation, without relying on desires. This contradicts Smith's commitment to the Humean theory of motivation.

Showing that the option leaves the possible Twin Earth case underdescribed – blocking Smith's objection to metaphysical objection because it becomes unclear that option really counts as a Twin Earth case

Showing that the option makes relativism impossible - blocking Smith's objection to metaphysical naturalism by making it impossible to make.

Showing that the option makes relativism possible but showing that folk theory would have to accept relativism too (albeit at the likely cost of error theory) - blocking Smith's objection to metaphysical naturalism by showing relativism is not inappropriately caused by merely theoretical considerations

The options vary relative to the assumptions you can make about the relevant psychological similarities between members of community 1 (Earthers) and community 1* (twin Earthers). The relevant extremes will be captured, I think, first by supposing that the maximum desiderative unity property is the same between community 1 and community 1* and then looking at what might be said (options 1 and 2) and then supposing that the maximum desiderative unity property is not the same between communities 1 and 1* and then looking at what might be said (options 3(a), 3(b), and 3(c)) . As I have noted above we will simply insist that the referent for right and right* remain non-identical.

6.2.1 Earther and twin Earthers are relevantly psychologically similar

The ideal populations generated from community 1 (Earthers) and 1* (twin Earthers) instantiate the same maximum desiderative unity properties in all their members. These options are the least friendly towards mounting Smith's objection to metaphysical naturalism because sameness in the relevant components of the psychologies of the idealisations of Earther and twin Earther populations makes it look impossible to maintain the stipulation of differential reference between the communities. We will differentiate option 1 and two by the explanations that a metaphysical naturalist might give for differential reference under the condition of relevant desiderative homogeneity.

Option 1 – metaphysical naturalists look for differential circumstances[138].

The way we should formulate the Smith objection here, I think, is to assert that since we have stipulated differential reference the metaphysical naturalist has to go along with the stipulation when folk theory offers no support for the move. The idea is that even though a metaphysical naturalist can qualify the relevance of cause in the acquaintance and reference fixing story of ethical terms they can't get rid of it entirely. So long as r-ness and r*-ness play the right causal role and are different then the metaphysical naturalist is required to posit relativism. But explicit folk theory tells us nothing important is different between the communities so the fact that the r-ness and r*-ness properties are not the same should not matter. Metaphysical naturalism must call for relativism when folk theory does not.

[138] That is the metaphysical naturalist tries to find a reason to suppose that the circumstances relative to which 'right' and 'right*' refer as they do turn out to be relevantly different, rather than relevantly similar as Smith's rationalism requires.

It is worth remembering what relativism is meant to be in the course of these objections. If metaphysical naturalism is right then two communities that look as if they share a concept (say, the concept 'the watery stuff of our acquaintance') – that is embedded in their use of a word like 'water' might find that different material kinds of stuff are being referred to in the two communities. The old refrains 'water is H_2O' and 'water* is XYZ' captures this difference. If the terms in play are natural kind terms then referring differently is enough for arguments over whether H_2O is 'water*' or 'water' is XYZ really are wrong headed. As Smith puts it

> "…even though our words 'water' and 'water*' play the same role, …., this will not by itself guarantee that we would be disagreeing if we [Earthers] said, of a certain stuff, 'That stuff is water' and they [Twin Earther] said of the same stuff, 'That stuff is not water*'. … Whereas the possibility of explaining such disagreements away is acceptable in the case of two communities who use natural kind terms - like 'water' and 'water*' – to play the same role in their lives, the possibility of explaining such disagreements away is unacceptable in the case of two communities who use a word to play the same role in their lives as the word 'right' plays in our lives. Yet metaphysical-but-not-definitional naturalism leaves open the possibility that we should explain such disagreements away." (TMP, p. 205 footnote 7)

The metaphysical naturalist in all the options we are considering is attempting to make a possible Twin Earth case where they obey the injunction in implicit this passage

> "… two communities who use a word to play the same role in their lives as the word 'right' plays in our lives "

The first option reply to Smith is to suggest that since they have complied with the injunction that the possible twin earth case communities use their words to play the same role in their lives as the word 'right' plays in our lives then she is within her rights to look for a flaw in the case description. The flaw, she might say, is the assumption that the circumstances are the same in the possible Earth-twin Earth case contexts. The contexts in the case (in the example locations) differ. By stipulating different referents for the term 'right' and 'right' and leaving the relevant psychological similarities in place and by stipulating that Dmax and D*max are the same we have effectively stipulated an error in the other important conditions. Since the relevant psychological similarities remain in play, we can look to circumstances. We are suppressing the relevance of context to circumstances and right is determined relative to circumstances.

The problem with this reply is that it is a merely a mistake if it turns out that description of the case of the Earthers and twin Earthers we are using is inaccurately described in the right way. and a mistake that we can either fix, or reasonably stipulate away. The general possibility of suppressed detail is, as a mere possibility, perfectly sensible. But it is irrelevant here. The Twin Earth case is formulated with just this sort of detail in mind. To succeed this line of thought would have to show every possible Twin Earth right/right* case *necessarily* suffers from this kind of problem. Clearly this is not achieved by this line of reasoning. The reply to the relativism charge we are entertain here amounts to changing the Twin Earth case into something different and

consequently rendering the consideration in the reply irrelevant. The metaphysical naturalist is adopting a fruitless strategy here.

The claim in the last paragraph might be correct (though as I have indicated above my intuitions favour the idea that under the conditions of psychological similarity in this case possible Twin Earth cases turn out to be impossible). If it is correct we don't have a general reply. However it does show that details of Smith's own theory can be used to block **some** putative possible Twin Earth cases. This gives us reason to consider the following:

For Smith's objection to work against metaphysical naturalism it has to be the case that Twin Earth cases are possible and that metaphysical naturalism must **in all such cases** make the wrong choice – that is it must be forced to assert in the face of folk theoretic objections that a Twin Earth case is a case of relativism. Have we shown, by showing that sometimes there are folk theoretically motivated ways to block the formulation of a Twin Earth case, that Smith can't mount his objection? No we have not, after all blocking some ways to formulate a Twin Earth case is not the same as blocking all the ways to formulate a Twin Earth case. But we also now see I hope that we can't also **assume** that metaphysical naturalism must make the wrong choice in favour of relativism under any of the remaining possibilities either.

We should conclude that option 1 leads to a standoff with no conclusive objection to, or definitive defence of, metaphysical naturalism.

Option 2

This is just a variant on option 1. We import Smith's objection. The difference here is the reply. The metaphysical naturalist will suggest that the reason the possible Twin Earth case where the ideal maximum desiderative unity property is the same is not a case of relativism is because there is something wrong with the stipulation that 'right' and 'right*' refer to r-ness and r*-ness respectively. Rather, possible Twin Earth involves a way that the reference of 'right' and 'right*' turns out to be the same: that is right and right* refer to the disjunction of r-ness r*-ness.

