ATTRIBUTIVISM

by

CASEY SEAN ELLIOTT

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy
School of Philosophy, Theology and Religion
University of Birmingham
September 2019

ABSTRACT

This is a thesis in three parts. It concerns the normative capacity of *attributive goodness*. Specifically, it critically evaluates *Attributivism*, the theory that attributive goodness is *fundamentally normative*, or that the distribution of that property determines when, whether, and in what way agents ought to act.

The first third develops, refines and defends *Attributivism*. Doing so is, in part, a ground-clearing exercise. I distil that theory from the arguments of many other philosophers. In doing so I isolate and precisify its core commitments. I defend it from a number of objections.

The second third analyses *fundamental normativity*. I stipulate that a standard or property is fundamentally normative if its distribution alone can ground normativity. I argue that for anything to be so fundamentally normative, it must minimally meet two criteria. It must be both *authoritative* and *regulatory*.

The final third evaluates whether or not attributive goodness satisfies the established criteria, and so whether or not Attributivism is correct. I argue that in its canonical form it isn't. I then develop a revised view of Attributivism that *can* satisfy the criteria. I argue, however, that the revised view is unsatisfying in almost every respect. Attributivism, revised or otherwise, should be rejected.

ACKNOWLEDGMENTS

My gratitude to Dr. Jussi Suikkanen. I can't overstate how valuable it is to have a supervisor who knows everything.

My apologies to Jen Elliott.

TABLE OF CONTENTS

Man might be stripped of everything he ever believed about moral restraints; he would still, in his constitution, find impediments to acting quite at random. For there is in him what might be called a 'natural machinery for self-control'.

W.D. Falk, 1986, 187.

1 – INTRODUCING ATTRIBUTIVISM

Here are two claims often found convincing. One, goodness is a normative property – as a rule we ought to do things that are good, and not do things that are not good. Two, it is *not* the case that, as a rule, we ought to do what is *naturally good*, and to not do what is not so. Together these theses prompt the following conclusion: *natural goodness*, goodness understood in terms of the satisfaction of some natural standard, is not good in the normative sense.

A particular view to the contrary has recently come into its own. It claims that not only is natural goodness normative, but that normativity should be understood entirely in terms of natural goodness, and that, in fact, goodness has only one intelligible analysis – a natural, *attributive* analysis. I call this view *Attributivism* accordingly. Attributivism claims that 'good' is an *attributive* adjective: there is no such thing as to be *just* good, rather anything that is good is good *as* some kind of thing – a good *x*. It claims thus that what it is for a subject to be good is for that subject to meet or exceed the functional standards of its kind. A good *x* is one that does what an *x* is *supposed* to do, in the way that a clock is *supposed* to keep time, a tree *supposed* to grow towards the light, or wolves *supposed* to hunt in packs. It claims finally that this sort of goodness – *attributive goodness* henceforth – is the central focus of ethics, and can serve to ground normativity – how we and all things ought and ought not to act and to be.

Attributivism as an idea is nothing new. Elements of it can be traced back to Aristotle, and its modern face has been developing for decades. Only recently though has it begun to come together as a distinct theory. A host of prominent ethicists have based their philosophies around its principles. And unsurprisingly, it has just as recently been confronted with a host of criticisms. In many ways that dialogue has been a messy one.

Despite the centrality of attributivist principles in recent work, those principles have most often been used as a new way of approaching old theories – a new way of grounding the importance of traditional *human virtues*, or of reinvigorating Aristotelian ethics. In fact I think only one work – Judith Thomson's *Normativity* – can be said to have considered attributivist principles in depth as comprising their own unique theory. And perhaps because of that, both critics and defenders of Attributivism can be accused of almost universally missing the mark. They have failed to appreciate the details on which Attributivism really rests.

So I find myself in a strange position regarding Attributivism. On the one hand, I don't think Attributivism is correct. Nor however do I think any of its critics are. I *do*, meanwhile, believe in the prospects for a naturalistic reduction of the normative. And what's more I believe that Attributivism gets closer to a plausible such reduction than any other theory. My instincts suggest I should both defend Attributivism, and critique it. Fortunately, the best critiques come from the most thorough understanding of the target. So I'll do both. This thesis is dedicated to rejecting Attributivism. Along the way however I will bring together the scattered elements of the theory into a cohesive whole to present what I take to be the most thorough account of Attributivism presented to date.

I won't, in this introduction, go into specifics about any particular attributivist theories, nor their criticisms. The rest of this thesis will do so exhaustively. I will though dedicate a few, short sections to (§1) why Attributivism is worth talking about, (§2) my intended conclusion, (§3) the process through which I will reach it, and (§4) a little bit about my methodology.

§1.1 – ATTRIBUTIVIST INTUITIONS AND NATURAL NORMS

Why, out of the many theories we could be evaluating, does Attributivism deserve attention? Like any plausible theory, Attributivism has at least enough intuitive pull not to seem immediately implausible. I think this is evident in the following anecdote.

Ten years ago I planted three fruit trees in a remote location, and left them to grow under their own influence alone. I returned to them this year to see how they'd progressed. The first tree was a study in success. Its branches grew tall and wide, its limbs were sturdy and its roots tethered it firmly into the soil. Its canopy was thick with leaves and bursting with fruit. Its sturdy branches and succulent fruit in turn tempted flocks of birds to nest in its branches, to eat the fruit and spread its seeds far and wide. That tree has grown well, I thought, it has grown into a good tree.

The second tree was sickly. I found out later it had a vascular defect that prevented it from adequately distributing nutrients throughout its body. It was stunted, barely a third the size of the first; it leaned, its roots only tentatively grasping the soil beneath it, and it had only one branch, on which grew a handful of withered fruits. Even as I watched a small bird, driven off from the first tree by stronger rivals, alighted on the branch, ate one fruit, and died. A shame, I thought, this tree has come from a bad seed.

The third tree was perplexing. Rather than accord with the conventional standard, its roots grew in the air, and its branches in the ground. Come Spring, instead of bursting into bloom it had burst into flame. It didn't burn. Rather than bearing fruit, it bore seven notes of music. Being familiar with trees, I could tell that something wasn't right. I couldn't say the tree had grown well or poorly, exactly, but whatever the case something had gone wrong.

I don't think my intuitions on those three trees are dramatically out of line with the norm. We might quibble terminologically – rather than saying the first tree was a good tree, we might say it had grown well, or strong, or healthy. We will typically agree though that

those ways are *good* ways for the tree to be. Moreover, to recognise that fact we needn't know anything more than what's involved in the life of a tree. We know how a tree is *supposed* to grow, how it *ought* to grow, when we know what a tree *is.* To know one is to know the other. Here though it seems that what we've done is discover a world of *natural norms*. Good, bad, right, wrong, and ought are all here derived from an understanding of the natural world, and purely naturalistic principles. Why not think all of normativity proceeds the same way?

Attributivism offers to make sense of this naturalistic intuition. It also offers some much needed ground clearing. Early attributivists were those who felt the language of normativity has become beholden to religious or moral notions that are no longer appropriate to assume, yet have embedded themselves into the discourse. They thought Attributivism offered a chance to approach normativity from a culturally and historically neutral position. Attributivism also makes sense of the apparent fact that norms change drastically depending on the kind of the subject – what is appropriate behaviour for a lamb is defective in a lion, and what a virus ought to do is quite different from what a human ought to do. It accommodates the intuition that what a subject *is* has a lot to say about how it *ought* to act. All of these are good things to do. Thus Attributivism is worth taking a serious look at.

§1.2 – PENDING CONCLUSIONS

However deserving of attention Attributivism is, this thesis is ultimately a critical one. The conclusion I develop is that Attributivism fails as a theory of the normative, because attributive goodness is not a *fundamentally normative* property.

A fundamentally normative property is a property that amounts to what Christine Korsgaard calls a "source of normativity" (Korsgaard 1996, 18). Such a property would be one through which all true normative claims were ultimately *made* true, a property from which the truth of normative propositions flows, and in which justificatory regression – the process of trying to ground the normative capacity of some *x* in some further normative consideration *y* – ends. It would be a property in which, alone, without appealing to any other normative properties or concepts, all-things-considered and non-optional truths about what we ought and ought not to do would be grounded. It would have to make it the case that overall truths about what ought and ought not to be done exist. It would have to account for what David Copp calls "the unity of practical reason" (Copp 1997, 103).

For any property to be a *fundamentally normative* property is a tall order. Nevertheless that's what attributivists have in mind for the property of attributive goodness. Yet any property that is supposed to be fundamentally normative will have to exhibit certain qualities. Part of this thesis consists in an analysis of the minimum requirements for fundamental normativity. I argue that, minimally, a fundamentally normative property must exhibit *Authority*, a type of overriding influence over agents' motivations, and a minimal degree of *Regulation*, the capacity for guiding action through the ruling out and ruling in of certain actions. I will argue that attributive goodness is unable to embody these qualities simultaneously.

Attributivism argues that for any agent, that agent will belong to a kind, and belonging to that kind will determine what that agent ought and ought not to do. If one is human, one

ought to do only what is good by human standards. I'll argue that the problem for Attributivism is as follows. Kinds are malleable. Agents can belong to multiple kinds, and can even change the kinds they belong to. They can come to instantiate new kinds, and cease to instantiate old ones. As a result, there's no reason to believe that the goodness of any *specific* kind will have the sort of overriding influence over agents' motivations that is required for any property to be fundamentally normative. If one can control whether or not one belongs to a kind, one can control whether or not that kinds standards apply to, or have any influence over, one's actions. Thus, I'll argue, the goodness of specific kinds cannot determine what one ought or ought not to do.

I reach several conclusions in this thesis, regarding the normative capacity of attributive goodness. The first is, as I've said, that attributive goodness as it has so far been developed fails the Authority test, and so cannot be fundamentally normative. In the interest of conducting the most thorough examination of its capacities as possible however, I go on to develop a new version of Attributivism I call *Global Attributivism*. I argue that this version of Attributivism *is* able to meet the demands of Authority. However I also argue that in doing so, it gives up any meaningful degree of *Regulation*. I argue that the resulting theory is so desiccated, so devoid of regulatory content, that it satisfies none of the motivations for adopting Attributivism in the first place. I conclude that we have no reason to accept any version of Attributivism. Yet I remain hopeful that Attributivism's near successes can inform future attempts at a naturalistic reduction of normativity.

§1.3 – LOOKING FORWARD

Including this introduction and a short conclusion, this thesis is conducted across 8 chapters. Chapter 1 is now coming to a close.

Chapter 2 explains Attributivism. The task of that chapter is to set the stage, to sufficiently inform the reader of Attributivism's history, and to establish the background against which the successive dialogue takes place. I begin by explaining the distinction between predicative and attributive adjectives, and how Peter Geach, arguably the first modern attributivist, took that distinction to motivate the attributivist project. I explain what it is to be attributively good, and why Geach understood such to be the *only* intelligible understanding of the concept, and how he understood it to be normative. I move on to discuss the problems that predictably arose in the face of Geach's, namely the problems of *Identification* and *Motivation*, or how we are to identify a subject's attributive goodness, and how it is supposed to motivate agents. I discuss with additional input from Elizabeth Anscombe, perhaps the most influential of early attributivists, how attributivists attempt to solve those problems. In the process I show how the works of modern attributivists like Philippa Foot, Rosalind Hursthouse and Judith Thomson have developed from the preceding.

Chapter 3 considers the new objections that have arisen in the face of modern, refined attributivist theories. These objections primarily concern (i) the inappropriateness or obviously unacceptable conclusions that seem to follow from basing normativity on natural norms or (ii) the inability of attributivists to ascribe plausible functions to kinds without relying on implausibly normative teleological or evolutionary function analyses. With some help from attributivist sympathisers like Micah Lott, I develop responses to a host of such objections, and demonstrate that Attributivism can accommodate all objections made to date.

Chapter 4 breaks from Attributivism for a time, to discuss normativity in general. As the goal of the thesis is to assess the capacity of attributive goodness to ground normativity, it

is required that we understand just what it is that it's supposed to ground, and what qualities it must possess to do so. I first evaluate a number of currently influential theories as to what constitutes the essential 'hallmarks of normativity'. I reject in order (i) Reactive Attitude theories, (ii) Naïve Instrumentalism and (iii) Reasons Fundamentalism. I then explore and argue for the qualities of (i) Authority and (ii) Regulation as constituting the minimal qualities for conference upon any property of fundamentally normative status.

Chapter 5 returns to Attributivism and, relatively briefly, explores and explains how Attributivism is supposed to meet the demands of Authority and Regulation. To the demand of Authority I develop a set of interconnected theses. The Necessary-Aim thesis develops the idea that motivation towards attributive goodness is necessarily interrelated with motivation generally, and that motivation towards one's own attributive goodness to at least some degree is inescapable. The Motivational-Functions thesis develops the idea that non-kind-defining dispositions are essentially secondary to kind-defining dispositions, by virtue of the latter grounding the function of the former. To the demand of Regulation, I explore the work of Rosalind Hursthouse and Judith Thomson, and explain how they develop two related ways of understanding the regulatory capacity of Attributivism.

Chapter 6 marks the end of my development and defence of Attributivism. I begin by further fleshing out the relationship between general desires and kind-defining dispositions introduced in the previous chapter. I argue that the best way of understanding this relation of dependence is by seeing all desires as conditional desires, conditional for their fulfilment upon being realised in a way that promotes some attributive good of the agent. I argue then that this conditional relationship fails to deliver Authority, because the range of kinds an agent may instantiate is wide and malleable enough to allow for *any* desire to be fulfilled in a way that's compatible with promoting one's attributive good. I develop two scenarios with which to challenge Attributivism – (i) the case of multiple instantiated kinds, and (ii) the case of

replaceable kinds. I demonstrate how Attributivism fails in each case to deliver normativity. I respond to a number of objections, before concluding that Attributivism as so far developed cannot capture Authority, nor thus normativity.

Chapter 7 concludes my critique of Attributivism, by looking towards what we should do in the face of Attributivism's failure. I argue that while extant attributivist accounts have failed, it's reasonable to look at the possibility of modification. I develop what I think is the only plausibly normativity-grounding attributivist account – Global Attributivism, or GA. I argue that GA *does* capture Authority, and plausibly grounds some degree of normativity. I argue too that its Regulation capacity is so limited – even such as to be potentially non-existent – that we still have no reason to accept it as a plausible theory of the normative.

Chapter 8 briefly summarises the work that has been done, and concludes.

§1.4 – TWO COMMENTS

Before proceeding I'd like to quickly make clear two features of my methodology, and the sort of thinking that underlies my approach. I hope in doing so to head off any distraction caused by the sudden discovery of dialectical disagreements, and to avoid the confusion that results from unidentified differences in initial assumptions.

In *Being Realistic about Reasons*, T.M. Scanlon observes that the dialogue in metaethics has shifted from being predominantly about morality to being significantly about normativity (Scanlon 2014, 1). I think he's right, and that such a shift is an appropriate one. I'm very interested in ethics. I'm not at all interested in morality. Frankly I'm not even sure what 'morality' is at this point, other than that it seems to be the concerned with specific characteristics important in certain historical ideologies – honesty, charity, piety, humility, etc. Dialogue about morality, as far as I can tell, assumes that those characteristics are worthwhile, or assumes that any worthwhile metaethical theory must validate their place in our practices. At the very least, moral dialogue seems to assume a distinction between general reason for action – the domain of practical rationality or practical normativity – and *moral* reasons for actions, and assumes the latter always trumps the former. I disagree. I don't believe there is any such distinction, and if there is, the relationship flows the other way – there can be no disagreement between moral reasons and general reasons where moral reasons come out on top. If there is reason for us to act upon moral reasons, it's because they are supported by reasons generally.

In 'Modern Moral Philosophy', Elizabeth Anscombe maintained that for ethics to proceed, 'morality' must be jettisoned from our dialogue, and our psychology (Anscombe 2005c, 169). I agree. So I understand metaethics, and even ethics more broadly, to be the study of the normative, the study of practical rationality, the study of *ought,* and indeed the science of reasons – what reasons are, what reasons exist, and what we are to do in light of

them. I very much hope that the actions we have often called moral ones come out to be those that we *ought* to do, but I am open to the truth being different, and the truth is what I want to know.