This reply has problems. For we might well ask what is different between Earthers and twin Earthers that explains why the former think they are referring to r and the latter think they are referring to r* when all along they are referring to the disjunction (something they would presumably discover if they were to discover each other and perhaps begin to mingle)? If we give any kind of relevant difference in the contexts of Earthers and twin Earthers we collapse this option into a version of option 1. If we suppose, as might be more plausible, that as it happens Earthers and twin Earthers have simply not completed enough cycles of deliberation to find the disjunction then we still find we collapse to option 1; we have supposed a contingent difference of a different sort, but the standoff still applies. The Twin Earth case is not about the contingent variations in knowledge, deliberation, circumstances or whatever of the possible populations. As we said in discussing option 1, to allow variations in features like the sameness of circumstance required by rationalism for terms like 'right' to refer is a mistake that should be reasonably stipulated out of Twin Earth cases.

The metaphysical naturalist could try a bit of bullet biting at this point. They might simply insist that any relativism incurred on their view is not the result

of the semantics of natural kind terms alone. Rather, it is that we discover that explicated folk morality turns out not to determine the right makers of acts in a determinate manner, but only in a manner that implies relativism. This is a more promising option. Metaphysical naturalists are effectively shifting the blame. And if things are as specified then it seems quite reasonable that they do so, given that we have stipulated Smith's explication of folk morality is playing the reference fixing role. What the metaphysical naturalist is suggesting is that relativism occurs in the twin earth case *because* the term 'right' only refers 'right' indeterminately due to given folk morality's underdetermination of its reference.[139] The twin Earth case shows this because the standards and conditions for folk morality to refer are being used to their best advantage in the Twin Earth case. We have stipulated as much to formulate the case. Twin Earth is not relevantly different from us in this regard – the Twin Earth case is one where folk apply *our* Folk morality (in its explicit rationalist format). Trying to insist, against this, that the populations in the Twin Earth case are relevantly unlike us (*actual* us) invites the reasonable and immediate demand for an account both of the difference and of why it is not one available to the metaphysical naturalist.

However, though indeterminate reference provides an interesting option for the metaphysical naturalist it does not show that every version of the objection Smith tries to make relative to Twin Earth cases is blocked.. Even if the metaphysical naturalist has shown that it is impossible for Smith's objection to be made against the position outlined above, one thing at least remains. We have not considered if it is possible to create a Twin Earth case using Smith's rationalism and the condition that Earther and twin Earther

[139] See Field (1973) and the ensuing discussion among philosophers of science of indeterminacy of reference as it applies to terms in the natural sciences (e.g., 'mass' and 'gene').

communities are relevantly psychologically *different*. That is, we have not yet considered what to make of the condition that the full rational idealisations of Earthers and twin Earthers instantiate different maximum desiderative unity properties. We can't conclude that Smith's objection to metaphysical naturalism can't be made until we consider Twin Earth cases under conditions of relevant desiderative heterogeneity.

6.2.2 Earther and twin Earthers are relevantly psychologically different

This set of options is the most friendly to Smith objections to metaphysical naturalism. Here we are imagining that community 1 and 1* in the possible Twin Earth right/right* case are relevantly psychologically different. The members of each community are similar with community cohorts and different from the members of the other community. We are as before imagining that Smith's explication of folk morality is used in the role of a reference fixing description and in that role it can be applied to all the folk in the possible Twin Earth case equally well. But members of community 1 when idealised, idealise to a desideratively convergent fully rational population who, for agents acting in circumstance C, desire that those acts instantiate the property r-nesss. And members of community 1* when idealised, idealies to a desideratively convergent fully rational population who, for agents acting in circumstance C, desire that those acts instantiate the property of r*-ness. Community 1's people idealise to a fully rational population whose desire sets instantiate the same maximum desiderative property, Dmax. Likewise community 1*'s people idealise to a fully rational population whose desire sets the same maximum desiderative unity

property, D*max. But it is not the same property between the communities and their idealisations: Dmax and D*max are not the same.

This looks like it is enough to force the metaphysical naturalist to agree that the possible Twin Earth case under this condition yields relativism. It is even explained by the facts of communities' psychology and ideal psychology. And there is no obvious reason to suppose this situation is impossible. We don't have a theory of desiderative unity to supply an account of how it is impossible. And even so surely a metaphysical naturalist should be indifferent to that fact. What matters is just that possible Twin Earth with relevant psychological difference between communities 1 and 1* looks as if it will be just the kind of situation where metaphysical naturalism has no reason to avoid calling it relativism.

Smith's objection then reminds us that

> "…in the case of two communities who use a word to play the same role in their lives as the word 'right' plays in our lives…"(TMP, p. 205 footnote 7)

it is unacceptable to explain away disagreements over the application of 'right' between members of community 1 and 1*, disagreements expressed when they say of some act 'That act is not right' and of that same act 'That act is not right*'. Explaining away these disagreements as talking past one another is relativism. So the problem is that if possible Twin Earth people really were like us and the words they use really were like ours playing the role our word 'right' does in our lives, then we should not accept that this is relativism. Yet metaphysical naturalism does.

At this point I think the metaphysical naturalist should refuse to follow along with the parochialism implicit in this objection. Now we appear to be insisting that the mistake metaphysical naturalists are making is evaluating the twin earth case from the point of its inhabitants when we should be evaluating it from 'our' point of view where 'right' is playing the role it does in our lives. But what is left of that role that the twin earth populations lack? What mistake is there in evaluating twin earth in the way that leads to relativism given that we have used the reference fixing description to import our folk moral concepts into the twin earth case whole sale? It seems to me that metaphysical naturalists should accept, under the condition that both populations instantiate at the ideal the same maximum desiderative unity , that they should call the case one of relativism. The differential reference in the possible twin earth case is not only relevant, it is relevant in a way that metaphysical naturalists should pay attention to, especially armed with Smith's explication of folk morality in the role of their reference fixing description.

The reply to Smith's objection, given all this, is to argue that calling relativism under these conditions is the right thing to do. After all how dissimilar can the twin earth communities members be from us, in any fashion that counts? Can metaphysical naturalists show that they have folk theory relevant reasons for describing possible twin earth as a case of relativism? I think that the problem with interpreting Twin Earth cases is working the scope of the term 'their' and 'our' for claims like

> "…in the case of two communities who use a word to play the same role in **their** lives as the word 'right' plays in **our** lives…"(Bold added. TMP, p.205)

when talking about possible Twin Earth cases.

There is a simple way to show why this feature of Smith's objection is relevant and potentially tricky. In possible Twin Earth why do we care about isolation between community 1 and community 1*? In a nutshell we should not care about the *isolation* of the twin earth case populations from each other for its own sake. What matters about the members of community 1 and 1* is whether or not they are different in an *important* way and we (the 'our' in the quote above) are the ones who define what is important in the first place (how else could it be after all).The twin earth case for ethical properties is always going to require a careful gauge of the role of our actual standards play in formulating and evaluating twin earth cases. But this observation motivates an immediate folk theory relevant reply to Smith's objection. Members of community 1 and 1* could just as well be members of one community. If that were the case then relativism would reflect a relevant psychological incoherence, of sufficient magnitude to yield 'folk theory induced relativism'. And the only question is 'by whose standards is this folk theory induced relativism?'. The metaphysical naturalist answers this question by pointing to the source of their reference fixing description which is an explication of our actual folk morality. Smith's objection to metaphysical naturalism was that it failed to pay attention to folk morality in the right way. That objection can't be made any more. .