Related to the above, I aim to approach this thesis with a minimum of underlying assumptions. I am largely anti-intuition. That isn't to say I make no use of or reference to intuitions at all. I'm not sure that's possible in this sort of dialogue. Nor do I entirely reject their usefulness. We all must *start* somewhere, and intuitions are the sensible place at which to do that. Likewise appealing to intuition may be the rational course when confronted with two seemingly equivalently meritorious yet contradictory positions. Yet an intuition should never be that which *decides* the merit of a position. The entire philosophical exercise, after all, is founded upon the fact that our intuitions can be mistaken. Thus an intuitive disagreement is, to my mind, of minimal philosophical interest. What interests me is to ask what the world would have to look like for one's intuitions to be true, to test to see if such is plausibly the case, and to revise one's intuitions if the answer is negative. As such I am open to radical revisions on issues of intuition – that some view is a challenge to commonly held intuitions is not an objection I find persuasive. Within this thesis then I attempt to limit my appeals to intuition to cases where facts about intuitions are already built into the existing dialogue.

I hope the rest of this thesis makes for an engaging read.

2 – THE ELEMENTS OF ATTRIBUTIVISM

No thesis develops in isolation. A thorough understanding of Attributivism, its merits and its flaws, must begin with an understanding of its origins, and the philosophies that have motivated its developments. This chapter will set the historical stage, so to speak. I will explain Attributivism – the what, the how and the why – by explaining where it's come from, and how, and why, it has developed.

In §2.1, *What Is Attributivism?*, I provide an initial formalization of Attributivism, with which to work.

In §2.2, *Predication and Attribution*, I review the distinction between predicative and attributive adjectives, and the properties they pick out. I explain the arguments for taking 'goodness' as an essentially attributive adjective, picking out the property of attributive goodness – a property that varies in its instantiation relative to features of the evaluated subjects. The discussion owes primarily to Peter Geach's seminal 1956 paper, "Good and Evil".

In §2.3, *Problems of Attribution*, I review two traditional problems for attributive goodness as a normative property. I call these the problems of **Identification** and **Motivation** respectively. §2.3.1 is dedicated to **Identification**; I observe that if Attributivism proposes that agents should do that which is in the good of their kind, it needs to be able to clearly identify just which behaviours actually are. The problem is that although it might seem clear what is involved in being, say, a chair or a knife, such that it's easy to recognise when something is a *bad* chair or a *good* knife, plausibly delineating what's involved in being *human*, and thus what's required of being a *good* human, is significantly more difficult. **Identification** asks for an answer to what makes a person a good or bad member of the *human kind*. It asks how

attributivists propose to identify, independently of any other ethical theory, a range of traits or behaviours essential in being *human*, such that a particularly *human* good can plausibly follow.

Even if it does end up plausible that there is some particularly human good, why should we care about it? §2.3.2 addresses that question and what I call the problem of **Motivation**. Apprehension of normative facts is supposed to be able to move us. Even if I accept that some *x* is in my good as a human however, why should I expect to be motivated in any particular way by apprehension of that fact, and how do attributivists explain that connection?

In §2.4, *Solving the Problems*, I review canonical approaches to resolving the problems reviewed in the previous section. Through §2.4.1-2.4.3, I explain the attributivist approach to **Motivation**, which argues both that one can't want an *x* without wanting a *good x*, and that one can't help but want what one needs. But what does one need? §2.4.4 explains the attributivist answer to that question and the problem of **Identification**, which relies on an analysis of kinds and lifeforms as categories which follow from, rather than precede, observable systems of behaviours. In a very real sense, attributivists believe one *is* what one *does*, and one *needs* that which facilitates the doing. What's essential in being human then is what is required to facilitate success both in those behaviours common to all organisms, and those behaviours which serve to distinguish the human lifeform from others.

In §2.5, I argue, prior to concluding, that through their responses we can see that Attributivism views the traditionally separate concerns above as facets of the same problem. The solution to one delivers the solution to the other.

§2.1 – WHAT IS ATTRIBUTIVISM?

Before anything else, let me clearly state what Attributivism is. This is the view that this chapter will develop and explain, and with which the majority of this thesis will deal.

Attributivism is the view that actions must be evaluated as actions taken by *kinds* of entities. An action is never purely a good or bad action, but only ever a good or bad *K* action, i.e., a good or bad *human* action. It is the view that those evaluations are normative. It is, in succinctly, the view that for any agent *x*, there will be some specific kind *K* to which that agent belongs, which determines whether an action Φ is good or bad for *x*, and so whether or not *x* ought or ought not to Φ. For example, if one belongs to the *human* kind, one ought to take actions that are good *qua* human, and not take actions that are bad *qua* human. If murder is a bad human action, bad *qua* human, and one *is* human, then one ought not to murder, or otherwise bring murder about. Formally,

Attributivism. (i) For any agent *x*, *x* belongs to some specific kind *K*. (ii) Goodness *qua K* is fundamentally normative for *x*. (iii) *x* ought to Φ *iff* Φing is good *qua K*.

Attributivism is the most common form of the view espoused by attributivists. Henceforth, whenever Attributivism is referenced, the above is the referent, unless explicitly said otherwise. The rest of this chapter will explain the history of Attributivism's development, its core commitments, and its apparent plausibility.

§2.2 – PREDICATION AND ATTRIBUTION

That goodness in some way plays an important role in our deliberations about what to do is uncontentious. If venturing into the mountains, one should have good hiking shoes, and walk in a good way, along good terrain, so as to avoid hazards. When building a house, one should buy good materials, and employ good contractors. One should read good books, and eat good food. One should *be* a good person; we should do what is good, and good things are what we should pursue.[1]

This discourse begins to seem almost tautological: if something is good it's the sort of thing we should do, and if it's the sort of thing we should do then it's a good thing; if it's a *bad* thing to do, then we shouldn't do it, and if we shouldn't do it, it mustn't be good.[2] Good is just the sort of thing we ought to do. What have we actually said though in these statements, so intuitively agreeable? Ought we to do something *because* it is good, or is it good *because* it's the sort of thing we ought to do? What is it that simultaneously makes something good, and connects that evaluation to reasons for action? Even if we agree that the idea of goodness is inextricable from what we ought to do, w*hat is it* to be good? In this section, I will unpack an answer to that question, provided by Peter Geach, to which the development of modern Attributivism can be traced. Although much of the following discussion will deal with linguistic philosophy, the ultimate upshot is a metaphysical one: in the metaphysics of normativity, the relation *x is good qua y* is more fundamental than the monadic property *x is good*.

In a renowned 1956 article "Good and Evil", Geach asserted that much of the confusion involved in the questions just presented stems from a tendency to talk about goodness as

[1] Similarly we might say we should refrain from and be motivated to avoid *bad* things; perhaps this is a meaningful distinction, but the scope of this project is limited primarily to an assessment of goodness, and won't be investigating its counterpart any further.
[2] Depending, one supposes on the existence of a better alternative.

something with independently apprehensible meaning, as if 'good' were a *predicative* adjective. A predicative adjective is something like the adjective 'square', where if a subject is square, its squareness holds regardless of any other substantive facts about the subject. If it is a square building, for example, it is a square *whatever-else* that building might be: a square piece of architecture, a square work of art, a square domicile. Geach, in contrast, argued that 'good' is always and only an attributive adjective, something like the adjective 'fast'. A subject cannot be fast *simpliciter*, but only relative to a comparison class. In this way we might have a 'fast snail': one which is fast as snails go, yet nevertheless a slow animal**.**

Geach's arguments and terminology set the stage for much in the way of significant philosophy.[3] In this section I will review Geach's proposal with three goals in mind. I aim (i) to clarify what Geach said, (ii) to explain why it was said and from where it developed, and (iii) to clarify and simplify certain terminological issues that will be relevant in the rest of this project.

To summarize Geach's view in the briefest possible manner: a predicative understanding of 'goodness' does not make sense. Rather, Geach argued that ascriptions of 'goodness' should be understood as uses of an *attributive* adjective. He has it, bluntly and unequivocally, that: "'good' … [is] always attributive" (Geach, 1956, 33). What does this mean?

Starting out, Geach first explains his predicative/attributive distinction by borrowing terminology from 'the grammars':

[3] What Philippa Foot once dubbed a "sadly neglected article" (Foot, 2001, 2) seems to have evolved into anything but. Recent appearances of the term 'attributive adjective' (helpfully provided by Rind and Tillinghast, 2008, 77) include Foot's own *Natural Goodness* (Foot, 2001, 2-3); Rosalind Hursthouse, *On Virtue Ethics* (Hursthouse, 1999, 195-196); Stephen Read, *Thinking about Logic* (Read, 1994, 176); and Judith Jarvis Thomson, 'The Right and the Good' (Thomson, 1997, 277).

> In a phrase an *A B* (*A* being an adjective and *B* being a noun) *A* is a (logically) predicative adjective if the predication is an *A B* splits up logically into a pair of predications is a *B* and is *A*; otherwise I shall say that *A* is a (logically) attributive adjective. (Ibid. 33)

Three things should be addressed here. Most important are the explanations of what exactly predicative and attributive adjectives are. I'll provide that explanation, with examples, in the following subsection. First however I'd like to clarify something about what Geach meant in claiming that 'good' is always attributive.

Geach's presentation takes a particularly grammatical approach. He speaks specifically in terms of adjectives and the logical structure of pieces of language. This might encourage in some the misconception that Geach is treating with goodness purely out of an interest in linguistic philosophy. Yet our interest in goodness is an interest in a *normative* phenomenon, in something that bears not just upon our linguistic practices, but in the exercise of practical rationality. We are interested not just in the classification of adjectives, but in what one certain adjective picks out, and how it bears upon practical deliberation. Geach *must* be speaking of something more substantive than a 'mere' linguistic device, and he is.

Miles Rind and Lauren Tillinghast claim that the correct way of understanding Geach here is as making the claim that "an adjective is attributive just in case it cannot be applied in a *truth-value-yielding fashion* unless combined with a noun" (Rind and Tillinghast, 2008, 77).[4] Here they are certainly correct. Geach is explicit in his understanding of the adjective 'good' as being primarily descriptive, and it is the adjective's relation to circumstances, the nature of *what* is being described, that should be seen as the ultimate motivation for his discussion.[5] Geach's interest in goodness is not primarily in the adjective 'good', but in the property of goodness picked out. His focus on the nature of the adjective 'good' is meant to illuminate

[4] Emphasis mine.
[5] "I totally reject [the] view that "good" has not a primarily descriptive force." Geach, 1956, 36.

particular features of the property it references – that the property manifests in a way that can only be accurately referenced in attributive adjectival terms. That is an important point to be clear on, so that anything that follows should make sense.

§2.2.1 – The Predicative/Attributive Distinction

Geach's theory is, succinctly, that (i) adjectives break down into two sorts, predicative and attributive, (ii) 'good' is essentially the latter, and (iii) the tendency to treat 'good' as a predicative has been the primary source of confusion in the history of moral philosophy.[6] Unfortunately Geach is not remarkably clear on the specifics of the attributive/predicative distinction. What we are given in "Good and Evil" as a technical definition is contained in the previously supplied quote. An adjective is *logically* attributive when it isn't *logically* predicative, and some adjective 'A' is *logically* predicative "if the predication 'is an *A B'* splits up logically into a pair of predications 'is a *B'* and 'is *A*'" (Geach 1956, 33).

This is hardly an exhaustive explication. Rind and Tillinghast (2008, 77-78) note that despite his distinction becoming part of the 'technical apparatus' of philosophers, Geach provides no satisfactory explanation of what an attributive adjective is, or of what it means for a predication to 'split up logically'. Fortunately Geach makes at least some attempt at aiding our comprehension.

Geach discusses specifically the adjectives 'big', 'small' and 'red'. The former two adjectives are, it is claimed, attributive with the last being predicative. The distinction is that the predicative adjective, 'red', is in some way intelligible, meaningful, or verifiable in the absence of any further information about the subject to which it is being applied, particularly regarding the subject's *kind* or *class*. In order to know that something is red, we don't need to

[6] "The apparent dissolution of [the concept of good] into a mass of ambiguities results from trying to assimilate it to the concepts expressed by ordinary predicative adjectives." (Geach 1956, 35)

know what *kind* of thing it is. The predicative adjective refers to a *substantive independent* property – the conditions for accurately applying a predicative adjective are independent of the noun to which it is applied. The same conditions for being red apply to cars, clouds, elephants and biscuits alike.

Attributive adjectives, in contrast, lack the aforementioned criteria. On the contrary, they are what Rind and Tillinghast and J.L. Austin have called 'substantive-dependent' (Rind and Tillinghast, 2008, 84) or 'substantive-hungry' (Austin, 1962, pg. 68) respectively. The conditions for applying an attributive adjective, along with what is described, vary with respect to the subject, and such adjectives are unintelligible or meaningless in the absence of some frame of reference. A subject being an *xy*, say a fast snail, will not imply that the same subject is an *xz*, a fast animal, even if the subject really is both a *y* and a *z*, and even if both *y* and *z* can be evaluated in terms of *x*.

So we can distil some idea of what it is for a predication to split up logically. A predication splits up logically into a pair of predications – and the associated adjective is thus a predicative one – when the predication 'is an *a b*' justifies the inference that the subject is a member of (at least) two independent classes, *b*s and *a*s. As an example – 'Floyd is a Canadian Surgeon' places Floyd as a member of the sets of Canadians, Surgeons, and Canadian Surgeons. Crucially, whether or not Floyd is Canadian depends in no way on his status as a Surgeon, nor on his membership in any other particular group at all, and his being Canadian intersects identically with his membership in any other groups. If Floyd is a pianist, he is a Canadian pianist; if Floyd is a fencer, he is a Canadian fencer, and so on.

On the other hand, a predication does *not* so split up – and the associated adjective is thus attributive – when such inferences are *not* justified – when the adjective-noun pairing *A-B* justifies only an inference to membership within the noun category *B*. The adjective in an attributive predication serves only to position the subject within a subset of the noun-category.

A *fast snail* is *not* a member within two independent categories, *snails* and *fast-things*, rather the adjective-noun pairing serves to restrict the subject's place within the noun-category alone. Crucially, the adjective does *not* interact with other, applicable nouns, and depends upon its noun in a way that predicative adjectives do not – that Floyd is a skilled Surgeon doesn't imply that Floyd is a skilled pianist, and the truth of his being skilled lives or dies with his being a Surgeon (or some other given class), in a way that being Canadian, or square, or red does not.[7] It's worth noting that, given this understanding, predicative and attributive adjectives correspond to what are known elsewhere as *intersective* and *subsective* adjectives.[8] In deference to Geach's role in the history of 'good' however, I will not be adopting those terms here.

To further illustrate these categories, let's return to Geach. Geach expounds on red cars, and big fleas. In the use of a predicative adjectival compound like 'red car', we will be able to identify two particular traits: substantive-independence in identification, and substantive-independence in application.

Being independently identifiable means that the characteristics predicated by an adjective are identifiable in ignorance of any additional substantive details about the subject being described. They (the characteristics described) imply and/or rely upon nothing about the subject other than the possession of those characteristics involved in the predication of the adjective, in this case 'being red'. In being independently applicable, the adjective will, if applicable in *any* substantive context, *remain* applicable howsoever we might further describe the subject.

We can see these features in Geach's red car example. Let us imagine that the vehicle is observed from a distance by two watchers, one colour-blind yet keenly sighted otherwise

[7] For more on dependence and presupposition, see Sven Danielsson's 'On Geach on Good', 2007.
[8] See, for examples, Heim and Kratzer (1998) on intersectives, Kennedy (2007) on non-intersectives, and Morzycki (2015), chapter 2, for an informative general review.

and the other the opposite. Our colour-sighted observer can tell clearly that the car is red even in complete ignorance of what type of object is so quickly approaching. His companion, in contrast, has no problem identifying the nature of the mechanical object speeding towards them, yet hasn't the faintest inkling of its colour.

Both of these observations can therefore be made in ignorance of the other – neither description depends on anything about the other. They can be combined to result in a guaranteed true compound description. In the case just outlined our watchers can ascertain that the object is a red car by sharing their observations, without needing to worry that the other's observation will invalidate their own. Crucially, it doesn't matter *what* the substantive part of a description of the subject turns out to be; identification and application of the predicative characteristics will remain valid regardless of any alternative substantive description of the object. Geach's red car is equally a red vehicle, a red machine, a red piece of art, a red feat of engineering, and a red object. The object's 'redness' appears entirely independent of any other descriptions that apply to the object, or any variations in its frame of reference.

To summarize - predicative adjectives describe independent, invariable characteristics. They describe attributes that will remain identifiable and evident no matter the context or terms in which the subject is presented.

Attributive examples behave quite differently. Take for example, as Geach does, some particular animal and the question of its size. Take too an analogue of our previous colour-perceptive but poorly sighted observer. Somehow blind to the *kind* of animal he is directed to observe, or worse, given no spatial context whatsoever, he's asked to judge its size. Is it large, or small? He is perplexed. Were the question regarding a predicate adjective, say whether or not the animal were red, or four-legged, it seems there would be no difficulty in providing an answer. Here, however, it seems meaningless, in the absence of some context or frame of

reference, to say whether the animal in question is large or small. This is essentially an issue of context; the answer cannot be determined independent of further substantive details.

Grant now, however, that *some* context is given: the animal in question is, unbeknownst to the observer, a flea, and is placed next to an elephant. Our observer gratefully concludes that the flea is small. The elephant is then removed, and the flea is placed in comparison to other fleas. Our observer is now faced with a conundrum: this flea, he has said, is small, but it turns out that in comparison to other fleas it is in fact very large, a giant amongst fleas. Is this flea then large or small? The observer has a problem if he tries to utilize these adjectives as one would predicative ones: if 'x is a small animal' broke down into 'x is an animal' and 'x is small', then it should follow that the description of 'small' should remain applicable no matter what other term might stand in place of 'animal'. Yet from the fact that the flea is a small *animal* it doesn't follow that it is a small *flea*. The truth of the matter is that this adjective is attributive. It can neither be asserted on set, substantive-independent criteria, nor can it be equally applied across frames of reference. Instead the veracity of its assertion is dependent upon further substantive considerations about the *kind* of object it is describing.

To summarize so far - unlike the substantive-independent predicative adjective, an attributive adjective depends upon certain contextual facts. Where a predicative adjective describes something evident in any context, an attributive describes some attribute that can only be observed and meaningfully understood relative to certain contexts.

Attributive adjectives seem generally to function in a comparative capacity: something is *large* or *small*, *fast* or *slow*, in comparison with other objects or relative to some standard derived from the relevant frame of reference. They are relational and this is a key sense in which they differ from their predicative counterparts. We might further our understanding by considering other relational and non-relational phenomena, such as the distinction in metaphysics between *intrinsic* and *extrinsic* or *relational* properties. According to Trenton

Merricks a 'mark' of intrinsic properties is their being "those properties that it is … possible that [an] object exemplify if that object and its parts … are the only objects that exist" (Merricks 1998, 61). This seems analogous to the characteristics described by predicative adjectives. We might furthermore in the same way claim that attributive adjectives resemble *extrinsic, relational* properties as being impossible for an object to possess if it were the only *thing* to have ever existed. There is no sense in qualifying some motion as 'fast' if it is the only motion to have ever existed, or some mass 'large' if it is the only mass. If there is no framework against which to qualify a phenomenon *as* something, then the phenomenon simply *is.*

To recap - Geach argued that 'good' is an attributive adjective. Like the terms 'fast' and 'successful', it describes a quality of a subject that is evident or meaningful only when the subject exists in some way relative to some context or standard. The characteristics involved are only evident in certain conditions – an object is fast when it traverses space more quickly than something else. Similarly then, the characteristics involved in something's being 'good' – that it ought to be pursued, say – intelligibly exist only when the object is compared with others in some relative context.

The discussion so far provides an intuitive understanding of Geach's distinction. Everything so far though has been primarily linguistic. The metaphysical upshot of the preceding is as follows. Goodness is fundamentally a relation between individuals and a context class. The property of being good is a property that can only be had relative to standards derived from one's kind. For any *x*, *x* is a *good x* only if it meets or exceeds some kind-specific standard. There is no such property as *goodness simpliciter*, but only of *attributive goodness*. It is that property with which the rest of this thesis will be concerned.

§2.2.2 – Origins and Motivation

Why did Geach consider it important to develop his theory, and what are we affirming when we assert his position? What does holding an attributive goodness theory commit us to? Geach's interests were not abstract; he believed the attributive nature of goodness implied consequences for the way we should live our lives. An understanding of his motivations will illuminate the role attributive goodness is meant to play.

Geach's work was motivated by his rejection of two, then contemporary, schools of thought, which he attributed to the Objectivists and the Oxford Moralists, in his terms. From the descriptions he provides of their respective theories, today their views would be recognised as versions of non-naturalism and non-cognitivism respectively.[9] Geach intended to present a naturalistic, cognitivist alternative to the dominant metaethical theories of his day. My aim now is to provide salient background and to explain Geach's motivation, by briefly reviewing his treatment of his rivals.

Geach's 'Objectivists' accepted two theses which Geach opposed. First, they believed that while there undeniably *are* attributive uses of 'good', instances of such are mere "trivial facts about the English language" and descend into a "complex tangle of ambiguities" (Geach 1956, 35). Second, they claim that if 'good' is to track a consistent, ethically meaningful property, there must be a predicative sense of 'good' that picks out to a single, consistent property. That property, they claimed, must moreover be a *non-natural* property, to avoid the threat of the so-called naturalistic fallacy[10]. Here Geach's Objectivists seem to agree with Moore, that if 'good' describes a natural property, then a further explanation is needed to link that property to moral obligation, and so the natural property can't, itself, be the source of that obligation. Indeed, the position Geach attributes to the Objectivists, which understands

[9] Well known contemporary proponents of these views include Schafer-Landau (2003) and Enoch (2011) for non-naturalism, and Gibbard (2003) or Blackburn (1998) for non-cognitivism.
[10] That is, the equivocation of goodness with merely coincident natural properties, about any of which it should seem sensible to ask 'is it good?'

goodness as a "simple and indefinable *non*-natural attribute" (ibid.) is almost identical to Moore's well known views.[11] Geach, then, should be taking as arguing against Moorean non-naturalism, and all other versions of non-naturalism whereby "good" references no natural property, nor subject of empirical sciences, nor anything existing within temporal boundaries (Moore 1903, §26).

The 'Oxford Moralists', meanwhile, agree with the Objectivists on their first point: attributive uses of 'good' are ethically inert. They claim that the primary force of 'good' in ethics must be unvarying, whereas the "infinitely varying descriptive force" of attributive adjectives leaves 'good' "merely ambiguous" (Geach 1956, 37). The Moralist's objection, which helps demonstrate the ambiguities that so concern the Objectivists, runs something like this: with the content described by attributive uses of 'good' varying from kind to kind, where the 'good' of a knife may express properties 'UVW' and the 'good' of some other object 'XYZ', there is no one property that 'good' picks out. So then there is nothing we can reliably and unambiguously pick out as being that which does what goodness is meant to do. Attributive goodness can't be *the* property which grounds and organizes ethics – there is no single property there to do the work.

Unlike the Objectivist however, the Oxford Moralist responds to the problem of ambiguity by denying that 'good' has any descriptive function at all. They claim instead that 'good' plays only a commendatory role.[12] This view, which Geach "totally reject[ed]" (idem. 36), is an early version of the familiar non-cognitivist position, whereby evaluative judgments

[11] § 15 of *Principia Ethica*, unambiguously – "the subject-matter of Ethics is … simple, indefinable, unanalysable". The assertion of non-naturalism may be drawn from the discussions in Moore's second chapter, or more directly, § 14, "Naturalism … offers no reason at all … for any ethical principle whatever".

[12] This isn't the only motivation the Moralist has to reject a descriptive interpretation; the alternative motivation however, to which Geach does respond, concerns the difficulty in moving from description to motivation, and I will review this difficulty later in this chapter.

about ‘goodness’ are taken to express desire-like attitudes rather than beliefs about the world.[13] We can assert thusly that Geach’s theory rejects what we now understand as non-cognitivism.

In the material just reviewed we can discern three principles which Geach was motivated to oppose:

1: Attributive uses of ‘good’ are essentially ambiguous and so ethically meaningless.

2: To avoid ambiguity, ‘good’ must describe a monadic, non-natural property.

3: To avoid ambiguity, ‘good’ has no ethically meaningful descriptive function.

Geach’s response to these claims was two-fold. First - as (2) and (3) were both motivated by accepting (1), Geach hoped to reject these claims by denying the assertion that attributive uses of ‘good’ are hopelessly ambiguous. Second - he objected to (2) and (3) directly. Geach’s position was not merely that attributive ‘goodness’ is intelligible and relevant in ethics, but that non-naturalist and non-cognitivist theories are not, and so that an attributive analysis of ‘good’ is the only intelligible analysis.

The rejection of non-naturalism and non-cognitivism is less essential to Geach’s position than his defence of attributive ‘good’ as ethically unambiguous, and so his objections to these views need only be briefly summarised. Of non-naturalism, Geach had this to say:

> Nobody has ever given a coherent and understandable account of what it is for an attribute to be non-natural [and so, non-naturalism] is only the pretence of a way out of the Naturalistic Fallacy: [giving no] account of how “good” differs in its logic from other terms, but only [darkening] counsel by words without knowledge (Geach 1956, 65-66).

[13] For modern non-cognitivist descriptions of this sort, see: Blackburn, *Essays in Quasi-Realism* and *Spreading the Word* (1993, 1998), Gibbard, *Wise Choices, Apt Feelings* and *Thinking How to Live* (1990, 2003), and Schroeder, *Non-Cognitivism in Ethics* (2010).

Poetically put, but in essence Geach views non-naturalism as aiming to avoid the threat of Moore's naturalistic fallacy, which he sees as looming over the assimilation of good to ordinary predicative adjectives. He denies that non-naturalism can accomplish this goal. Non-naturalism is, he claims, devoid of actual content and unable to provide an explanation of how a non-natural attribute should differ from ordinary predicative adjectives in the logic of its application.[14]

On non-cognitivism, Geach's criticisms are less thorough, and his rejection of the non-cognitivist relies primarily on the strength of his positive arguments against the ambiguity of attributives. Where he does criticise non-cognitivists he does so by rejecting what he takes to be their core assertion. Geach attributes to the early non-cognitivism the following view: in order to explain the role which 'good' plays, loosely a sort of action motivating function, the primary role of 'good' must be commendatory, not descriptive. To refer to something as 'good' gives voice to one's desire-like attitudes towards it, rather than describing any attribute. In this, Geach seems to have in mind something like an *expressivist* form of non-cognitivism. Further, he attributes to the belief that where 'good' *seems* descriptive, i.e., when the speaker isn't meaning to recommend the object personally, it is only describing what some group of judges would commend who *were* in the business of recommending such objects.[15] While Geach targets his expressivist analogue specifically, since this view is presented as

[14] Modern accounts of non-naturalism and analogous theories are less clearly motivated by the sort of Humean or Moorean motivations Geach has in mind. For modern advocates of such, see: Enoch 2011, Parfit 2011, Shafer-Landau 2003, Wedgwood 2007, and Huemer 2000 and 2006. Although these authors approach non-naturalism from a different direction than that which Geach found so unpromising, it's nevertheless plausible that his objections remain relevant. Many of the arguments put forth recently are either abductive (Enoch) or eliminative (Parfit, Huemer) in nature, or where more of a positive account is attempted (Wedgwood, Shafer-Landau, Huemer) it isn't immediately clear how they account for a difference from ordinary predicative adjectives in the logical application of their concepts. It may be that many of these theories do not *intend* to argue for a predicative 'good', but their approaches still plausibly entail one or the other sort of scenario that Geach found concerning, and as these authors (with the possible exception of Shafer-Landau) still rely on 'good' referencing the same sort of non-natural characteristics Geach so clearly opposed, his position seems still opposed to modern non-naturalism.

[15] Where this is supposed to extend an expressivist-like position, Geach is probably being unfair - the idea of what some panel of judges *would* recommend in the right situation seems to imply a sort of response-dependency to which expressivists might justifiably object. Geach seems to be running together an analogue of modern expressivism with general *speech-act theories*, where the meaning of 'good' is in what it is used to do. This is no necessary conjunction however – See Schroeder 2010, 74-76.

ideologically opposite to his own view – that 'good' *is* in fact descriptive and representative of some disposition-independent properties – we can take his pseudo-expressivism as representative of non-cognitivism in general.

Geach denies central elements of non-cognitivism – that good is not descriptive, and that judgments about something's goodness must accompany some sort of active disposition. He insists instead that one can sensibly judge of some object that it is a good example of its kind without holding any disposition towards it whatsoever, and without needing to reference any other existent group bearing any such disposition. According to Geach, when we say of something that it is "a good eye or a good stomach, [the] remark has a very clear descriptive force and has no reference to any panel of eye or stomach fanciers" (Geach 1956, 37).[16] For Geach, the primary force of 'good' is not to influence action in the way his non-cognitivists intended. Rather, when we say of something that it is 'good', as in a good toaster or good chair, we say something like that it is 'successful' or 'effective' in the function it performs, and this is an entirely sensible use of 'good' without any recourse to any dispositions, existent or imagined.

Of course modern non-cognitivism has made significant advances since Geach's target example. Still, however much more sophisticated modern non-cognitivist claims might be, the central point remains that Geach intends that his theory allows for ethical evaluations while referencing no mind-dependent phenomena. His position is firmly a cognitivist one.[17]

Where does this leave us? Even if we accept Geach's objections to non-naturalism and non-cognitivism, the charge of ambiguity still looms over his own proposal. It remains then

[16] This has become something of a standard argument – Charles Pigden (Pigden 1990, 131-133) makes use of nuclear-tipped ICBMs, that we may consider to be perfectly good ICBMs without suggesting that anyone should ever recommend one. Similarly might use the example of some virus, terribly deadly to humans – we can sensibly recognize, it seems, of the virus that it is a very *good* virus, without in any way suggesting that it is something desirable, or desired by anyone. See also Michael Smith's 'On Normativity' (Smith 2010, 719).

[17] It is also worth noting here, as evidence of Geach's cognitivism, that his legacy includes the tellingly named *Frege-Geach Problem*, a long-standing obstacle for non-cognitivist theories that remains relevant today (see Geach 1964 and Skorupski 2012).

for Geach to offer a positive account of how attributive 'goodness' should answer the concerns of its opponents. Here then is a review of Geach's response to the charge of ambiguity, and with it the crux of his positive argument against his opponents.

Geach doesn't deny that attributive 'goodness' describes different characteristics of different subjects. What he *does* deny is that these differences indicate that what attributive 'goodness' refers to is ambiguous, that it describes some loose, unrelated set of properties with no common feature, thus incapable of grounding some distinguishing feature in ethics.

"It is mere prejudice to think that either all things called 'good' must satisfy some one condition, or the term 'good' is hopelessly ambiguous" (Ibid 35). Geach is terrifically clear in his dismissal of such concerns. Yes, the term 'good' describes varying properties – 'UVW' of a knife and 'XYZ' of a tree. So what? That the properties described by 'good' vary depending on the subject doesn't imply that those descriptions don't, in each case, reflect something in common about each subject. Although what counts as good may vary by subject, how each subject's good is determined may owe to a common process. So in every case, what it means for *x* to be *good* is for *x* to have undergone that process, even if the result of the process varies with variations in input. Geach offers a mathematical analogy by way of clarification.

Attributive uses of 'good' vary in the content they express, true, but so too do the numerals '4', '9' and '16'. Each describes a different object, yet nevertheless they can all be given a common description: they are all 'squared numbers', numbers that are produced by squaring a smaller number. Now what's involved in one's being a 'squared number' appears, superficially, to vary in the case of each subject. When we say that 4 is a squared number, we mean it is the double of 2. That 9 is a squared number describes it as the treble of 3. This hardly means though that what it is to be a 'squared number' is ambiguous, that the description just sometimes means double of and sometimes treble of without rhyme or reason. On the contrary, given a certain number we know exactly how to determine whether or not it's squared,

what specifically that describes, and why it is so. That's because being squared is determined through the application of a consistent, underlying mathematical principle – a definite process, formula or standard through which varying inputs produce varying, yet determinate and predictable, outputs. So too, says Geach, with being good.

Just as we say of some number when we call it a 'squared number' that it adheres to some specific mathematical principle, so too when we call something a 'good', we say that it is adhering to, or fulfilling the demands of, some underlying process that determines what it is for something to be good, and so how it ought to be. 'Good' may attach to varying descriptions, but in each case those descriptions are representative of the fulfilment of set demands. Given some object x, what is required of the x to be 'good', and what will be described of the x insofar as it *is* good, is determinate.[18][19]

Undeniably Geach is unclear about what exactly this specific underlying standard *is*, what particular feature of ethics it should account for, and how. It's one thing to talk about mathematical principles, but what do we refer to with talk about hypothetical underlying standards of acting and being? Geach provides some vague guidance here. He says we can know what the 'good' of an hygrometer is once we know what it is an hygrometer is *for* – the function that, when performed, entitles us to call something an hygrometer (Geach 1956, 38).