There is only one residual worry – telling the story I just told involved treating the members of the twin earth case populations (community 1 and 1*) as relevantly similar and claiming that the members of both communities are the agents from whom the relevant idealised population should be developed. Should we allow this move?

The metaphysical naturalist can motivate using members of both twin earth communities as the base from which one idealised (non-convergent) fully rational population is developed. Roughly, folk morality, according to the metaphysical naturalist, tells us that any agents who can realise the reference fixing description derived from Smith's rationalist meta-ethics should count as part of the same community. for the purposes of developing fully rational idealisations.[140] If possible Twin Earth is a relativism case it is because the important group from whom a full rational idealisation is indexed turn out not to converge in the right way. And, as Smith himself acknowledges, this is a way for his rationalisms to turn out false – that is if we found for us (actually) that there was no desideratively convergent fully rational population of us (all of us, actually) then Smith's anti-Humean rationalist ethical theory would turn out to be false. How can it be objectionable to keep the possibility of error theory because of relativism open then?

As it happens I don't think it is objectionable to keep open the possibility of error theory. Diagnostically, I think what was objectionable about a metaphysical naturalism was failure to pay appropriate attention to folk

[140] This is roughly correct. In chapter 7 I will point out that metaphysical naturalism could allow historical accidents of the evolution of agents (humans for example) to make a difference to what properties the term 'right' refers to. This complicates matters but not in a way that bears on these considerations I think.

morality. Building an explication of folk morality into a reference fixing description seems to solve this problem.

I think that we can show that this is reasonable by briefly considering three different ways we can interpret the proposal that twin earth communities that are sufficiently psychologically similar to realise Smith's anti-Humean rationalism used in the role of a reference fixing theory. The first is just the one canvassed above. The remaining two (one which attempts to reject the one relevant community proposal, and another which accepts it but tries to maintain that we should still describe the twin earth case as a non-relativist one by folk moral standards) are the only ways I think we could even attempt to persist in insisting something is wrong from the our actual point of view (the one from which folk morality comes). Both rely on a kind of extreme parochialism that can't be justified. I doubt Smith would see either in a positive light but then examining them will show why I think this.

6.2.3 Counting communities in possible Twin Earth under the condition of ideal desiderative heterogeneity[141]

I think there are two counts and three options here. Either you count all the agents in a possible twin earth as members of the same community (because they are relevantly similar from the point of view of folk morality) or you count them as members of two communities (because they are relevantly dissimilar from the point of view of folk morality). Each count is then

[141] Desiderative heterogeneity is just the divergence of the relevant desires of the relevant fully rational population that, given Smith's rationalism, would were it to occur actually show Smith's theory is false (and if Smith's theory were also the correct explication of implicit folk morality it would thus show error theory is the case – it would show there are no moral truths)

evaluated. We have conceded that under the condition of psychological divergence framing these options metaphysical naturalism has to take the twin earth case as a case of relativism. That means for each of the two community-counting proposals metaphysical naturalism has to take the twin earth case as a case of relativism. As far as I can tell this gives us three options to consider. The first option (3(a)) is that we propose the community-count is 1 and allow that this proposal is endorsed by or fits with folk morality. This proposal is the one I closed the previous section discussing and is I think the only sensible option. The second option (3(b)) accepts that the community-count (of the number of relevant communities in the twin earth case) is 1 and asserts that taking this as an instance of relativism (as the metaphysical naturalist must) is not endorsed by or consistent with folk morality (and so metaphysical naturalism remains objectionable from the point of view of folk morality). Rather than relativism, the twin earth case is one where neither community instantiates 'right' as we understand it. The third option (3(c)) asserts that the community-count should be two **and** that this should not be taken to be an instance of relativism. Both claims are supported by folk morality and moral naturalism must deny both (and so metaphysical naturalism remains objectionable from the point of view of folk morality) Rather we know that one or other of the communities can instantiate 'right' as we understand it but not both. The only motivation I can see that could be offered for considering options 3(b) and 3(c) is the blunt intuition that the objectivity platitudes of folk morality will always give sufficient reason to reject the possibility of relativism. This intuition should not be held onto if we think it at all possible that folk morality could be inconsistent to the extent that it might fail to determine uniquely the referents of ethical terms like 'right'. Options 3(b) and 3(c) are untenable because one way or another they rely on this intuition. If we imagine this is not the case

we fall into the parochialism, I have argued above, we should at this point refuse to accept.

Option 3(a)

The members of communtiy1 and communty1*, in the possible Twin Earth case, are relevantly similar and both are included in the starting set of agents from which fully rational idealisations are generated.

This is just the option where a metaphysical naturalist can call the case a case of relativism but point out that the population of possible Twin Earth are simply internally incoherent, in the right way to lead to relativistic error theory. They, like us (actual 'us'), really are just one population and the differential reference for what should be seen as one term 'right' is in their circumstances a result of a relevant heterogeneous psychology.

I think there are only two styles of counter to this position that might allow a proponent of Smith's anti-Humean rationalism attempt make Smith's objection against metaphysical naturalism. They are option 3(b) and 3(c). Option 3(b) and 3(c) can be seen as attempts to make good on the claim that rather that talking past one another (as relativism requires) folk in twin earth cases have to be involved in real disputes about matters of fact to even be examples of folk who are using a term like our term 'right'. That is, the folk of twin earth cases must be such that one, the other or both communities are wrong about the facts concerning 'right' – that is 'right' as it is used in our mouths. The problem with this formulation is obvious I think. Smith's rationalist meta-ethics makes room for the possibility that 'right' as we use it fails to refer because as it happens 'right' as we use it allows relativism. In

what follows we will see that the antagonists of metaphysical naturalism will end up doing no better than this formulation of the 'problem' with twin earth cases and so fail.

Hereafter I will use C1 and C1* to refer to the two communities in possible Twin Earth with heterogeneous maximum desiderative unity properties.

Option 3(b)

C1 and C1* are not relevantly similar and neither populations members form a part of the starting set of agents from which fully rational idealisations are generated.

We have already argued that under the condition of ideal desiderative heterogeneity to the sort in play here we have a good reason for the metaphysical naturalist to accept that her view insists on relativism in this case. So to mount a version Smith's objection we have to suppose that folk theory find a way to reject this interpretation of possible Twin Earth.