[18] Another way of putting Geach's view is to say that although being good implies different things for different subjects, the view still respects compositionality. Some ambiguous words don't – 'bank' and 'bat' are ambiguous words, and *don't* respect compositionality; what is meant whenever those words are used is determined by context alone, there is no connection between their multiple extensions. Not so with 'good'. Learning how to determine goodness in one context leaves us able to determine goodness in every context – once we know how to determine what counts as a good knife, we know the sort of process for determining what counts as a good toaster, tree, or human. The extensions of the ambiguous word 'good' *are* connected by a common process, and don't vary arbitrarily with context.

[19] Gilbert Harman, in *Moral Relativism Explained* (Harman 2012) has recently offered another framework which might help clarify the sort of thought Geach is expressing, via comparison of morality with sport. Although there might be a variety of ways in which football, as an example, might be played, every valid description of a sport which answers to 'football' will adhere to an underlying structure of rules, referenced by every valid example of the sport. In so much as each sport is accurately described as 'football' it is being described as fulfilling set criteria, even while differing in instantiation,. While Harman intends to further the cause of moral relativism, his argument seems well suited to Geach's perspective: Harman agrees that relativistic descriptions of moral conditions might nevertheless be consistent with the existence of certain *moral universals*. While this alone doesn't guarantee that these universals are *important*, there's equally nothing to demand otherwise.

Similarly then, what is good for a human being is in some way related to human *function*, what it is that humans *do* in virtue of being *human* (Ibid, 40). Geach's underlying standard, then, has something to do with what is *required* of any *x* in the process of *being* an x. What is it to be good is not ambiguous, even if the property is properly referenced by an attributive adjective: being a good *x* is systematically linked to what it is to *be* an *x*.

There are undeniable difficulties that arise here: not the least of which involve whether or not there *are* clear standards of *being* we can derive for any given subject, and whether any relevant kind of standards exist at all. These difficulties will be addressed more thoroughly in part 2.4 of this chapter. For now, however, to summarise *this* segment: Geach was motivated to develop his theory in opposition to predominantly non-naturalist and non-cognitivist theories of his time. Geach denied the core claims of rival positions, and a significant objection to his own project – the objection that the attributive 'good' is too ambiguous to represent anything coherent. In asserting a Geachean position, we are asserting that the term 'good' describes a variable set of natural conditions *in terms of their fulfilment of* the determinate application of some underlying standard, derived in turn from the natural characteristics required of some 'x' to exist as an example of its kind.

§2.2.3 – Recap

Let me then briefly summarize what has been discussed in this section. According to Geach, 'Good' is an attributive adjective, which describes a determinate, natural and mind-independent condition or set of conditions, which are derived through the application of a consistent standard. Attributive adjectives are contrasted with predicative adjectives. Examples of predicative adjectives are 'red' or 'square' or 'composed of 128 parts'. These descriptions describe the sort of properties objects might possess regardless of the object's kind, or of external relationships. A predicative adjective thus describes a definite set of invariable

characteristics that persist unaffected by any further variable details of the subject, and remain identifiable regardless of the subject's context or classification. Attributive adjectives in contrast describe the opposite. Examples of attributive adjectives are 'fast' or 'bright'. These adjectives describe relational properties that are held only relative to some other standard or comparison class. A subject will possess the property described by an attributive adjective only in virtue of the fact that subject exists relative to certain contexts.

According to Geach, 'good' describes a property or characteristic that is evidenced by a subject relative to certain contexts. As examples go, one might be a good archer without being a good blacksmith; a good book yet simultaneously poor building material.

If we follow Geach, 'good' is a naturalistic, cognitivist and determinate concept. It is a description of something as adhering to the requirements of a determinate, if relative, standard. 'Good' is attributive, the characteristics entailed by its description are evident only in certain contexts. This standard, then, entails or is representative of some natural context, some class or kind that an object belongs to, which requires specific behaviour of the subject if the subject is to succeed as an example of such a kind.

§2.3 – PROBLEMS OF ATTRIBUTION

The previous section introduced the distinction between predicative and attributive adjectives, as well as Geach's suggestion that 'good' is an instance of the latter. He argued that the property that 'good' picks out – the property of goodness or of *being good* – is one that varies with the kind of the subject being evaluated, and describes them relative to standards derived from that kind. Geach offered compelling arguments as to why we shouldn't take the fact that this *attributive goodness* varies from kind to kind as a mark against it. However, to this point, Geach's attributive good has been presented as a primarily classificatory property – it positions a subject within a certain context. Yet goodness is supposed to be a *normative* property, one that tells us what to do, and which has implications for our actions. Geach's positive account there though is, self-professedly, vague.[20] Two problems arise from that vagueness.

Here is the first. Before we can determine what an *x* ought to do, we need to be able to tell what it is to be a good *x*. Geach though hasn't actually told us *how* to determine what goodness *qua x* consists in, but only that it must involve fulfilling the standards of an *x*. We might however be sceptical that there *are* any such standards, or any way of identifying them. The joint question of whether it's plausible to believe in such standards, and how to identify what they are, is the problem of *Identification.*

Here is the second problem. Goodness isn't supposed to be just a property by which subjects might be classified. It's supposed to be a property one *ought* to obtain. Whereas one might be indifferent about being fast, or tall, one should *care* about being good, or otherwise be *motivated* to be so. An explanation as to why goodness differs in that respect from other attributive properties must be provided. That is the problem of *Motivation*. This section will review those two problems, in turn.

[20] "I shall not here attempt to explicate the descriptive force of 'good (bad) human action'" (Geach 1956, 40)

§2.3.1 – Identification

In this section I address the problem of identification. I lay out the history of that problem, i.e., why we might be sceptical about the existence of natural kinds standards, and/or the prospect of identifying them. I argue that it is plausible that natural kind standards do exist, but that the problem of how to identify them remains a significant one.

I'll start with a broad claim. Being a toaster involves toasting bread. Similarly, being a map involves depicting geography. Being a match involves making flame. Being a knife involves cutting, a ship involves sailing, etc. Doing these things, or otherwise being *aimed* at doing so, is what it means to *be* such an object. Doing them *well* is what it means to be a *good* one.

Geach's theory relies upon that sort of idea. He believed that being a member of a kind involves being subject to certain standards. If something fails to uphold those standards, then it is an inadequate example of its kind. This seems intuitive enough when we're given simple examples like those just mentioned – a toaster is only a toaster insofar as it is involved in or aimed at the toasting of bread; to do such is its *function*, it's what it is *for*.[21]

Toasters though, and the other objects I've mentioned, are artefacts. When it comes to artefacts, it's a simple matter to determine their essential characteristics – those that are required for being those *kinds* of objects – for such requirements are imposed upon them. They have what Larry Wright (1973, 142) called *conscious function*, the product of design or conscious effort. What happens when we consider more complicated subjects? Wright

[21] The term '*function'* is contentious: what it is explicitly for z to be the function of x is not immediately clear. Early on in function literature, Wright offered the following definition: "the function of X is Z means (a) X is there because it does Z, (b) Z is a consequence (or result) of X's being there" (1973, 161). Here is an example of this: the claim that the function of the human heart is to circulate blood means that the heart has come about because it circulates blood, and said circulation is a product of its being. For many reasons Wright's definition is not widely accepted today. It is an example of what is sometimes called an 'historic' function theory, and modern proponents of that style of theory have much more to say on the relevant processes of selection. However many function theorists – including attributivists it will turn out – eschew historical based perspectives entirely. Much more will be said on the topic throughout this thesis – for now let's accept that 'function' here means vaguely 'that which some *x* does or is aimed at doing/disposed to do insofar as and in virtue of which it is a member of kind *y*'.

contrasted his *conscious function* with *natural function,* and the latter class is a dubious one – the task of identifying a function in self-determining or naturally occurring objects is less simple. Still, it seems we can with some plausibility say of a tree that, to be a good tree, to meet the requirements of its kind, it must *grow*, it must have strong roots, take in nutrients, mind the seasons. If it fails in these requirements, it is a poor tree.[22] As complexity develops, however, and as the capacity for self-determination grows, the difficulty of identifying a meaningful standard of being seems to increase in turn. What demands apply to an object that can determine its own course? What, crucially, is the proper function of a *human being*?

The question of identification splits up into two sub-questions. The first question asks, is there any sense *at all* to understanding individuals as fitting into categories or *kinds* which in turn impose standards upon their behaviour? The second question asks, in the event that there are such standards, are they determinate, how are they determined, and plainly, what are they? Let's consider the first question.

The attributive position relies on there being some generalized or objective standard that any given subject might be subject to, that arises in virtue of what the subject *is*. Certain standards of behaviour are imposed upon a subject by virtue of its *being* some kind of thing. We judge the object as more or less successful as an instance of its kind by judging how well it adheres to those behaviours. And this, of course, is exactly the question – are there any such natural essences, any objective, standard-implying categories into which subjects should be – non-arbitrarily – understood as fitting?

A natural way to frame the discussion of Geach's position is to consider the idea of the *essences* or *natures* of objects. Very broadly speaking, the essence of an *x* is thought to consist in (i) those properties or characteristics that *x* possesses in virtue of which it can be analysed

[22] See Hursthouse 1999, 198: "a tree … should have certain sorts of roots [that] help it survive, by keeping it anchored and taking up nourishment … its leaves should curl when it is short of water … it should produce [good] seeds at a certain time of the year…" etc.

as a certain *kind* of thing, (ii) are *necessary* for such an analysis, and (iii) are manifested necessarily by *x* insofar as it *is* a member of such a category. Being *human*, for example, is thought to be in the *essence* of being Socrates, and being *unmarried* essential to being a bachelor.[23] These terms invoke a familiar issue however – can we consistently, coherently, and in a principled way, analyse individuals as belonging to specific categories? If so, how? Can we conceive of an essential nature for something like a human being, such that we can say that there are certain traits one *must* embody to *be* human?

Certainly we might. One way of reading Wittgenstein would have us understand the *essence* of a kind as a sort of linguistically created category of classification.[24] If categories are merely conceptual classificatory constructs, we might easily *say*, i.e. *decide*, that certain traits need to be evidenced by something if it is to *count* as human. Whether or not some particular subject *counts* as human however, under a practical conceptual framework, is far from the sort of criteria Geach requires.

The Wittgensteinian criteria involves when and how a subject meets some *externally* imposed standard, and such a standard might say more about the imposing agents and their concepts than about any supposed subject. There is nothing immediately apparent in the process of linguistic classification that *demands* anything of the subject itself. Geach needs to assert that the relevant standard bears meaningfully on the *subject*, not merely that it meets someone else's, possibly arbitrary, criteria. The question Geach needs resolved then is not merely whether or not essences exist, but whether or not they can be identified *naturally*. Can we identify, in individuals, properties or behaviours that *inherently* place them within a generalizable kind, and indicate *internal* commitments to the associated standards? Can we identify, as it were, essence *within* the individual?

[23] For a review of modern essentialist doctrines see Sonia Roca-Royes' 'Essential Properties and Individual Essences' (Roca-Royes 2011).
[24] "[The essential is] the mark of a concept, not the property of an object" (Wittgenstein, 1994, I, 73). Similarly: "If you talk about essence – you are merely noting convention" (Ibid. I, 74).

Wittgensteinians might deny the possibility of natural kinds and natural rules of development. Generalized categories, they say, cannot be derived from instances of specific individuals. An object alone cannot instantiate *itself* with varying success, there is nothing *general* in an individual. Individuals simply and always are what they are, there is no question of *degree*. Therefore, any systems of classification must be created concepts, externally imposed.[25] This sort of thought seems intuitively plausible in the case of ordinary common objects: a toaster that doesn't toast hasn't failed to act according to some innate or natural obligation. If it has failed to embody some standard, it is one imposed upon it, and its failure means nothing in terms of the object itself. However, to say that essences are *created* categories shouldn't imply that the act of creation is entirely arbitrary. Existence precedes essence, so Wittgenstein or Quine might assert, but that shouldn't mean essence isn't to be found *in* existence.

Wittgenstein himself recognized that in the process of creating classificatory categories, we still intend to capture something about real states of affairs: we are conceiving a picture of the world, how it really is and can possibly be, and such a process "can at most do what a painting or relief or film does, [it] can't put there what is not the case" (Wittgenstein 1953, I, §520). The Wittgensteinian can admit that some degree of arbitrariness is entailed in the creation of essence through semantic practice: we decide what characteristics are originally relevant to some category, or which certain characteristics will represent identity with a class rather than being merely coincidental. In trying to describe the natural world however, we discover subsequent natural facts about our creations.

[25] Quine, in *Word and Object*, held a similar view. He claimed of essential properties that what is essential, and what is contingent, comes down to referential bias. "[I]nsofar as we are talking … with no special bias … there is no semblance of sense in rating some … attributes as necessary and others as contingent. Some … attributes count as important and other as unimportant … some as enduring and others as fleeting; but none as necessary or contingent." (Quine 1960, 199)

The concept of a 'geometric square', which allows us to point at to two things and say of each in turn 'this is a square, and so is this', might be initially arbitrary – we decide to base a classification around four equal sides joined at the corners, and rule out other considerations, like what those sides are made of, or what is nearby. Having decided the criteria, however, we soon discover new truths: necessarily a geometric square will split into two isosceles triangles, even though that criterion was not initially selected for. The fact is nevertheless in the *essence* of those characteristics we *did* select. In this way we might say we discover the *natural essence* of that thing we call a square. We have discovered a further, non-arbitrarily selected, condition an object must abide by in being a square.[26] Whether or not we ever actually get around to defining those natural properties as constituting square-ness, they are nevertheless *there*, 'waiting' to be defined.

Geach's project requires this sort of natural essence, but it also demands something further – not simply facts about the way an object necessarily must be, but in addition general facts about how it *should* be. A square either is or is not a square; a *human*, however, is said to be able to be *more* or *less* successful as an instance of its kind, without *ceasing* to be a member of its kind. The Wittgensteinian might still balk at the idea of natural, *generalizable* facts, and deny this aspect of Geach's theory. Again however, even philosophers generally amenable to Wittgenstein's ideas make room for certain exceptions. In the case of certain sorts of objects, the idea goes, natural essences are teleological in nature, and place the object within a greater framework.

G.E.M. Anscombe was profoundly Wittgensteinian in much of her writing, yet argued nevertheless for the existence of persistent, inherent "patterns of development" (Anscombe,

[26] Similar arguments continue to be made, recently in the first few chapters of Ted Sider's *Writing the Book of the World*. In one sense Sider continues a trend of responses to Quine or Wittgenstein, similar to those of Anscombe, Kripke (1972), Putnam (1977) or Fine (1994).

2006, 32) in certain types of objects.[27] These represent generalizable and development-directing characteristics of a class. These arise naturally in the individual, reflective of and inherently relevant to a historical pattern the individual continues. An individual may thusly be judged as an example of their relevant type by how well they realize these patterns.[28] By virtue of a biological history, for example, we can say that a tiger should have four legs and sharp teeth, should hunt and be carnivorous. Similarly a man should have two legs and two arms, and develop the powers of speech and rationality. These are not just facts about how men and tigers *are*, rather they are facts about how they *should* develop – consistent even with there being currently *no* examples that actually have the relevant attributes. Even should every living human be suddenly struck blind, this is *aberrant*, and subsequently born humans *should* still develop sighted.

So we are offered the idea of characteristic features of kinds - truths about how individuals *should* and do develop - and necessities that hold in virtue of those truths. In attributivist literature, such characteristic features are typically paraphrased from Anscombe as 'Aristotelian Categoricals'.[29] Likewise the necessities that exist in virtue of such features are 'Aristotelian Necessities'.[30] Anscombe introduced these thoughts as the natural progression of Geach's position, offering 'virtues', those traits that realize Aristotelian Necessities, as the teleological focus of natural standards of development.[31] In so doing, she brings the discussion

[27] "The definitions of *animals, plants and chemical substances* might be regarded as forms which certain things fit." (Anscombe 2005b, 29) Emphasis mine.
[28] For a thorough account of Anscombe's views on reconciling Wittgenstein and natural essence, see: 'The Question of Linguistic Idealism' in *From Parmenides to Wittgenstein* (Anscombe, 1981a, 112-133) and 'Human Essence' in *Human Life, Action and Ethics* (Anscombe 2005b, 27-38). .
[29] See Anscombe, 'On Promising and Its Justice' (1981b, 15, 18-19), 'On the Source of the Authority of the State' (1978a), and 'Rules, Rights and Promises' (1978b), Foot, *Natural Goodness* (2001, 46), and Thompson, *Life and Action* (2008, 73).
[30] Roughly: that "without which good cannot be or come to be" (Anscombe 1981b, 15)
[31] The line between Aristotelian categoricals and necessities is often a hazy one. As an example of the interplay between Aristotelian Categoricals, Necessities, and Virtues: categorically a tree grows towards the light. Necessarily in this, it requires drawing nutrients from the soil, and so necessarily it develops roots, which in turn is another categorical. The deeper and stronger a tree's roots are, the better it realizes the categorical demand represented in the necessity. Depth and strength of roots then are (some of) a tree's virtues.

to the *second* question of identification: can we identify any *specific* human necessities, and, if so, what are they? Anscombe didn't pretend that her talk of necessity and virtue did away with the problem of identification – in fact her arguments only further illustrate its difficulty, for what, exactly, are *human* virtues? What is necessary in being *human*?