One way this might be done is if some support can be found for the claim that the best interpretation of possible Twin Earth under conditions of maximum desiderative unity heterogeneity is that it is an example where no agent instantiates the right kind of psychology. Neither C1 nor C1* have a term like our term ''right'. The problem for this view is that we either claim that the assertion is grounded in a substantial reason or it is not but rather is simply motivated by the application of the objectivity platitudes. If the assertion is motivated substantially then the metaphysical naturalist deserves an account of why this substantial reason has not turned up in the explication of folk

morality in play, that is Smith's anti-Humean rationalism must explain why possible Twin Earth under the condition of maximum desiderative heterogeneity is really not an example of relativism. The appeal to the objectivity platitudes in this context, I think, is vain. We know that under option 3(a) we have an explanation, from the explication of folk theory, why folk theory should accept that the possible Twin Earth case under conditions of ideal desiderative heterogeneity *is* an example of relativism: the objectivity requirements of folk morality have not been meet. We also know why the members of community 1 and 1* are relevantly the same (their members can realise Smith rationalist theory cast in the role of a reference fixing description) – to continue objecting in the manner proposed to metaphysical naturalism requires that an explanation of why members of community 1 are different from members of community 1* *and* why this is not a difference that a metaphysical naturalist can make use of be given. Appealing to the objectivity platitudes of folk morality will not work since the metaphysical naturalist has a good account of why they are not met - Option 3(b) collapses to 3(a). If there is supposed to be another reason to distinguish the notions of 'right' in play between member of twin earths populace and our folk moral notion of 'right' we should simply demand to know what that reason is. Option 3(b) is unmotivated and so untenable.

Option 3(c)

Here we attempt to imagine the antagonist of the metaphysical naturalist asserting that both C1 and C1* are not relevantly similar, and that one populations members count as the starting set from which fully rational idealisations are generated and the other populations members do not.

If folk morality gives us reason to believe this then we can, as we did in 3(b), require an explanation both of what that reason is **and** why it is unavailable to metaphysical naturalism. 3(c), taken charitably is really much like 3(b) and either collapses to 3(a) or is untenable.

But really I think that 3(c) is much worse. The problem with it is just how brutally parochial it really is. After all which of community 1 or 1* is appropriately considered the better group from which to create a fully rational idealisation? And if the problem is rather not in the differences between the possible communities each to the other but with the possibility itself matters get worse. What else could be wrong with the possible people of the possible Twin Earth case relative to us that is not already in play in the case itself? Their lack of actuality? Less ridiculous but bad enough would be attempting to suggest that there is a difference between us and Twin Earth folk that cannot be captured by any reference fixing description. This idea would run the risk of failing to allow reductive naturalism as a possibility. I think that we should reject 3(c) as an option for Smith, it is doubtful he would welcome it.

6.3 Possible Twin Earth is no objection to metaphysical naturalism

Even without a theory of desiderative unity we can amply demonstrate that Smith has no secure objection to metaphysical naturalism, at least none of the kind found in TMP. That is to metaphysical naturalism with a plausible reference fixing description got from an explication of our implicit folk morality. If we allow that the populace of possible Twin Earth is relevantly psychologically homogenous then we have licence to look for an explanation

from Smith for how differential reference is possible given the role his rationalist theory is playing as a reference fixing description. As we saw, the very best we can do for Smith's objection under this condition is an indefinite deferral of his objection. Under the condition of relevant psychological heterogeneity things are much worse for Smith. The charge of attributing relativism in a possible Twin Earth right/right* case can be made against metaphysical naturalism but only at the cost of acknowledging that folk morality supplies the reason. And if we read relativism as error theory, the possible Twin Earth case is error theory due to a contradiction of the platitudes of objectivity that is permitted by the same source as supplies the objectivity platitude in the first place. The possible Twin Earth case can't be used against metaphysical naturalism, where metaphysical naturalism uses Smith's rationalism[142] as the source for a reference fixing description. Any intuitive pain found in this conclusion is, I argue, the pain of imagining that folk morality is incoherent, not of the inappropriateness of metaphysical naturalism for ethics

6.4 Summary

We can reiterate now the dilemma argument of Chapter 4. Smith must either accept that his theory collapses into a definitional network analysis of the kind that fits into Jackson's moral functionalist framework or adopt a natural kinds treatment of ethical terms. In particular we can motivate the use of the natural kinds approach for desiderative unity. We can also demonstrate that using the appropriately complex reference fixing description for 'right', where Smith's rationalist meta-ethical theory is used to fill in the parameters

[142] I suspect the same might be said even if we use other explications of implicit folk morality and the force of twin earth objections to metaphysical naturalism. We can't explore the issue here.

of the reference fixing description, Smith's objection to this approach fails. The choice then of adopting definitional network analysis and reduction or metaphysical naturalism remains in place.

Chapter 7 The Dilemma for Smith's anti-Humean rationalism, how to choose, and closing remarks

We begin this chapter by reminding the reader of the background to the thesis and then summing up what has been accomplished.

7.1 Background to the thesis and conclusions reached

Early in the thesis we saw that Smith rejects definitional network analysis reductive naturalism for ethics, where a definitional network analysis supposes that there is an explication of implicit folk moral concepts with sufficient information to interdefine all relevantly normative terms in a network and which will allow a Ramsey-Lewis-Carnap style reduction to be effected. The key feature is just that a definitional network analysis supposes that all the terms relevant to morality can be explicitly included in a network analysis (analysis because the relevant facts are in principle a priori) and collectively reduced to non-moral properties.

Smith also rejects metaphysical-but-not-definitional naturalism, which he defines as the view that the word 'right' refers to the property of acts that is causally responsible for our uses of the word 'right' (see chapters 4 and 5). He offers an alternative — a third way, as it were — that he calls a summary style network analysis and reduction. Smith's primary example of a

summary style analysis is the dispositional analysis of colour terms, an analysis that is explicitly non-reductive and circular and provides a usefully informative premise that figures in what we have been calling a narrow reductive argument. In the colour case, this yields a reduction of colours to particular surface reflectance properties. This reduction is narrow because the nature of the circular premise used in this context means that the physicalism secured by this reduction needs to be 'squared' with a broader physicalism. In the colour case this means a physicalist theory of 'looking red to normal perceivers under standard conditions' has to be supplied.

Smith thinks that the same kind of story should be told about ethics (various parts of this thesis have supplied the details of this story). He thinks this third way is distinct from, and preferable to, definitional and metaphysical reductive naturalism in the case of ethics, and to implement it he provides a two premise reductive argument to secure naturalism in ethics. The first conceptual premise, which relies on a summary style analysis of right, links right actions to desires of relevant populations of fully rational agents generally. The second substantive premise secures the details of the relevant circumstance-relative convergent desires of this fully rational population, details that effectively specify natural features of acts that make those acts right. The reduction then identifies right action, relative to a circumstance, with a property of acts — according to Smith the *same* property for all agents in the relevant circumstances.

What I have argued in this thesis is that Smith does *not* have a third way option in the case of ethics. Instead he must choose either definitional

naturalism or metaphysical naturalism or face his anti-Humean rationalist meta-ethics collapsing into non-reductive naturalism or non-naturalism, each of which is as bad as the other, with the latter, at least, explicitly rejected by Smith. In the remainder of this chapter I first highlight the important staging posts of the rather complex argument given in the earlier chapters against this third way, and I then conclude with some more speculative observations about the argument and its aftermath.