One sentence in 'Modern Moral Philosophy' neatly introduces Anscombe's view, which I here paraphrase: it is not profitable for us to do moral philosophy until we have an adequate philosophy of psychology.[32] Anscombe specifies a philosophy of psychology because she is explicitly focused on the evaluation of *voluntary actions*.[33] She saw human essence as wrapped up exclusively with the mental. There is not much in the way of specific rationale provided by her for that view, and it's not immediately necessary that we agree with her. The relevant point is that before we can sensibly speak about what is *good qua* human, we need to know what the human *is,* and how it operates. The problem is: what *is* human? The naturalist, cognitivist attributive position demands a determinate set of factual considerations that make up what is essential in being human. Self-directing humans, however, can do so many different things, and develop in so many different ways. Which human traits and actions, then, are *essential*, and which only secondary or accidental?

Certainly we don't want to say that *every* trait we might grant a human being is essential in *being* human. Not only would this trivialize the issue of ought – for we ought to do then whatsoever we might do – but many traits seem clearly of little consequence. It's a categorical truth that humans develop sight, but it's also a truth that they develop eyes of various colours. We should be hesitant however to class these developmental truths as equally important. Having particularly coloured eyes seems arbitrary and of little concern in human affairs. Likewise, presumably the concept of human kind would easily survive should arm-hair cease

[32] Anscombe 2005c, 169

[33] "I am so using 'human action' that nothing is a human action unless it is a voluntary action on the part of a human agent. Otherwise, like digesting your food and breathing and sweating, your acts are the acts of a human agent but are not what I call human actions" (Anscombe 2005a, 203)

to be associated. What then differentiates these from *important* traits that *are* constitutive of the human essence? How do we determine what *is* important, in a non-arbitrary and convincing way? Moreover, what justifies respecting one set of characteristics at the expense of others? The traits we are searching for are intended to represent *the* good of an individual: they are to be *supersessionary*, their consideration supersedes other, non-essential considerations.

Anscombe expressed the problem with hierarchical traits poignantly in the following:

> Essentially the flourishing of a man *qua* man consists in his being good (e.g. in virtues); but for any X to which such terms apply, X needs what makes it flourish, so a man needs, or ought to perform, only virtuous actions; and even if … he flourishes less, or not at all, in inessentials, by avoiding [inessentials] his life is spoiled in essentials … [however] there is a huge gap … which needs to be filled by an account of human nature, human action, the type of characteristic a virtue is, and above all of human 'flourishing'. … For it is a bit much to swallow that a [perfectly 'virtuous'] man in pain and hunger and poor and friendless is flourishing.[34] (Anscombe 2005c, 192)

What above all is in the interest of a human? Are there any traits that can't, at least sometimes, be sacrificed for others? Social virtues, physical virtues, mental virtues, we easily warrant sometimes sacrificing one for the others in varying degrees. Of what particular set of characteristics are we justified in saying that they are *essential* – worth pursuing even at the cost of suffering or failure in other areas – and if so, *why* are they essential? A good man is said to be a virtuous one, yet it is difficult to see a good man in one who consistently fails in most of their goals. It is likewise difficult to identify virtues with something other than 'those things that help an agent attain its goals', but such an identity resolves nothing. Which goals are we interested in? Which goals are worth attaining? Which actions are therefore virtuous?

[34] Discussion of attributive goodness is often parsed in terms of an object's 'flourishing', roughly: the state in which an object's virtues, those traits fulfilling the necessities of its kind, are attained and exemplified.

These questions are at the heart of the identification problem. A way to differentiate the *right* goals from the *wrong* ones is required, and talk of virtue offers very little in its absence.

These issues are persistent - current work on attributive value sees J.J. Thomson, in *Normativity*, talking about 'goodness-fixing kinds' – the sort of things that can be evaluated in terms of goodness because they 'fix' a relative evaluative domain in virtue of their kind. Thomson has drawn criticism for being unclear on just what differentiates goodness-fixing kinds from their evaluatively neutral counterparts – on what it is in virtue of which a kind counts as goodness-fixing. T.M. Scanlon, drawing on Paul Ziff,[35] puts forth in turn that, in any important sense, what makes a kind 'goodness-fixing' is not an objective standard, but rather that it identifies certain interests – interests "relative to which things can be [evaluated] according as they 'answer to' [them]" (Scanlon 2011, 445). Such interests need not be determinate or necessary, and as such would create trouble for the idea of firm, cognitivist goodness-fixing kinds.

Conveniently, the connection between kinds and interests serves as a nice segue into the next section. For now though, this section has explained an objection to Geach's attributivist proposal for goodness, namely that the sort of kind standards required might not exist, or might not be determinable. I've argued that it's plausible that they do exist. It remains still unclear though just what they are, or how to determine them.

§2.3.2 – Motivation

In comparison to the problem of identification just discussed, the problem of motivation is familiar. Generally speaking, the problem is predicated on the observation that evaluations of goodness are not supposed to be inert with regards to our motivations and dispositions. Ethical, normative properties are not like, say, rocks – just out there to be acknowledged or not, and

[35] See Ziff, 1963, *Semantic Analysis*.

implying nothing in particular for our dispositions towards them. On the contrary, we expect someone who acknowledges that *x* is good for them, or that they ought to do *x*, to be at least somewhat motivated to actually do or promote *x*, at least insofar as they are rational. Goodness is supposed to entail reasons to act. It must have what Parfit describes as a 'reason-implying sense' (Parfit 2011, ch.1).

Explaining the above is an obstacle every metaethical theory must overcome, but is a particular challenge for views like Geach's. Attributivism as it's so far been explained focuses on descriptive, comparative evaluations. A fast snail is one that is faster than other snails; a large flea is one that is larger than other fleas. Similarly, some x is a good x if it possesses certain characteristics that classify it favourably compared with others of its kind, or successfully in regards to a relevant general standard. How though are we to account for a consistent and reliable connection between descriptive facts like those, and motivations to act on them in particular ways? Why imagine that a subject having or not having some particular set of characteristics should mean anything for the way it ought to behave? A step seems to be missing.

That the problem of motivation is well established doesn't mean that it's clear cut. Two distinct problems of motivation fall out of the above. They are as follows.

The Normative Problem. Why *ought* agents be motivated by attributive goodness?

The Practical Problem. How *are* ideal agents motivated by attributive goodness?

The reason for the above questions are straightforward. In the case of **The Normative Problem**, we are entitled to ask what the *normative* significance of attributive goodness is – the mere fact that we might be classified in some way relative to some standard doesn't seem

immediately to entail any reason to *care* about that fact. In the case of **The Practical Problem**, it follows from any answer to **The Normative Problem** that an agent who actually *was* how she *ought* to be *would* be reliably motivated by attributive goodness. If we are expected to believe that such would be the case, we deserve an explanation of just *how* such a reliable connection between attributive goodness and the *actual* motivations of ideal agents comes about. In the rest of this section, I'll explain why, historically, those problems have been seen as particular problems for views like Geach's.

The Normative Problem has the more storied legacy. Views like Geach's attempt to ground normativity directly in *natural*, *classificatory* properties. Historically there has been doubt about the possibility of doing so. Although times have moved on somewhat, it will be worth understanding the origins of this problem. Moore and Hume developed early canonical explanations of it. Moore, with his 'naturalistic fallacy', already partly discussed in §2.2.2, and his 'open-question argument', criticised the tendency of some philosophers to equate goodness with morally inert properties that only accompany, instantiate or facilitate a moral aspect.[36] He famously held to the insensibility of identifying goodness with a property about which one could intelligibly ask 'is it good?' That is to say, if some property is supposed to be identical with what we think of as the good, then it shouldn't conceivably be the case that goodness should fail to be apparent in apprehension of the property. As an example, once we define being fluid as having the properties of being disposed to flow and change shape in response to pressure, there is no further sense in asking of something that possesses those qualities whether or not it is fluid. But there is, Moore contends, a step always missing in moving from a reductive identification of good with some property to an appreciation of goodness itself. It is always intelligible, he thinks, to ask whether any set of properties we

[36] Note that, in a case of misleading nomenclature, the naturalistic fallacy isn't aimed solely at natural properties. Moore argues against reducing moral properties to *any* other properties, natural or supernatural; if it is not inherently moral, a property is thought to be subject to the open-question argument.

identify goodness with are *really good*. Moore's 'open-question', he thinks, can be intelligibly asked of any identification of goodness with some natural property.[37]

Hume makes a similar case with the well-known 'is-ought problem', also occasionally known by the moniker 'Hume's Law'.[38] In an oft-quoted passage from *A Treatise of Human Nature*, Hume takes umbrage with the tendency of philosophers to move directly from observations about what *is* to statements about what *ought* to be done as a result of such observations, without explaining the transition. This is a problem:

> For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it. (Hume, 1739, 3.1.1.27)

Hume's idea is that there is something implicit in normative evaluations, a directive to action, which is not found in natural properties. Normative judgments and classificatory evaluations differ in kind; the former cannot validly be inferred from the latter without some additional detail being provided; one cannot derive a valid conclusion that is not implicit in the premises.

The underlying idea behind both Hume's and Moore's arguments is that there is something implicit in normativity that is lacking from purely descriptive or classificatory evaluations. There are at least two ways of proceeding from this observation. One way is to take Hume as calling for an explanation of the missing premise, and to attempt to offer one:

---

[37] We might also understand Moore's argument, and the feeling of 'openness' that accompanies grounding normativity in natural properties, in terms of reasons. Theoretically there is a commitment to goodness entailing reasons to act. *x* is good <-> *x* entails reasons. In other words, reasons, both practical and normative, are built in to our concept of goodness. Natural properties though don't generally seem to contain reasons – the mere existence of a natural property seems, *prima facie*, compatible with nobody ever being motivated by it. So grounding normativity – an essentially motivating concept – in such properties seems, *prima facie*, dubious.

[38] R.M. Hare, *Freedom and Reason*, p.108.

what *is* the link between description and normative force? Another way, more explicitly in line with Moore, is to accept that an ought-evaluation simply cannot be inferred from a set of non-evaluative facts. If goodness is identical with or implicit in some other property it should be impossible to fail to appreciate that relationship. Thus no reduction of goodness to some set of natural properties can suffice.

The arguments of Moore and Hume, while enduring, are hardly conclusive. Today philosophers recognize that identity of one property with others might be synthetic, as opposed to the analytic relationship Moore found so problematic. The relationship between goodness and the properties it reduces to may be too complicated for even a well-informed agent to easily recognize, thereby weakening the open-question argument.

Kripke gives the canonical example of the above. Water may well be identical with hydrogen dioxide, yet for much of our history the question 'is this water hydrogen dioxide' would have been perfectly intelligible. Even someone familiar with the concepts of both water and hydrogen dioxide might find himself, in the absence of sufficient tools for examination, unsure if a given pool of transparent liquid qualified as either. Similarly, if even aware of their identity, one might fail to appreciate exactly *how* or *why* the familiar macroscopic qualities of water are instantiated by hydrogen dioxide. Yet, nevertheless, there is no open question as to whether or not water is hydrogen dioxide. It is.[39]

Still, central elements of Moore's and Hume's arguments continue to resonate in philosophy today. If anything they have evolved. Modern philosophers may be amenable to the prospect of drawing ought from is, but think there is more to be done. They ask not only how to reductively identify reasons, but, having reductively identified reasons, how to explain why they are *our* reasons. Why, that is, do we have reason to care about the reasons there are? Moore's open question, 'is it good', has evolved into "the familiar question 'Why be moral?'"

[39] See Kripke (1972), 116–144, particularly 128.

(Thomson 2011, 476). Even having established the identity of some moral property, why should we care? What is the link between recognizing 'good' or something we *ought* to do, and being *actually* disposed or obliged to *do* it? Gilbert Harman puts it quite succinctly.

> Why should I care about what's wrong? Why should I care about what I ought to do? Why should I care about what I have most reasons to do? In fact, *why should I care about what I should care about*? No view seems immune from this sort of worry. (Harman 2011, 441)[40]

It is tempting to read Thomson, Harman et al. as reiterating the thought behind Moore's open question. I think the more appropriate way of reading them is as illustrating the *practical problem of motivation*. It's one thing to establish that we should do something. It's another altogether to explain how or why we should expect to find within ourselves any motivation to act upon what we should do, or to avoid what we should not. Absent any such explanation though, it's difficult to say what is meant to be meaningful about having done right or wrong. If even a perfectly rational agent can expect to find themselves completely unperturbed by the fact that they have sinned, so to speak, then what is it to them whether or not they ever act well? Why should they care about what they should care about?

That such a connection between any purported property of goodness and motivation needs to be explained in *some* way is now widely granted, even if no single account has met with universal acceptance. It was Bernard Williams who observed as a condition upon *x*'s being a reason for *y* to *z* that *x* be capable of – at least partially – explaining *y*'s actually doing *z*.[41] Responses to that sort of observation are varied. Some like Williams think reasons must be at least partially constituted by motivations. Others believe it is sufficient that there be a connection between judging something as good and being motivated towards it. Some will

[40] Emphasis mine.
[41] See Williams 1979 and Finlay 2009.

grant that this connection might be external, that we might be contingently motivated towards goodness due to some unrelated bridging desire. These make up the judgment externalists.[42] Others, perhaps more popularly, postulate judgment internalism, and advocate for a direct, internal connection between judging something as good and being motivated accordingly.[43]

In conclusion, even if we take the Open Question Argument to be discredited, and accept the possibility of grounding the normative in the natural, it still remains for anyone who insists on a cognitivist, naturalist reduction of normativity to answer the problems of motivations. This is a particular challenge for attributivists, who seem to be basing their evaluations upon an empirical, comparative classification of agents within a given domain – that of their kind. Such classifications seem easily divorced from our motivations: so what if I am a bad human, why should I care about being human at all? In the very next section, I'll explain how it is historically that attributivists have answered that question.

[42] See Railton 1986, Brink 1989, Sturgeon 2006.
[43] See for example Mackie 1977, Williams 1979, Smith 1994, Korsgaard 1996, Shafer-Landau 1998, Björklund 2012.

§2.4 – SOLVING THE PROBLEMS

In the previous section I have reviewed the two problems that have been predominant in the history of attributive-goodness theory. The problem of *Identification* concerns the ability to identify specific, generalizable standards, reflected in the behaviours of the individual, against which to judge human actions. The problem of *Motivation* concerns why or how, having identified such standards, we should care, or expect our actual motivations to align with our supposed obligations. In the following sections I will review canonical attempts at resolving those problems.

Responses to the problem of *Motivation* have historically been more forthcoming, so I'll attend them first. This will primarily involve a review of Geach's own response to the problem in 'Good and Evil', and Anscombe's elaborations in 'Modern Moral Philosophy' and 'Good and Bad Human Action', with both authors focusing on the relation between human goodness and desires.

Responses to the problem of *Identification* will consider more recent work done by Judith Jarvis Thomson, Philippa Foot, and Rosalind Hursthouse. Thomson's work serves to frame a general sort of response, while the latter two authors serve to illustrate attempts at a more specific formulation of identification, specifically focused on the identification of human essence with the essence of the human *species*.

§2.4.1 – Providing Motivation

Geach was aware of the problem of motivation even while writing 'Good and Evil', and didn't ignore it. What is the connection between recognizing attributive goodness, judging that we should act upon it, and actually being so disposed? Geach recognized the implications of Hume's Law for his descriptive account of good: if Geach's good is descriptive, the connection

to motivation can't be analytic. "From a mere description, advice cannot be logically inferred" (Geach 1956, 38), so what additional element is involved?

Geach's solution was to draw upon a link between descriptive goodness and desire. He proposed that the connection between goodness and motivation "belongs to the ratio of want, choose, good, and bad, that, normally, and other things being equal, a man who wants an *A* will choose an *A* that he thinks good" (Geach 1956, 38). Good, for Geach, entails motivation when coupled with choice. Recognition of goodness alone seems to demand nothing: recognition of a chair as a good chair means nothing for our behaviour in regular circumstances. When we are in the business of *choosing* a chair, however, the goodness or badness of the chair bears inherently upon our actions. We will choose the chair that we think is good, and goodness will generate such action whenever we choose, whether "the *A*'s we are choosing between are knives, horses, or thieves; quidquid appetitur, appetitur sub specie boni" (Geach 1956, 38).[44]

When we are motivated to choose anything, the thought goes, we are motivated to choose in positive accordance with the goodness of the thing. This is not, for Geach, a mere empirical observation – rather, this is an essential aspect of choosing. Geach tells us that we would soon find ourselves talking nonsense "if we [tried] to describe a people whose custom was, when they wanted *A*'s, to choose *A*'s they thought bad and reject *A*'s they thought good" (Geach 1956, 39). I take Geach here as imagining that, in continuing this thought, we would inevitably conclude that this strange people must, in some way, see the bad as worthwhile. In making their strange choice their *practice*, they introduce some regular demand, standard or criteria; they analyse goodness as that which ought not to be chosen, in that it is detrimental to their ends, and analyse the bad as being that which meets their criteria. There could be no other

[44] Literally – "Whatsoever is desired, it is desired under the form/appearance/aspect of the good". This might be taken to imply that desire follows from an appreciation of goodness, but this would be a problem for Geach's acceptance that recognizing goodness alone doesn't imply anything for an agent's motivation. Instead, we should understand this to mean, for Geach, that, for any *A* desired, desire for *A* will accord with an understanding of its particular goodness.

explanation for their actions, but such a practice smuggles into their motivations goodness in the guise of badness, and vice versa. In some way or another they think the bad is good. The bad is what they choose because it is good in terms of their standard; they may be confused or misguided in their understanding of goodness, but a connection remains between their unrecognised concept of goodness and their actions. This seems generally right to me, but I don't set out here to explicitly defend Geach, only to point out that in his thought there seems to be a necessary connection between recognizing good and choosing in virtue of it.

We might agree with what's been said so far, yet still object that Geach has failed to really answer why we should care about attributive goodness: being motivated by attributive goodness when choosing chairs or boots or knives is fine, but what if we don't *want* to choose amongst those things? These goods mean nothing if we don't care about chairs or boots or knives and so want to choose among them. So what if we don't care about *human* goods? Or, if we tend towards the extreme, what if we don't care about anything? Why be good?

Geach has a ready answer: the question 'why be motivated to choose human goods' is the same question as 'why be motivated to choose how to be human', and in this we have no choice. Under Geach's framework, a necessary connection exists between being motivated to choose amongst *A*s and caring about choosing good *A*s. The question 'why care about chair goods' seems sensible because we can reasonably ask 'why care to choose a chair?'. That is, we can ask whether or not we have any reason to be choosing a chair – sometimes chairs are not relevant concerns. Choosing how to be *as a human* however is different. *Whenever* we make a choice, we are by necessity choosing how we will be as a human. Every action we take is by necessity a *human* action, one that stands to make us better or worse as humans, and so whenever we choose to act, we make a choice on how to be as a human. We can't help but do so; that's is a consequence of *being* human. So the question 'why be motivated by human goods', for Geach, is the same question as 'why be motivated to choose human ways of being'.

The answer is simply that we cannot do otherwise, and so human good is unavoidably relevant to our motivations.[45] [46]

§2.4.2 – Wanting and Choosing

Geach's attempt to link evaluation of human ends necessarily to the act of human choosing is crucial to his project - however, he might reasonably be charged with having overlooked a particular difficulty. Geach claims that human goods must play a part in our motivations by virtue of our being required to make human choices. The problem is that, on the logic so far provided, that is no valid conclusion. Geach's original claim was that goodness belongs to the 'ratio of want', whereby someone who wants something, when choosing such things, will choose a good one. This can be rephrased as: *if* someone wants an *A, then* they will choose a good *A* when choosing among *A*s. Put that way, it's clear that the goodness of *A*s becomes relevant in virtue of one's being in a state of *wanting* an A. But Geach also maintained explicitly that the normative relevance of Human goodness in Human affairs owes to no "individual peculiarities of desire" (Geach 1956, 40). As I've shown above, he explains such by appealing to the unavoidability of choosing among Human actions. In doing so though, Geach's claim has become that being in a state of *needing to choose* among *A*s makes the

[45] Geach tells us specifically – "what a man cannot fail to be choosing is his manner of acting; so to call a manner of acting good or bad cannot but serve to guide action. As Aristotle says, acting well, εὐπραξία is a man's aim simpliciter, ἁπλὼς, and qua man; other objects of choice are so only relatively, πρός τι, or are the objects of a particular man, τινός; but any man has to choose how to act, so calling an action good or bad does not depend for its effect as a suasion upon any individual peculiarities of desire" (Geach 1956, 40).

It is odd here that Geach references calling *acts* good or bad, given that Geach has also argued that "'Event' … is too empty a word to convey either a criterion of identity or a standard of goodness" (Geach 1956, 41). The general 'act' seems quite similar in this regard. In order to properly understand Geach here, we should replace each instance of 'act/action/acting' with 'human act/action' or 'acting qua human'. It is not good or bad acts in general that factor in human motivation, but good or bad *human* acts, acts that are judged by how they further, hinder or otherwise express particular human goods.

[46] There is a similarity here between Geach's view and the modern *constitutivism*, wherein our status as acting agents itself is taken to entail normative ends. There is, however, an important distinction in that constitutivists typically aim to ground normative judgments in facts about action *qua* action, where for Geach such normative truths stem from facts about kinds. The similarities may be informative nonetheless. See Katsafanas, 'Constitutivism about Practical Reasons', forthcoming.

goodness of *A*s relevant. That is quite different from the original claim, and not something we should accept.

It is no great stretch of the imagination to picture being made to make some choice we don't care about. Imagine being forced at gunpoint to make an arbitrary choice between three chairs. Whether or not you pick a good or bad chair is there totally irrelevant, regardless of your being stuck making the choice. Geach needs to supply a reason why being positioned in such a way as to make a certain choice inescapable will *also* demand that our desires reflect the standards of the options involved, in order to make the goodness of the options relevant. He hasn't.

Of course we should object to my above example: in that scenario we aren't *really* choosing between chairs – we're choosing between devices that will save our life. That the goodness and badness of the objects don't *seem* to factor is explained by each chair being equally good in *that* respect. Correspondingly, a charitable – and probably correct – way to take Geach is as saying that every (purposeful) choice must involve *some* goal, and so *some* standards, and so make relevant the goodness of the options *qua* those standards. I wager that's true, but it *still* doesn't do what Geach proposes and guarantee a place for Human goodness in Human action. Geach intends the argument that, *as* Humans, wanting to choose any *x* is wanting to choose a human action. My response is that that's technically true, in the same way that, above, wanting to choose between life-saving options *incidentally* involves wanting to choose between chairs. But just as we needn't want to choose between chairs *qua* chair, we needn't want to choose between Human actions *qua* Human.

Even granting that *some* standards must be involved when choosing between Human acts, it doesn't follow that those standards must be *Human* standards. This is because making a choice between options doesn't demand that we must be choosing those options *as themselves*. The options may serve our goals in a variety of ways – although we might

technically be involved in choosing between fruits, for example, the goodness involved needn't be the goodness of fruits *qua fruits*. We might be choosing them *qua* ammunition to throw at some dire comedian. Likewise, although we may be unavoidably positioned to choose between Human options, Geach hasn't yet provided a reason why we should need to choose between them *as* Human options, rather than with some other goal in mind.

Geach has proposed, strictly, that since we *must* choose between Human actions that we must embrace the standards of Human actions. That isn't true. That the objects of our choice happen to be Human actions might be incidental and have nothing to do with our purposes in choosing among them. For Geach's project to succeed, a way must be supplied to connect the unavoidability of choosing between Human actions to an equally unavoidable *interest* in Human affairs. Geach requires not only that we are necessarily confronted with Human options *in* our choices, but that we can't help but hold some specifically Human ends. Geach doesn't offer such an argument. Others have.

§2.4.3 – Necessary Wanting

Anscombe's views on goodness were very much in line with Geach's. In many ways she developed and refined Geach's original ideas. In *Intention* she echoed Geach's view on the connection between desiring and goodness, arguing that a man who wants something – say a saucer of mud – without also thinking that it is *good* in at least *some* way, is a "dull babbling loon" (Anscombe 2000, 70).[47] Notably though, Anscombe was sensitive to the fact that such a connection wasn't enough to guarantee the sort of priority for *human* goodness that both she and Geach desired. So, while Geach tried to ensure a role for human goodness by focusing on the supposed necessity of making certain *choices*, Anscombe recognized the difficulty

[47] Many others have since come to be convinced of this sort of thought. Thomson endorses Anscombe's argument directly in *Goodness and Advice* (Thomson 2001, 40) and similar points have been made by Warren Quinn, 'Putting Rationality in Its Place' (1993) and Richard Kraut, 'Desire and the Human Good' (1994).

previously discussed, and sought a way to establish the necessity of *having* certain *desires.* Her attempt is the focus here.

As I've so far explained, both Geach and Anscombe rely on the idea of underlying, teleological standards involved in *being* certain kinds of entities. Furthermore, both philosophers rely on the idea of *necessity* or *need* to move from the existence of these standards to obligation on our part to act, at least at times, in accordance with them. For Geach this is the fact of being necessarily confronted with certain standards in the process of our pursuits, choices or actions. For Anscombe these necessities are the aforementioned *Aristotelian Necessities*, necessary traits that develop in the process of *being* some kind of thing, and the needs that arise in virtue of these traits. To some degree these stances answer the *normative problem of motivation* – Geach and Anscombe think we *ought* to do what is in our good, because doing so will realise our ends, fulfil our needs, or otherwise make our endeavours go well. What Geach seemed to omit, however, and what Anscombe was very aware of, is that needing to do something – whether it's in virtue of our kind or in virtue of our situation – doesn't seem to guarantee actually *wanting* to, and particularly not in any way that makes anything's *goodness* particularly relevant. The *practical problem of motivation* remains to be answered.

In 'Modern Moral Philosophy', Anscombe describes the problem of moving from the recognition of something's attributive goodness, to being motivated to promote it. Very often we can recognise what is good for something without caring about it in the slightest. Though she thinks, for example, that it is 'not at all dubious' to move from an understanding of what a plant *is* to an understanding of what is *good* for it, and thus how it *ought* to be, all of these facts mean nothing unless we also happen to have some interest in the plant's well-being (Anscombe 2005c, 177-178). Recognition of the plant's needs and corresponding good don't necessarily correspond to motivation – why should we think recognition of our particular *Human* goods

will entail anything different? If it doesn't – if a rational agent can be confronted with all the fact and fail to be moved, whence normativity?

Anscombe's response may be found in the following passage:

> Certainly in the case of what the plant needs, the thought of a need will only affect action if you want the plant to flourish. Here, then, there is no necessary connection between what you can judge the plant 'needs' and what you want. But there is some sort of necessary connection between what you think *you* need, and what you want. The connection is a complicated one; it is possible *not* to want something that you judge you need. But, e.g., it is not possible never to want *anything* that you judge you need. This, however, is not a fact about the meaning of the word 'to need', but about the phenomenon of *wanting*. (Anscombe 2005c, 178)

To understand Anscombe's point, it's worth taking particular note of these specific statements: "there is [a] necessary connection between what you think *you* need, and what you want", "it is not possible never to want *anything* that you judge you need" and "this [is a fact] about the phenomenon of *wanting*". Here, Anscombe expresses an idea that will play a crucial role in attributive normativity. Simply put, our desires and interests don't arise in isolation, appearing through smoke and magic and according with no consistent rule. Rather, apprehension of our needs fundamentally informs our wants. Insofar as we are rational, she thought, our interests will be shaped by facts about what we are, and what we need as such.

Anscombe is no longer expressing the earlier simple sentiment that the phenomenon of wanting is necessarily tied up with judgments about goodness. We've seen that this is insufficient to ensure the sort of role for goodness that Geach et al. require – namely, that *our* goodness features necessarily in our desires. Instead Anscombe is now making the stronger claim that one's goodness will always be relevant to one's desires, because one's desires are invariably aimed to some extent at what one *needs*. It isn't entirely clear whether or not

Anscombe believed that because she thought judgments about needs *create* corresponding desires, or merely make the rational agent aware of what is already driving them, or what grounds the phenomenon of desiring.[48] Either way, Anscombe denies the possibility of a rational agent who is aware of her needs and attributive goodness yet completely unmoved by any judgments about them.

Anscombe's justifications for that claim are unfortunately sparse. Additionally, what she *has* provided seems to allow that the needs which shape our interests and impose standards needn't necessarily be *human* needs. After all, we can judge that we need many things, for many reasons, including perhaps *because* of our desires. If it is possible not to want some things we judge we need, then, at least conceivably, our *human* needs might never happen to be those that we end up wanting.[49] Nor does she seem to have guaranteed that wants reflecting our particular human needs should supersede other wants in cases of conflict. These sort of difficulties, and possible solutions, will be returned to later. For now, however, Anscombe has laid out an essential attributivist principle – that interests are, in some way, a reflection of needs, and align with the standards they imply.

To recap, Geach argued that Human goodness has a necessary connection with Human motivations, because when one wants an *x* one will be motivated to choose an *x* they think is good, and one can't help but be involved in choosing how to be as a Human. The problem in Geach's account is that the connection he proposed was between goodness, choosing and *wanting*, yet while it is unavoidable that one chooses to act as a Human, it isn't prima facie unavoidable that one *wants* to. Anscombe's solution to that problem was to posit a necessary

[48] For the latter type of explanation, see Rosalind Hursthouse. Hursthouse argues that facts about one's kind, about what *sort* of agent one is, go all the way down – informing desires, interests, and dispositions to act (Hursthouse 1999, 123).

[49] Anscombe does, in fact, seem to have seen this as a real issue for any sort of project, such as this, which eliminates the absolute quality of 'moral ought', and makes it contingent upon natural facts. The difficulty this sort of theory has in providing moral absolutes was a particular complaint of hers – see 'Modern Moral Philosophy' 182-183.

connection between evaluation of one's needs and what one desires. Together Geach and Anscombe make up the canonical attributivist view on motivation: the desires of even rational agents can't fail to correspond to at least some degree with the needs of their kind, as evaluation of one's needs influences the formation of one's rational ends. Since one can't help but desire to fulfil one's needs, one can't help but be motivated by one's own attributive goodness.

I'd like to make two observations to close this section. The first is that we can see from the above that attributivists seem committed to some form of judgment internalism. There exists a necessary connection between normative judgments and motivation – judgments about one's own goodness are at least in part evaluations of one's needs, which can't fail to produce corresponding motivation in rational agents.

The second observation is that, as established, there are two problems of motivation, **The Normative Problem** and **The Practical Problem**. Geach and Anscombe focused primarily on the *practical*, but it may seem that having done so – having answering why we *are* motivated by goodness – hasn't answer the *normative* concern – why we *should* be.

There are two ways we might take attributivists as responding to the above. One way is to take attributivists as identifying one's needs with one's *objective interests* – the sorts of things that, when fulfilled, lead to one's flourishing or otherwise just makes one's life go better. We ought to be motivated to fulfil our needs because that's what makes for a good life, the sort of life we should want.[50] It's hard not to see this move as circular – unless they want to

[50] Judith Thomson seems at times to favour this approach. Across several 'replies to critics' she disagrees with philosophers like Scanlon and Smith, who believe we only have reason to be motivated by kind standards in cases where we have a pre-existing reason to act upon them, represented by an interest that is met through their fulfillment, or where they have the higher order property of 'deserving to be desired' (Scanlon 2011, 447, Smith 2010, 726-731). Thomson argues instead that while it is true that something's being in the good of your kind entails having reasons to act towards it, those reasons are not prior to facts about one's kind, but *are*, or are grounded in, those facts (Thomson 2010, 749, 2011, 475). She argues that such is what explains the difference between interests we have reason to pursue – like our interests in sustenance or relationships – and those we have *no reason* to pursue, like an interest in smoking a thousand cigarettes (Thomson 2011, 474). Scanlon sums up their disagreement thusly: "the difference between Thomson's view and mine has to do with the relative priority we assign to truths about reasons and truths about [kinds]. She thinks the latter are primary … I of course believe that there are truths about reasons for action that … give content to claims about goodness of a kind. She believes

introduce some further understanding of good, then what makes for a good life for the attributivist just is one that's attributively good. We wanted though to know why an attributively good life really is one that we *should* be motivated towards. Identifying needs with objective interests just repeats that the needs of one's kind are what we should be motivated towards, without providing any further justification.