7.2 The argument

As we saw in chapter 2 Smith rejects definitional naturalism on the basis of what he calls the permutation problem. To formulate the problem and his solution to it Smith uses the analysis of colour resemblance relations and the dispositional analysis of colour and the two-premise argument for a narrow reduction of colours (of objects) to physical properties. But his arguments based on the colour case depend on there being analogies between the colour and ethics cases. I show that on the contrary there are important disanalogies between the cases that make these arguments ineffective. The colour case uses a posteriori premises in both the narrow and broader squaring reductive arguments when according to Smith the parallels in the ethics case are a priori. As I argue in chapters 2 and 4 the a priori nature of Smith's theory provides very good reason to include the analysis and reduction of rationality into the analysis and reduction of right and that would appear to effect a definitional network analysis reduction to natural properties. He also provides an inductive argument against definitional naturalism. This is

blocked because the evidence that the induction is based on is highly contentious. In particular, the failure to date of consensus on a reductive definitional naturalism is used by Jackson to give an account of the absence to date of a mature folk morality and its on going negotiation. What Smith takes as evidence of the failure of a reductive project Jackson takes as the necessarily vexed negotiation about which explication of which component of our implicit folk morality should play the role of explicitly fixing our moral concepts. Smith provides no reason for us to prefer his account of the evidence. This contestation of evidence makes an admittedly weak inductive argument even weaker.

As I showed in chapter 2, what the permutation argument was meant to give Smith was a reason to keep separate the analysis and reduction of right from the analysis and reduction of rational (what I have called the semantic gap), thereby preventing a collapse of his position into a kind of definitional naturalism. In the colour case there were various reasons – to do with the a posteriori nature of the identifications and theories in play – for the analysis and reduction of 'looking red' or more accurately states of colour sensation (as Lewis 1972 puts it) to play an indirect role in the reduction of colours in objects to physical properties. But the a priori and non-causal nature of Smith's anti-Humean rationalist meta-ethics makes reasons like those found in the colour case unavailable, or so it appears.

Having removed Smith's own arguments for a semantic gap between the analysis and reduction of right and the analysis and reduction of rational, in chapter 3 I consider where Smith could turn to find the kind of theory he

needs to maintain the gap and avoid the collapse to definitional naturalism. I argue that since Smith argues that rationality is a central determinant of which properties are the right making properties of acts and that the central component of rationality that allows it to perform the role it needs to for the purposes of Smiths rationalist ethics is the feature of desiderative unity then this is where we should look for an alternative justification of the semantic gap. In short, desiderative unity has the job of supplying truth makers for moral beliefs because of the effect maximum desiderative unity is supposed to have on the contents of the desires of fully rational versions of us. An agent's actually instantiating desiderative unity provides both the appropriate (and appropriately defeasible) attitude toward right making properties and the connection between actual agents and fully rational idealisations of those agents – the connection needed to make fully rational idealisations **relevant** to actual agents. Finally, reference fixing on actual desiderative unity allows the correct interpretation of counterfactuals involving the desires of rational versions of ourselves and also allows for Smith's putatively correct explication of our moral concepts to also possibly be false of us.

Smith provides a sketch of a theory of desiderative unity in TMP. It is the Rawlsian account of moral belief formation transplanted into an explanation of or analogy for desiderative rational change. I argue that a Rawlsian account of moral belief is no help in providing a theory of desiderative unity since facts about the way desiderative unity increases between sets of desires are what make moral beliefs true or false. Used more directly, the Rawlsian account of desiderative unity is a comprehensive failure. Normatively minimal accounts like the subsumption of specific desires under general

desires is demonstrably incomplete and richer accounts like the idea that desiderative unity tracks explanatory relationships between the contents of desires are question beggingly empty. The upshot is that Smith's theory as presented does not provide the reason we need to prevent it collapsing into definitional naturalism.

Chapter 4 turns to Smith's own squaring arguments to see if they are capable of staving off definitional naturalism. It turns out that Smith's own squaring argument is profoundly flawed because, as stated, it permits primitively normative natural properties that either just are non-natural properties in disguise or are as pathologically mysterious as non-natural properties. Examination of squaring arguments for the colours case offers no help for Smith, because of the disanalogies mentioned above. Quite simply, there are no background theories, philosophical or scientific, upon which an anti-Humean rationalist theory of ethics can rest a partial account of desiderative unity. We find that there is simply no theory of desiderative unity. Smith needs a reductive squaring argument that gives an account of desiderative unity. If he does not have such an account he must provide a plausible promissory note about how to find one. I close chapter 4 by suggesting that treating desiderative unity as a natural kind provides just the right balance of promise and ignorance that Smith's meta-ethics needs to solve the pending problem. For if desiderative unity considered a natural kind, without leaving open inappropriate possibilities for relativism, then it involves necessary a posteriori identities and so will effectively block, in a fashion similar to that found in the colour case, definitional naturalism.

Smith's initial set of problems included the ethics case being significantly different from the colour case. The proposed natural kind solution involves decreasing the degree of relevant difference between the cases.

Chapter 5 considers Smith's objections to natural kind treatments of ethical terms – metaphysical naturalism. If Smith's objections succeed then treating desiderative unity as a natural kind is not an option. There are two objections. The first takes up and extends Hare's "cannibals and missionaries" case objection to metaphysical naturalism. The second makes an analogy to the water case and in particular the possibility of twin earth water/water* cases. Both fail. The first objection depends on an effectively inaccurate portrayal of metaphysical naturalism (by focusing solely on the cause of term use and not on the constraints a reference fixing description place on the causes of natural kind term use) and compounds this by allowing Hare's use of a dictionary to supply the reference fixing description for 'good' to pass by unremarked. The second argument relies on an analogy to the twin earth water/water* case. Here the analogy is misguided. Smith's argument supposes that the appropriate reaction to a twin earth water/water* possibility, where two different kinds play the same reference fixing role for the term 'water', is indifference. This is an assumption and I suggest that it is only defensible to the extent that our interests in substance referred to by the term 'water' are exhausted by the features of our relationships to it that are captured in the reference fixing description of water. And when we look carefully at Smith's objection to metaphysical naturalism it turns out the possibility of relativism that a twin earth right/right* case allows is not the problem. Rather it is that metaphysical naturalism must treat the possible twin earth right/right* cases as relativistic when our folk morality (by way of

the objectivity platitudes according to Smith) disagrees with this interpretation of the case. However in the base case of a twin earth water/water* possibility the permissibility of a kind of 'water relativism' depends entirely on the nature of our interests in watery stuff and its nature. These sorts of facts about our interests should effectively figure in the reference fixing description of water. Similarly, if we use the correct reference fixing description in the right/right* case then the interests we have in the referents of the term right will figure in the reference fixing description for 'right'.

Chapter 6 aims to solve this underdescription problem by taking up Smith's meta-ethics wholesale and using it to provide the reference fixing description that a metaphysical naturalist can use. There are number of features or parameters of Smith's theory that need to be part of that description and using them has to be discussed carefully. The chapter uses the resulting picture to construct what I take to be an accurate and appropriately detailed versions of twin earth right/right* cases. The exhaustive considerations of chapter 6 show that either a twin earth right/right* case can be shown to be impossible, in concert with folk morality, or indeterminate. In either case, it is clear that Smith can adopt metaphysical naturalism.