A second and better way of understanding attributivists is as answering **The Normative Problem** *through* **The Practical Problem**. Attributivists should, I argue, be seen as committed not just to judgment internalism, but to a form of *existence* or *reasons internalism*, on which not only is there an internal link between normative judgment and motivation, but that what makes something normative owes at least partially to facts about what we are and can be motivated to do. The reason philosophers like Geach and Anscombe focus almost exclusively on the practical problem of motivation is because, for them, answering the practical problem *does* answer the normative problem – facts about what we *are* motivated by at least partially explain what we *should be* motivated by. Further arguments for this interpretation will be given later in this thesis, particularly across §5.2 and §6.1.

§2.4.4 – Towards Identification

Here again is the problem of identification. I've granted for the sake of argument that natural kinds exist, and that belonging to a kind entails having certain defining dispositions or functional aims. Say then that I am a member of the Human kind. Say also that there are many things I am disposed to do, or at which I can be said to aim. Which of those things are those that mark me out as *Human*? Certainly not *everything*; if that were the case then doing well at *anything* would be doing well at being Human, and so the classification would be meaningless.

that the reverse is true." (Scanlon 2011, 449). Thomson's views on the matter will be returned to later in this thesis, particular across early sections of chapters 5.

By what right, though, do I decide that behaviour *x* is in the essence of being Human and behaviour *y* is not? *Prima facie* it seems that, because I *am* Human, anything I do is done *as* a Human, and so if I am successful at anything then I am being successful as a Human. The attributivist must argue that some of my successful actions *don't* count as successful *Human* actions, but it isn't clear why, once we accept that I *am* Human, that any of my actions should count as *less* Human than others.

In this section I present how attributivists canonically respond to the above problem. The intuitive force of the problem, I argue, comes from a tendency to try to move from kinds to functions – to begin by accepting the existence of kinds, and then trying to demystify which of the many things its members do are the *characteristic* behaviours of the kind. I argue that attributivists must and do reverse that process. Attributivists *begin* by observing the existence in the world of certain behaviours, and go on to identify a kinds with the class of objects that account for those behaviours. They observe, for example, that in certain circumstances bread gets toasted, and say that a toaster is whatever accounts for that phenomenon. Something is a toaster only to the extent that it is aimed at accounting for toasting. In that way they limit what gets to count as an action of a specific kind. They argue that a functional, kind-defining act is just one that explains the phenomenon in virtue of which a kind is posited in the first place.

Judith Thomson provides a good example of the attributivist view on identification. In *Normativity*, Thomson argues that normativity is grounded in the standards of what she calls 'goodness-fixing kinds'. What is a goodness-fixing kind? Kind K is a goodness-fixing kind, says Thomson, "if and only if K is a kind such that what being a K is itself sets the standards that a K has to meet if it is to be a good K" (Thomson 2011, 473). Thomson provides a ready list of the sorts of kinds she takes as 'goodness-fixing': toasters and lawnmowers, seeing-eye dogs and tennis players, beefsteak tomatoes, tigers and humans, etc. These are all kinds that Thomson thinks are marked out – that *exist* even – in virtue of there being something that they

do, at which they can be good or bad. There are no tennis-players save that some agents play tennis, and nothing is a lawnmower unless it's in some way aimed at producing mowed lawns. Being one of these things involves existing relative to certain standards. Thomson contrasts her goodness-fixing kinds with kinds like 'pebble', for which there is no good or bad. Any pebble is as good as the next, *qua* pebble – not so for toasters, tomatoes, or Humans.[51]

I think Thomson has asserted something profound in the above, but it may be hard to notice. Thomson might seem to have reiterated the belief that some kinds are marked out by having essential standards, when what we want to know is what they are, and why they get to count where others don't. Thomson has said something more though. The impression that she hasn't comes about if we mistakenly paraphrase her as saying that 'a goodness-fixing kind K is a kind such that *x*'s being a K sets functional standards for *x*'. That would be wrong. Such a definition leaves open the option of seeing the kind as coming *first*, prior to the standards – the kind exists, and *then* in doing so somehow gives rise to certain standards by which it may be evaluated. That though leaves us just where we were, asking which of the myriad things a member of a kind might do are important. Thomson's goodness-fixing kinds work the opposite way – it is better to rephrase her as saying that what it is *to be* a goodness-fixing kind *is* to be the sort of thing that embodies certain standard-implying behaviours. Her point here is best taken not as being that the existence of a kind entails standards, but that the existence of standards entails a kind. "Being a toaster", she says, "is being an artifact manufactured to toast, and *that* itself sets the … standard … *qua* toaster" (Thomson 2008, 21).

Toasters, tennis players, and seeing-eye dogs are all kinds that are brought about *in response to* or *in accordance with* certain standards. Tennis players and toasters don't exist prior to the standards they correspond to – something simply cannot *be* an artefact manufactured to toast prior to the standards of the goal of toasting, and a tennis player *is* a

[51] See Thomson 2008, 19-22.

tennis player only when it acts according to the standards of playing tennis. To be a member of such a kind *is* to subscribe to its standards. What all goodness-fixing kinds have in common, Thomson claims, is that they are brought about, in some way, by the imposition, internal or external, of a standard.

So the step Thomson has taken by way of solving the problem of identification is to identify kinds in response to the existence of demarcating, standard-implying behaviours – something a subject does or is aimed at doing in virtue of which it is considered a member of a kind. Doing so explains why we are justified in assuming kinds have essential characteristics, for what it is to *be* such a kind is *essentially* to act according to such standards. Doing so also helps us understand how to identify the standards of a given kind – we only need to identify the standard-implying behaviours in virtue of which we stipulate the existence of a distinct kind. Thomson has captured here a crucial aspect of attributivist theory – although she is the most clear in expressing that commitment, we can see the same idea in the work of other prominent attributivists, as they attempt to answer the obvious remaining question.

What are the behaviours and standards in response to which we assert the existence of, and membership within, the *Human* kind? Or indeed of any complex, semi-autonomous life-form? It's easy to know the defining behaviours of toasters and lawnmowers – they are artefacts, they have simple, straightforward functions we impose upon them in creation. Organisms though, even ones much simpler than Humans, are more complex.

Philippa Foot and Rosalind Hursthouse have both attempted to explain the defining features of organisms. They both offer similar theories.[52] Hursthouse characterises the kind

[52] Foot's *Natural Goodness* and Hursthouse's *On Virtue Ethics* argue for very similar positions. In this thesis I focus more heavily on Foot, whose work is more concise and who I take to present most effectively the predominant species-centred position amongst Neo-Aristotelians, and attributive-goodness advocates in general. This view is at least partially informed by the fact that Hursthouse, unlike Geach, Anscombe, Foot or Thomson, doesn't view her project as intended to produce motivating reasons, or to convince hard-headed anti-moralists, but rather to provide a natural explanation, justification or validation of presupposed ethical virtues – despite her belief in a link between virtues and emotions, as previously mentioned. See Hursthouse 1999, 194. See also Nussbaum 1995, where it is argued that the sort of naturalism advocated by Neo-Aristotelians is informed by and aimed at providing a natural framework for presupposed ethical understandings.

'organism' in general as defined by the pursuit of four ends: (i) individual survival, (ii) continuance of the species, (iii) freedom from pain and participation in enjoyment, and (iv) the good functioning of the social group. She thinks we further categorise individual living things as members of species by evaluating how they employ their (i) parts, (ii) operations/reactions, (iii) actions and (iv) emotions/desires, towards the aforementioned ends. The way they go about pursuing those ends – the teleology of *the species* – determines just what *kind* of organism they are.[53]

Similarly, Foot argues that "plants and animals have … an 'autonomous', 'intrinsic', or … 'natural' goodness and defect … notably different from what is found elsewhere in other things" (Foot 2001, 26). This is, for Foot, a sort of 'first-order' goodness, contrasted with 'secondary goodness', or what Thomson calls 'goodness modified' (Thomson 2010, 27) – the goodness that is ascribed to something in accordance with how it meets or is employed to meet some external standard (Idem, 22-24). Specifically, this 'natural goodness' is evaluated against the framework of 'self-maintenance' which Foot takes to define the teleology of life-forms (Foot 2001, 31). Specific kinds of lifeforms then are determined by the distinct ways certain organisms go about the process of self-maintenance. Bees are defined by cooperation within a hive, and by stinging to drive away threats. Humans are likewise defined by the characteristic ways we pursue success as a species – through, thought Foot, embodying social and rational virtues. Foot took quite literally Geach's quote: "Men need virtues as bees need stings" (Geach 1977, 15).

Both Foot and Hursthouse have proceeded in their analyses in just the same sort of general way that Thomson outlined. The tactic of the Neo-Aristotelian, and of attributive-goodness identification in general, has been dominated by a focus on species and biological teleology, but the underlying theme is consistent. Certain behaviours are *essential* in being a

[53] See Hursthouse 1999, particularly 197-202).

specific kind of thing, because to be a kind is *by definition* to be disposed to act in certain ways. Attributivism draws standards from the existence of end-oriented systems of behaviour or dispositions, identifies kinds as the categories which correspond to those standards, and sort individuals into those categories in virtue of their demonstrating the corresponding dispositions. For attributivists, you are what you do.[54]

Is all the above convincing? Many objections have been raised against attributivists, not only for how they view species, but for how they derive *moral* standards from systems of behaviour – how certain things are disposed to function – or from the teleology of biological self-maintenance. Many of these will be dealt with extensively in the next chapter. I'd like to end this section however by noting one concern, which I think informs some of the objections to come. Thomson et al. may have contrived of a way to identify essential features in kinds, but what of individuals? Individuals can certainly belong to certain kinds, and *insofar* as they are a member of a kind *K*, they can be evaluated as good or bad *K*'s, but are they *only* a *K*, or might there be more ways to classify them? If so, can attributivists establish *priority* among any specific kinds an individual might instantiate?

The basis for identifying any particular organism as a member of kind seems to be that the hallmarks of the kind are involved in the organism's operations. A kind is said to be instantiated by a specific entity or object because that entity demonstrates certain behaviours or dispositions oriented towards specific ends. That's fine if and when the entity being picked out as a member of that kind evidences only one identifiable set of standard-implying behaviours. In many cases however, even perhaps in the majority of cases, that's unlikely to be so. If one's evaluative standards are implied by one's kind, and one is classified as a kind

[54] In further support of this, see Michael Thompson (1995 and 2003) whose work is referenced by Foot. Across Thompson's work, what it is to *be* an organism is to be situated against a broader normative framework – the identifying of an organism entails the identification of certain 'vital' processes, which cannot be done by examining any particular event in isolation (Thompson 1995, 275-276). In a very real sense, to identify an organism as a kind of life is to pick out and identify a set of teleological behaviours.

in response to evidencing specific dispositions, what happens when one evidences multiple, even conflicting, dispositions? To how many kinds do they belong, and by which standard are they to be evaluated?

Foot was fond of talking about the blue tit, which in addition to possessing developmental dispositions related to survival and self-maintenance, is also innately disposed to develop a blue-coloured head – a trait that supposedly has no influence on its life. Foot has decided that the *important* kind is the species, the kind defined by active self-maintenance. She claims thus that the traits to focus on in evaluation of the individual are those involved in the teleology of the species. The normatively *relevant* kind, she claims, is 'bird' and not 'blue-headed organism', and so the colour of the blue tit's head has no bearing on evaluations of its goodness (Foot 2001, 30-33). But what is her justification for this prioritisation? She doesn't offer one – rather she seems to select this kind as normatively relevant out of mere preference, or a desire to avoid prioritising 'unimportant' traits. Surely though the natural disposition to develop blue-coloured heads is just as 'intrinsic' as the disposition to lay eggs or fly south or whatever it is blue tits do. If these sorts of dispositions are supposed to justify identifying the individual organism with a kind, why then should this sort of arbitrary distinction be accepted? Why consider the blue tit organism as being first and foremost the kind 'bird life-form' rather than 'blue-headed life-form'? It seems one developmental disposition has just as much right to priority as another.

Perhaps this doesn't seem like an important concern when the issue is restricted to birds and coloured heads. The weight of the concern might be more apparent though when we turn to humans. Granting that we can identify a kind with a set of standards, what particular kind has the right to be called *human*, to be identified with the organism we are so intimately concerned with, and why? Thomson, for example, chooses to identify the human kind not merely with the standards of physical fitness – as with tigers – but with moral standards: a good

tiger is a physically fit tiger, but a good *human* is a *morally* fit human (Thomson 2010, 20). Why, however, should we accept *moral* traits as having priority in defining *the* human kind, over any other trait? How does she justify drawing this distinction? It turns out to be quite hard to say. She says vaguely that the difference in evaluation of a human over evaluation of a tiger is due to the fact that "[u]nlike tigers, human beings *can* act morally well or badly, and that is why their being good *qua* human beings consists in their acting morally well" (Thomson 2010, 21). Surely though this is insufficient justification – human beings, unlike tigers, can also rob banks well or badly, yet we should hardly want to say that the goodness of a human *qua* human consists in robbing banks well.[55]

I've said the above is a concern – I hope I've given some indication of why. If the selection of the essential teleological goals of an organism – that is, the selection of its predominant, normatively relevant kind – is arbitrary or down to mere prejudice, then what reason does any rational agent have to abide by the standards of any particular designation? If, on the other hand, we grant that every end-oriented disposition, and its associated kind, has equal claim to normative importance, this should be expected to pose significant problems for any theory of normativity that expects to meaningfully inform action and choices. These issues will be returned to in later chapters. For now, I will conclude this section.

The problem of identification is the difficulty in non-arbitrarily designating certain behaviours as essential in being a given kind of thing. Attempts at solving the problem revolve around identifying natural or goodness-fixing kinds *in response to*, rather than prior to, the identification of end-oriented dispositions in individuals. Identifying kinds as categories created and defined in response to certain behaviours explains why those behaviours are

[55] Similarly, replies of the sort that the fact that moral action is a possible consideration for humans – it's a standard that arises uniquely in the operation of the 'human' organism – makes moral standards relevant in their evaluation, or that moral behaviour comes up unavoidably in human life, should be dismissed. Bank robbing is, again, a possible consideration for humans, and hair growth comes up unavoidably, but a good bank robber is not a good human, necessarily, and a bald human isn't thought, by Thomson, to be meaningfully defective.

essential to being such a kind – acting in such a fashion is literally what justifies membership. The predominant focus of attributivists has been to prioritise the *species* of organisms as the normatively relevant kind, that which determines the standards for normative evaluation of the subject's behaviour. It's unclear whether that prioritisation is justified – the attributivist strategy may only have swapped the difficulty of identifying the essential standards in kinds, with that of identifying essential kinds in individuals.

§2.5 – MOVING ON

The intention of this chapter was to instil in the reader an understanding of Attributivism, its origins, its aims, and its elements. I hope that I've accomplished as much.

To recap, in §2.2 I explained Attributivism's origins, through the lens of Geach's observations on the nature of good. Geach argued that 'goodness' is essentially attributive; it picks out a property that varies in relation to the kind of object which instantiates it. Against contemporaries who objected that such a property would be too ambiguous, implausibly naturalistic, or insufficiently commendatory to ground normativity, Geach argued against the coherence of any other account of goodness, and that variations in what 'goodness' describes imply no ambiguity, but correspond to a consistent, systematic process.

In §2.3 I explained two problems which have long characterised objections to Attributivism – the problems of (i) *Identification* and (ii) *Motivation*. (i) Asks, even if we grant that some entity *x* is a member of kind *y,* how to determine what it means to be a *y*, such that some determinate standards are implied for *x*'s actions in virtue of their membership therein. (ii) Asks, having identified such standards, why we (a) *should be* motivated to promote the standards of our kind and (b) actually *are* so motivated.

§2.4 explained the elements of attributivist responses to those two problems. To answer *Identification*, attributivists assert the existence of a kind in response to the existence of observable systems of behaviour and end-oriented dispositions, and attribute kind membership to subjects corresponding to their instantiation of said systems. To answer *Motivation*, attributivists argue for a necessary connection between being wanting an *x* and being motivated to choose a good *x*, and for a necessary connection between one's needs, the needs of one's kind, and one's wants.

It's interesting to note that in their responses attributivists blur the line between *Identification* and *Motivation*. The answer to either question entails the answer to the other.

For attributivists, what you *are* is a function of what you *do* – one's existence is defined through manifesting certain dispositions. Thus to exist as a member of a kind *is* to be motivated in certain respects, and to be motivated in certain respects is to exist as a member of a kind.

In this chapter I've dealt with the history of Attributivism, its canonical arguments and challenges. Attributivism is, however, very much a live and breathing theory. In the following chapter then, I explore the challenges that have faced Attributivism in recent years, and the sort of responses attributivists should and do offer.