Smith then can choose among the following – definitional naturalism (with an unresolved worry about what desiderative unity is), non-naturalism (or an arguably worse primitively normative natural property theory), or metaphysical naturalism. The last option is the most conducive to his rejection of definitional network analysis reductions. It has the benefit of

explaining why there is no theory of desiderative unity in a manner that makes a promise of the provision of such a theory in the future at least provisionally tenable. It does this in a way that is consistent with Smith's anti-Humean rationalist meta-ethics. The cost is that it makes Smith's theory hostage to empirical fortune in a way that may remain unpalatable to him, even though (so I have argued) it is not objectionable in the way he indicates in his own arguments.

7.3 Final observations

A number of questions naturally suggest themselves. First, is there any hint in Smith's later work, especially work that is not explicitly meta-ethical, as to how he might respond to the concerns the thesis has raised? Secondly, what are the prospects for the kind of metaphysical naturalism that we have identified as perhaps Smith's best hope for his meta-ethical project, given his background assumptions?

There is some reason to think that any answer to the first question will have to take on board Smith's recent 'action theory' turn. In 'A Constitutivist Theory of Reasons: Its Promise and Parts' (Smith 2013a), Smith formulates a theory of action and anti-Humean rationalist agency that in effect addresses the challenges to an anti-Humean rationalism about ethics, although its argument is formulated in a very different and (according to Smith) more foundational way. Here is what Smith has to say:

> "[Constitutivists] insist that Hume's characterization of an ideal agent is inadequate because he fails to see that certain final desires are constitutive of what it is to be an ideal agent. More precisely, they think that all ideal agents have certain dominant final desires in common, where these desires are dominant in the sense that their realization is a condition of the realization of any other desires that an ideal agent might happen to have. The final desires that are constitutive of being ideal therefore make it the case that certain things are finally good no matter which agent final goodness is indexed to. The Constitutivisits's [sic] account of the dominant final desires that are constitutive of being an ideal agent thus provide the much needed link between rational requirements and moral requirements that we've been looking for." (Smith: 2013a, pp. 19-20)

Smith summarises the way Constitutivism does this as follows:

> "In conjunction with the Inheritance Thesis [the thesis that "reasons for finally desiring something inherit their status as reasons from their being reasons that support the truth of the proposition that that thing is finally good" Smith 2013a, p. 18] and the standard story of action, it entails that there are certain final desires that everyone has reason to have, and so certain actions that everyone has reason to perform, and it further entails that agents with the requisite rational capacities are responsible for failing to have these dominant final desires and performing these actions when their failure to do so is a result of their

> failure to exercise these capacities, and it identifies these actions with those that are morally required. … The question that remains is how Constitutivists manage to deliver on this promise. (ibid, p. 20)

Space prevents detailed description of Smith's argument, parts of which I applaud (especially his objections to Parfitian accounts of agency and resulting accounts of moral requirements). But the new theory invites the question of whether or not it adds any significant advances to what is found in TMP. At first sight it might well look as if it does since Smith argues that general considerations of coherence are enough to generate dispositions to resolve doxastic and desiderative conflicts in an effectively rationalistic way. But on closer investigation we find that the kind of disposition Smith has in mind is explained as an effect of trying to maximize psychological coherence, with the goals of an evidence-responsive belief formation system in agents being dominant. That is, the kind of disposition Smith argues for is one that is focused on preserving the most ideal functioning of a system of beliefs when they conflict with what Smith calls idiosyncratic desires to believe contrary to evidence. The move to the idea that the parallel desire system, which has the goal of maximizing desire satisfaction, can also generate dominant dispositions to correct itself towards an ideal is only supported by analogy to the belief case. This weakness is compounded when the role of the notion of good-fixing kinds, which plays a crucial role in 'A Constitutivist Theory of Reasons', is made clear. What grounds the idea that there are any ideals at all for the doxastic or desiderative components of agents, taken independently or together, is the supposition that agents are like toasters or barometers. That is, like toasters and barometers they have a function and that entails ideal ways to achieve that function. But we have no theory of

what naturalizes the function of agential psychologies and no account of why we should believe that agents as good-fixing kinds idealise in the way Smith describes. And this feature of Smith's new agency-oriented theory and its problems seem like straight parallels to Smith's notion of desiderative unity in TMP and the problems for that notion we have identified in this thesis.[143]

We next turn to the question of the prospects for the kind of metaphysical naturalism that we have identified as perhaps Smith's best hope, a question that was left dangling at the end of chapter 6. The first thing to say is that Smith reconfigured as a metaphysical naturalist about desiderative unity will still fall under Jackson's moral functionalism framework. At least this is the case if we accept Jackson's views on how to deal with natural kind terms and a posteriori necessary identifications, or metaphysical necessity as it is sometimes called. Jackson, in FMtE, argues at length that metaphysical necessity, at least in the case of natural kinds understood as Jackson argues they should be understood, can be adequately accommodated using only

[143] The parallel runs deep. Smith says

> "The Dispositional Theory of Value in effect uses the fact that agent is a goodness-fixing kind in order to provide an analysis of a different concept of final goodness (Smith 1994, Smith 2010). According to Dispositional Theory, what it is for something to be finally good in this different sense, as indexed to some agent A, is for that thing to be the object of a final desire that A's maximally good counterpart has. There are thus two quite distinct concepts of goodness in play. The latter concept of goodness is the one internal to goodness-fixing kinds. The former is the one that we have defined in terms of the latter." (Smith: 2013a)

We should notice that the pattern of dependence on undefined and unreduced terms found in Smith's dispositional analysis and reduction of colour terms, and likewise in his analysis of rationality in terms of desiderative unity in TMP, is simply repeated in this passage. Final goods are by analysis the objects of the final desires (non-derivative desires) of rational idealisations of agents, where what determines the idealisation just is whatever it is that is good-fixing about the kind *agent*.

logical necessity and the two dimensional modal semantics. Should Smith wish to avoid this he could adopt Cornell realism – the position that at least in the ethics case Jackson is wrong about necessity, that it comes in two kinds (logical and metaphysical) and a natural kinds treatment of ethics perforce uses the metaphysical kind of necessity in its reductive identifications.[144] Jackson in FMtE engages in argument on this matter and Smith, unsurprisingly, does not. Though it is an option I think it is much less interesting than adopting a natural kinds treatment of desiderative unity in the first place. Treating desiderative unity as a natural kind leaves open what theory will explain how and why desiderative unity can perform the tasks a Smiths rationalist meta-ethics require of it. We can give a description of what it would take for actual agents to have a psychology capable of counting as 'oriented' or 'structured' in the way needed to support morality as Smith conceives it. And despite the complexities and attenuation of the armature linking current states of mind to ideal states of mind there remains a simple enough sense in which Smith's anti-Humean rationalist theory of the facts about right, should there be any, involve shared desires and tendencies to desire between our actual selves and idealisations of us. The important facts for Smith's theory, I have argued, are all determined by human psychology. And as we have seen the most important feature of all of this is whether or not we can find a theory to support the idea that human desiderative psychology has at least tendencies to form similar enough desires in relevantly similar circumstances across a wide enough variety of humans.

---

[144] Smith in verbal communication finds falling in with Jackson's framework undisturbing. More importantly, it is unclear whether joining those who think metaphysical necessity is characteristically different from logical necessity for whatever reason is something Smith would find appealing. I will simply leave further discussion to one side since my main interest here has been to evaluate how viable Smith's summary style analysis and two-premise reductive argument method is in the case of ethics. The question of whether to adopt or reject Cornell realism or views in a similar spirit is really a new topic. Jackson discusses the view in FMtE, pp. 144-146. For an account of Cornell realism, see Boyd 1988.

The metaphysical naturalist version of Smith's rationalism sets a kind of explanatory adequacy threshold for a theory of desiderative unity. It is sensitive to contingent or merely historical variation in the population relative to which Smith rationalist theory is evaluated for conceptual correctness and truth. The story about psychological tendencies needed to make this version of Smith's rationalism true do not involve anything obviously normative at all. For example, the evolutionary history of the human animal - either in virtue of individual or group selection causing psychological characteristics that have spread into populations – could provide enough psychological common ground for metaphysical naturalist version of Smith's rationalism. By way of a simple example of how this might be[145] we might expect a tendency for cooperative behaviour with identifiable cohorts to evolve in an environment that permits group competition. The idea is that we might be supplied with the right kind of psychological characteristics to tend towards some set of related desires for reasons that have to do with contingent, explicitly non-normative, evolutionary facts. Evolution only plays the role of supplying an account of why it is reasonable to expect either the wide distribution of particular types of desires in our primate or hominid line or the reasonable expectation of desires being negotiable in a way that tends towards a convergent overlap after sufficient negotiation. But of course any story might well do. Just so long as there are facts that fix actual and ideal psychology and relate them in the right kind of way we could realise a metaphysical naturalist version of Smith anti-Humean rationalist meta-ethics.

[145] What follows, of course, is a just so story. But it is no worse off for that since it neither claims the world is a way nor even that it could be a way. Rather it is more a prod to thought.

This kind of 'historically accidental' feature of metaphysical naturalist versions of Smith's anti-Humean rationalism might be thought of as part of the cost of ethical theories remaining appropriately defeasible. Certainly it is part of the cost of adopting metaphysical naturalism. What is accidental in an evolutionary sourcing of the right kind of psychological properties for Smith's rationalist ethical theory is not just that there is such a story to be had. Supposing that there were such a story to be had, the particulars will only be relevant to actual human or hominid evolutionary history. In effect, in this kind of story desiderative rationality is relative to populations and their histories.

Though I like making desiderative rationality a by-product of biology, a natural kinds treatment of desiderative unity does not require this approach. We simply hold in place the idea that there is a nature to aspects of psychology that we suppose underpins, explains, and plays the roles we require of it for an anti-Humean rationalist ethics and then search for facts about psychology that will settle whether things are as we suppose or not. What a theory of desiderative unity will then look like is open.

Whatever that theory might be, one thing I suspect we should be pessimistic about is retrieving anything more than an only *apparent* explanatory relation between the contents of desire sets. Put simply, I think we should be pessimistic about desiderative unity being about explanatory relations between desire contents. What a theory of explanation for relationships between the contents of beliefs might be is hard but perhaps not too hard. By

contrast, a theory of explanation that accounts of such relationships between the contents of desires looks altogether odd. The contents of beliefs, because they are explicitly about the way the world is taken to be, invite at least two dimensions of explanation. True beliefs about the causes of events or other things we have beliefs about explain the latter by way of a causal account of their existence. Relatedly, and perhaps separately, our beliefs can have as their contents theories of the way world is and those theories can be held to perform better or worse as explanations depending on the extent to which they embody or instantiate a variety of theoretical virtues. At this point I don't care what those virtues are, nor that this sketch of explanation is clearly incomplete. The important thing is that at least part of the notion of explanation for beliefs contents depends on how beliefs characteristically represent how the world ***is***. Desires characteristically do not do this.

Of course, with the imposition of means ends reasoning relative to some set of true beliefs, you might have the view that you should desire the means to your ends, and that your desires for ends explain your desires for means. But this is not the kind of conditioning on sets of desires that Smith's rationalism requires. 'Desires for ends' must, for Smith, be subject to rational scrutiny. You might then again think that the subsumption model of specific desires under more general desires will do the explanatory trick, but we know now that this cannot be the complete account of desiderative unity and so cannot serve to ground an explanatory model of desiderative rationality. And we know this for much the same reason as we know means ends theories are inadequate for the kind of rationalist ethics Smith has in mind. Desiderative unity increases between desire sets should be able sometimes to motivate

eliminating one equally general desire in favour of another while holding the more specific desires they subsume in place.

Perhaps all we can hope for is something like this: That if there is a contingent theory of human desiderative evolution that supports the view that we converge on the same desires (or tend to under ideal conditions), then desiderative explanation amounts to reflecting this common history in reasoning exchanges between people that aim for changes in someone's motivations.

This kind of approach to an explanatory theory of desiderative rationality is a distal one at best. The desiderative explanatory relations that desiderative unity tracks would not be a simple feature of the desires and their contents but rather would really be dependent on, and thus reflect, whatever theory of desiderative unity we ultimately come up with. And, as I have indicated above, that theory is importantly contingent. I would be content with such a view. But if you are not, then it is hard to see how you could give a contingently true story of the explanatory relations between the contents of desires without postulating that explanatory relations were just primitive features of relations of the contents of desires within sets of desires. And this looks like postulating primitive normativity – something we have shown Smith should avoid.

Bibliography

Bigelow, J., & Smith, M. (1997). How not to be muddled by a meddlesome muggletonian. *Australasian Journal of Philosophy, 75*(4), 511-527. doi:10.1080/00048409712348081

Boyd, R. (1988). How to be a Moral Realist. *Essays on Moral Realism* (pp. 181-288) Cornell University Press.

Brink, D. O. (1989). *Moral realism and the foundations of ethics* Cambridge University Press.

Brandt, R. B. (1979). *A theory of the good and the right*. Oxford : New York: Oxford : Clarendon Press ; New York : Oxford University Press 1979.

Copp, D. (1997). Belief, Reason, and Motivation: Michael Smith's "The Moral Problem". *Ethics, 108*(1), 33-54.

Denham, A. E. (2000). *Metaphor and moral experience*. Oxford ; New York: Oxford ; New York : Oxford University Press 2000.

Döring, S., & Andersen, L. (2009). Rationality, Convergence and Objectivity. Originally presented at Eberhard Karls Universität Tübingen April 06, 2009 Philosophisches Seminar. Retrieved from http://scholar.googleusercontent.com/scholar?q=cache:Na3Oex9yLbMJ:scholar.google.com/+%22Rationality,+Convergence+and+Objectivity%22&hl=en&as_sdt=0,5

Field, H. (1973), 'Theory Change and The Indeterminacy of Reference', *The Journal of Philosophy* 70(14): 462-481

Hare (Richard Mervyn), (1952), *The language of morals*. Oxford : Clarendon Press 1952.

Hesse, R.Michael Smith's Conception of Morality as an Instance of Moral Realism. Retrieved from http://scholar.googleusercontent.com/scholar?q=cache:kp-9opCXV6IJ:scholar.google.com/+%22Michael+Smith%E2%80%99s+Conception+of+Morality+as+an+Instance+of+Moral+Realism%22&hl=en&as_sdt=0,5

Holton, R. (1996). Reason, value and the muggletonians. *Australasian Journal of Philosophy, 74*(3), 484-487. doi:10.1080/00048409612347451

Jackson, F. (1998). *From metaphysics to ethics: a defence of conceptual analysis. Clarendon* Oxford.

(2004). Why We Need A- Intensions. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 118*(1-2), 257-77. doi:10.2307_4321466

(2010). *Language, names, and information*. Malden, MA: Malden, MA : Wiley-Blackwell 2010

Jackson, F., & Smith, M. (2006). Absolutist Moral Theories and Uncertainty. *The Journal of Philosophy, 103*(6), 267-283.

Kennett, J. (2003). *Agency and responsibility a common-sense moral psychology*. Oxford: Oxford : Clarendon 2003.

Kieran, S. (2007). *Reasons without rationalism*. Princeton, N.J.: Princeton, N.J. : Princeton University Press c2007.

Lewis, D. (1972). Psychophysical and Theoretical Identifications. *Australasian Journal of Philosophy*, *50*, 249-258.

(1997). Naming the colours. *Australasian Journal of Philosophy*, *75*(3), 325-342. doi:10.1080/00048409712347931

Marino, P. (2010). Moral rationalism and the normative status of desiderative coherence. *Journal of Moral Philosophy*, *7*(2), 227-252.

Mcfarland, D. M. (1998). Response- dependence without reduction? (Moral judgements). *Australasian Journal of Philosophy; Australas.J.Philos.*, *76*(3), 407-425.

Nolan, D. (2015). The A Posteriori Armchair. *Australasian Journal of Philosophy*, *93*(2), 211-231. doi:10.1080/00048402.2014.961165

Norman, D. (1996). *Justice and justification : reflective equilibrium in theory and practice*. Cambridge England] ; New York: Cambridge England ; New York : Cambridge University Press 1996.

Pettit, P., & Smith, M. (2006). External Reasons. In C. Macdonald, & G. Macdonald (Eds.), (pp. 142-170). Malden, MA: Malden, MA : Blackwell Pub. 2006.

Randel Koons, J. (2003). Why Response- Dependence Theories of Morality are False. *Ethical Theory and Moral Practice*, *6*(3), 275-294. doi:10.1023/A:1026090102604

Sayre-McCord, G. (1988). *Essays on moral realism*. Ithaca, N.Y.: Ithaca, N.Y. : Cornell University Press 1988.

(1997). The metaethical problem. *Ethics*, , 55-83.

Smith, M. (1994). *The moral problem*. Oxford, UK ; Cambridge, Mass. USA: Blackwell.

(1995). Reply to Ingmar Perrson's critical notice of The Moral Problem. *Theoria, LXI*, 159.

(1996a). Normative Reasons and Full Rationality: Reply to Swanton. *Analysis, 56*(3), 160-168.

(1996b). The Argument for Internalism: Reply to Miller. *Analysis, 56.3*, 175.

(1997). In Defense of "The Moral Problem": A Reply to Brink, Copp, and Sayre-McCord. *Ethics, 108*(1), 84-119.

(1998a). Ethics and the A Priori: A Modern Parable. *Philosophical Studies, 92*(1/2, A Priori Knowledge), 149-174.

(1998b). Response-dependence without reduction. *European Review of Philosophy, 3*, 85-85-108.

(1999). The Non-arbitrariness of Reasons: Reply to Lenman. *Utilitas, 11*

(2001). The Incoherence Argument: Reply to Schafer-Landau. *Analysis, 61*(3), 254-266.

(2002). Exploring the Implications of the Dispositional Theory of Value. *Noûs, 36*(, Supplement: Philosophical Issues, 12, Realism and Relativism), 329-347.

(2004a). The Structure of Orthonomy *. *Royal Institute of Philosophy Supplement, 55*, 165-193. doi:10.1017/S1358246100008675

(2004b). Instrumental Desires Instrumental Rationality. *Proceedings of the Aristotelian Society, Supplementary Volumes, 78*, 93-129.

(2004c). Internal Reasons. *Ethics and the a priori : selected essays on moral psychology and meta-ethics* (pp. 17-42). New York: New York : Cambridge University Press 2004.

(2006). Is that all there is? *The Journal of Ethics, 10*(1), 75-106. doi:10.1007/s10892-005-4591-9

(2007). Is there a nexus between reasons and rationality. In S. Tenenbaum, & I. NetLibrary (Eds.), *Moral psychology* (pp. 277-296). Amsterdam ; New York, NY: Amsterdam ; New York, NY : Rodopi 2007.

(2009a). Reasons With Rationalism After All. *Analysis, 69*(3), 521-530. doi:10.1093/analys/anp082

(2009b). The explanatory role of being rational. In D. Sobel, & S. Wall (Eds.), *Reasons for action* (pp. 58-80). Cambridge, UK ; New York: Cambridge, UK ; New York : Cambridge University Press 2009.

(2010). Beyond the error theory. In R. Joyce, & S. Kirchin (Eds.), *A world without values essays on John Mackie's moral error theory* (pp. 119-139). Dordrecht ; New York: Dordrecht ; New York : Springer c2010.

(2011a). Beyond beleif and desire: or, How to be orthonomous. In N. A. Vincent, I. v. d. Poel & J. v. d. Hoven (Eds.), *Moral responsibility beyond free will and determinism* . Dordrecht ; New York: Dordrecht ; New York : Springer c2011.

(2011b). Deontological Moral Obligations And Non-Welfarist Agent-Relative Values. *Ratio, 24*(4), 351-363. doi:10.1111/j.1467-9329.2011.00506.x

(2012a). Agents and Patients, or: What We Learn About Reasons for Action by Reflecting on Our Choices in Process-of-Thought Cases. *Proceedings of the Aristotelian Society, 112*, 309-331.

(2012b). Naturalism, absolutism, relativism. In S. Nuccetelli, & G. Seay (Eds.), *Ethical naturalism : current debates* (pp. 226-245). New York : Cambridge University Press 2012.

(2013a). A Constitutivist Theory of Reasons: Its Promise and Parts. *Law, Ethics, and Philosophy, 1*, 9-30.

(2013b). The Ideal of Orthonomous Action, or the How and Why of Buck-Passing. *Thinking about Reasons: Themes from the Philosophy of Jonathan Dancy*, , 50-75.

Smith, M., & Sayre-McCord, G. (2014). "Desires… and beliefs… of one's own.". In M. Vargas, & G. Yaffe (Eds.), *Rational and social agency; The philosopy of Michael Bratman* (pp. 294-343)