3 – Modern Attributivism, Challenges and Tactics

In the previous chapter I explained the history of Attributivism and its central elements. I explained also the solutions attributivists offer to a set of canonical problems. In doing so however, attributivists have *not* flouted all historical precedent and settled the matter – the dialectic continues. In this chapter I will look at a series of objections that have been levied against Attributivism in recent years. I will explain how attributivists can, do and should respond to each objection. The process of doing so should help to illuminate subtler details of Attributivism than could be accommodated in chapter 2.

What falls out from the responses to the problems of the previous chapter is an attributivist reliance upon natural teleology, to teleological standards predicated upon subjects in virtue of their membership within natural kinds, and from which, it is supposed, normativity derives. Put another way, attributivists are committed to some account of natural *function* – they derive the *proper functioning* of a subject and its parts through appealing to its kind. Almost all modern objections to Attributivism arise specifically because of the sort of functional accounts attributivists must endorse, and the consequences of doing so.

In §3.1 I will explain an objection made primarily by William FitzPatrick, based on the thought that natural teleology can only be the teleology of evolution, and so that evolutionary norms are the only natural norms. Since evolutionarily sound behaviours can often be *detrimental* to individual welfare, FitzPatrick argues that natural teleology can't plausibly be used to ground the good of individuals. I argue that FitzPatrick's objection stems from a failure to distinguish between *categorical* needs and *categorical* behaviours – that is to say, the needs and behaviours that are definitive of membership within a kind – and from incorrectly taking attributivists to prioritise both equally in the life of organisms.

In §3.2 I explain a second objection from FitzPatrick, who argues that in order to avoid implausibly objectionable results, attributivists must take an *ahistorical* approach to biological function, and that doing so leaves them incapable of distinguishing in a principled way between what a subject does incidentally and what it does as a function of its kind. In response I explain how ahistorical functional analyses allow for such principled distinctions. They do so by making relevant only behaviours and environmental effects that are *entailed* or *required* in the operations of a given subject. I argue that any further objection along these lines must beg the question against attributivists.

In §3.3 I look at a range of objections which argue that the norms of natural teleology can endorse behaviours that are obviously implausibly *vicious*, both (i) by intuitive moral standards *and* (ii) by the internal standards of organisms themselves. I argue that (i) is unpersuasive – attributivists are already prepared to accept major revision to our moral intuitions and can accept a degree of moral relativism – and that (ii) is built upon an insufficiently fine-grained view of organisms and their development.

In §3.4 finally I explain the objection that in order to account for the role of rationality in Human affairs, attributivists must appeal to Humans as essentially rational creatures, but that doing so grounds Human normativity in Reasons, not natural norms. I respond that this is only a problem if we take rationality to entail rationality *full-stop*, or as being constituted in responsiveness to Reasons for which there is no informative reductive analysis. I argue that since attributivists take reasons to be analysed in terms of natural facts about kinds, their view does not entail the proposed objectionable result.

The result of this chapter is to demonstrate that Attributivism is able to answer all of the prominent modern arguments against it, and to leave us with a more sophisticated understanding of its commitments, which allow it to do so. We should accept as a result either that Attributivism should be taken seriously, *or* that new objections need to be made against it.

§3.1 – Evolutionary Norms and Detrimental Teleology

Perhaps the most prominent strain of modern objection to Attributivism focuses on the reliance of modern attributivists like Philippa Foot on the teleology of species. The objection is constituted in two claims. First, that the function of a natural kind must be evolutionarily determined. Second, that evolutionary functions come apart from the good of the individual. Thus the good of the individual can't be constituted in achieving the functional standards of one's kind.

Arguments of that sort have been put forth most notably by Robert Adams and William FitzPatrick. Adams takes this sort of argument so seriously that, in *A Theory of Virtue*, it motivates him to dismiss Foot's Attributivism entirely. "If there is a teleology intrinsic to *our* biology," he writes, "it is one in which the *telos* served … is the propagation of … genes; and efficacy in serving *that telos* has … no plausibility as a measure of ethical virtue" (Adams 2006, 51). The sentiment of Adams and other likeminded philosophers is broadly that any account of species teleology must be an evolutionary account. As such any traits or dispositions we possess *qua* that species membership can only be traits that persist due to their evolutionary fitness, for how they contribute to success in gene replication. What it is, then, to be a good member of one's species is to exemplify traits aimed at gene replication. The teleology of gene replication though is no *moral* teleology. Attributivist theories that rely on biological function or species membership to provide moral standards, then, must fail to get off the ground.[56]

The above sort of argument has been influential. Intuitive views of species as tied up with the telos of gene replication have even been credited for the 'cool reception' of Foot's particular application of attributive goodness.[57] The most explicit argument of the above sort

[56] See FitzPatrick 2000 for the most explicit formulation of this argument.
[57] See John Hacker-Wright, 2009, 309.

has been developed by William FitzPatrick in *Teleology and the Norms of Nature*. Like Adams, FitzPatrick accepts that organisms just *are* gene-replicating systems – "functional systems that have as their general and ultimate biological end [the replication of certain genes], with lower level functions and ends all geared towards this end" (FitzPatrick 2000, 186). More, FitzPatrick helpfully explains just *why* he thinks these evolutionary ends aren't of any moral use. He takes it that to determine moral oughts, fulfilling one's evolutionary ends must also promote one's welfare. He then uses certain animals, elephant seals particularly, to demonstrate how fulfilling evolutionary standards fails to line up with the welfare of the individual.

Elephant seals, FitzPatrick argues, adhere to a particularly problematic dominance hierarchy. Male seals battle for exclusive breeding rights to the female population – the females refuse to mate with any but the dominant male. Participation in this particular hierarchy is what's called an 'Aristotelian categorical' for elephant seals. *Categoricals*, or *categorical behaviours*, are (i) those traits which, when manifested, classify an organism as a member of a kind, and (ii) those behaviours which are typically employed in satisfying the standards of that kind. A tree, for example, *categorically* grows leaves and puts out roots, etc. Male elephant seals, likewise, *categorically* battle for mating rights.

FitzPatrick characterises attributivists as saying that being a good member of a kind means manifesting its categoricals. The problem is that many categorical behaviours, like elephant seals battling for mates, have nothing to do with "the good of [the] organisms, whether considered individually or generally". Battling for mates reduces welfare for the individual elephant seals, says FitzPatrick; even the dominant male is impoverished by this system, as regular battle can lead to significant injury, and shorter life-spans amongst male seals in general. Although the evolutionary role of such behaviours is clear, it is simply implausible, says FitzPatrick, to suppose that elephant seals "fighting desperately with their peers simply to

out-reproduce them are thereby acting 'for their good', or … making themselves better off. On the contrary, it seems elephant seals could … *get on* just as well without these traits – perhaps even better" (Idem 72-73).[58] There is no plausible *benefit* to these traits, says FitzPatrick – the individual animal is hampered, and their society could survive just as well, if not better, in different circumstances.

In short then, the evolutionary-norms argument holds that modern attributive-goodness theories, like Foot's, are non-starters. The standards of biological kinds, or species, are clearly unrelated to the goods of any organisms therein, and so such attributive standards are normatively irrelevant. We can formalise FitzPatrick's argument as follows.

1. According to Attributivism, if *x* is a member of kind *K*, what's good for *x* is what being a good *K*.
2. A good *K* is one that does what a *K categorically* does.
3. Doing what a *K categorically* does can be bad for *x*.

Therefore

4. Being a good *K* is not what's good for *x*.

The following section will explain how attributivists can respond to FitzPatrick.

§3.1.1 – Attributively Good vs. Categorically Exemplified

The objection of the previous section consists in two claims: (i) natural teleology is exhausted by the evolutionary end of gene replication, and (ii) natural teleological standards are not always good for – and sometimes even actively detrimental to – individual organisms. There

[58] Similar points can be made about the Peacock, or the Bird of Paradise, where the males of each species spend resources on fancy displays to win mates, displays that in turn impede their actions, and make it significantly more difficult for the male to survive, to out-compete rivals in finding food, or to avoid threats.

are two ways to respond to the objection, depending on how closely you take those claims to be related. The predominant method of response has taken them to be very closely related, and has responded by addressing (i). That response has argued for the plausibility of *non-evolutionary* natural teleology. For what it's worth, it seems likely that there *are* plausible ways of deriving non-evolutionarily based natural teleologies.[59] I don't, however, think that that has ever been a good line of response. That's because claims (i) and (ii) can also come apart. That the attributive goodness of organisms *qua* species is not plausibly related to the welfare of organisms may be true and troublesome even given some plausible non-evolutionary natural teleology. For example, regardless of how we derive their teleology, elephant-seals are still lifeforms that categorically engage in behaviours that seem detrimental to their welfare. Surely *that* is the salient feature of FitzPatrick's objection, and so I will develop a different response. I will argue here that FitzPatrick's objection (a) incorrectly conflates being attributively good with exemplifying *any* and *all* categoricals, and otherwise (b) begs the question against Attributivism.

In the preceding objection, the thought was that many categorical traits and behaviours are not *good* for the organisms that manifest or perform them. FitzPatrick's view is that they aren't good because they fail to deliver the organism's welfare. To set the standards for when

[59] See Thompson 2004 and 2008, and Lott 2012, 22-23, and 2014, 763, who primarily references Thompson. Their argument is roughly that non-evolutionary natural teleologies are not only plausible, but that we cannot do without them – they are *required for* any understanding of organisms at all. A 'top-down' view is required for us to even understand something as a life-form, to interpret vital processes and characteristic actions, and to evaluate them. Thompson provides a salient example: the same biochemical process – mitosis – counts as reproduction for single-celled organisms, and self-maintenance in the more complex. To be able to make that judgment, we need to be able to understand how mitosis functions in the maintenance of the organism here and now (Thompson 2008, 55). Characteristics such as eating, drinking, mating, etc., rely on an interpretation of an organism as a system wherein certain behaviours represent certain forms of self-maintenance, regardless of how those behaviours came about or were selected for. If we were to understand biological ends in exclusively evolutionary terms, we couldn't classify organisms as doing *anything* but replicating genes.

To the existing arguments above for the salience of non-evolutionary ends in biological evaluation, I would only add the following. It seems implausible to suppose that the standards of organisms can't come apart from their origins, as if the ends of a child must be exhausted by the purposes of its parents in choosing to bear it. Even if organisms are indeed built up *by* genes, *for* genes – vehicles employed for a purpose – that shouldn't mean said organisms must embrace or recognize that original purpose, or that the social ends, desires and standards imbued are any less 'real' – and evaluable – independent of their role in gene replication. A teleology built for an additional purpose is a teleology nonetheless.

welfare has been met, he appeals to our "ordinary, if somewhat fuzzy, conception of organismic welfare" (FitzPatrick 2000, 69).[60] Too fuzzy I think. For sake of ease I'll understand FitzPatrick as arguing that many categorical behaviours are not ultimately *in the interest* of the organisms which partake in them. Even attributivists should want to agree that realising one's attributive goodness will be in one's ultimate interest, so there should be no difficulty there.[61]

A second difficulty however in parsing FitzPatrick's argument is that there are a number of different ways to take it. The objection might be, for example, (i) that certain categorical behaviours are *entirely contrary* to the interests of the organism. It might also be, however, that (ii) even if exemplifying a categorical benefits the organism in some way, it would still be *better* for organisms not to engage in certain categorical behaviours. Each variant of the objection may seem plausible, and each requires a different response.

Take variant (i). This objection argues that attributive goodness is not *good* because categorical facts about organisms – from which natural norms and standards are drawn and in the exemplification of which organisms are supposed to be good – are often *detrimental* to the interests of the individual organisms themselves. That seems *prima facie* plausible. That elephant-seals fight for mates is contrary to their interests in avoiding pain and pursuing self-preservation, yet, supposedly, an attributively good elephant-seal is at least in part one that readily engages in fighting. The objection here then is that we cannot countenance a theory of good that is detrimental to one's interests, and that attributive-goodness is detrimental in that way.

[60] At times FitzPatrick also, seems to think of welfare as in terms of what would be delivered by a *benevolent designer* – see FitzPatrick 2004, pp 72 and 79. I won't pursue that line of thought here.

[61] It's worth noting that some philosophers have, in the course of defending neo-Aristotelianism from these sorts of objections, argued that accounts like Foot's aren't even *intended* to be welfare-based, in what I take to be this vague sense of interest satisfaction. See Hacker-Wright 2009, 312-313. In a very technical sense they are correct, but I have reservations – divorcing goodness from welfare entirely is apt to seem unconvincing and, as I say, even prominent attributivists balk at the idea. Remember Anscombe – "it is a bit much to swallow that a … man in pain and hunger and poor and friendless is flourishing." (Anscombe 2005c, 182)

The flaw in that objection is that it misunderstands the role of categoricals in Attributivism, and how they are thought to connect to interests. It does so in the following way. FitzPatrick's objection characterises Attributivism as saying that exemplifying categoricals – all categoricals – is supposed to be good, and that organisms *ought* to do whatever they categorically do. That though isn't what Attributivism says – at least, not exactly. Attributivists draw a distinction between *types* of categoricals. Some categoricals are what Anscombe called *Aristotelian Necessities*.[62] They are, broadly speaking, what is required for the ongoing maintenance of the organism. Humans, for example, categorically *need* to breathe, act rationally, and maintain healthy social relationships. *Those* are the categoricals a *good* Human should exemplify, because if he doesn't, he won't be able to maintain his way of life. There are many other categoricals though. Humans categorically develop coloured irises, grow hair on their arms, and die when deprived of food. Nobody, however, thinks that exemplifying the latter traits are part of what it is to be a *good Human*. Rather, what Attributivism should be taken as arguing is the following principle.

> **Categorical Relevance.** For any individual *x*, kind *K* and categorical *c*, exemplifying *c* is good for *x iff x* is a *K* and exemplifying *c* is part of how a *K* maintains itself.

In other words, categorical *c* ought to be exemplified by a member of a kind, according to Attributivism, *iff* doing so is required for/conducive to the maintenance of the lifeform. That is what is meant when philosophers like Allyn Fives say "what a person should do all things considered is that which he or she should do because human good *hangs on it*"[63] (Fives 2008, 172). It's not that a person ought to do just whatever is categorically true of humans; they ought to do that on which the human lifeform *depends*. Organisms like humans and elephant-seals are complex

[62] See §2.3.1.
[63] Emphasis mine.

systems of interrelating traits, their good is not manifested in the exemplification of just any particular categorical trait, but by the exemplification of those categoricals *insofar as* they maintain the system.[64]

With the above in mind, we can see the flaw in objection variant (i). Yes it's true that some categoricals are entirely detrimental to organismic welfare. It's false however that Attributivism requires their exemplification. It's true that male elephant-seals categorically engage in violent combat to decide mating rights. *If* it is likewise true that a *good* elephant-seal is one that does so, it's true *because* of how doing so supports other aspects of their lifestyle – because doing so is how an elephant-seal satisfies its social or reproductive dispositions, for example. Were there *no* connection between the elephant-seal's fighting and the further interests of the organism, or if elephant-seals were able to maintain themselves *better* in some other fashion, it *wouldn't* be the case that elephant-seals ought to fight. The objection that Attributivism demands the exemplification of entirely detrimental behaviours is unfounded. [65]

Let me turn to variant (ii) of FitzPatrick's objection. What might still seem troubling is the intuitive thought that it would be *better* for elephant-seals and analogues if they could live another way, and achieve their ends with less trouble. If true, and surely it is, then that seems to lend credence to the thought that the way elephant-seals *are* can't possibly ground their good, because at some possible time they *ought* to act contrary to what is involved in being an elephant-seal. An elephant-seal that could achieve its essential needs *without* having

[64] In fairness to FitzPatrick, *many* philosophers are unclear about expressing this point, including those who attempt to characterize Foot positively. Jennifer Frey, for example, seems *prima facie* to support FitzPatrick's view, when she says "an action, disposition, thought, or feeling is naturally good insofar as it exemplifies the life that is characteristic of the species" (Frey 2018, 48). It's only when she says further that "virtues like justice and prudence are naturally good for human beings, since they are *necessary* to carry out the activities that constitute human life" (emphasis mine) (ibidem.), that the salient details are made clear: that which is good is that which is *necessary* to support the functioning of the system.

[65] There is, in fact, a third method of response that has recently been developed against similar objections. Ulf Hlobil and Katharina Nieswandt have recently argued that attributivists like Foot don't even claim that one ought to exemplify *any* categoricals specifically, but rather that categorical features merely ground what one ought to do (Hlobil 2019). I don't pursue that line of argument here, as I don't agree with their argument that Attributivism doesn't entail, for example, that humans ought to be good humans. My reasons why are explained in ch.6.

to fight would be better off, so clearly their current way of living isn't a good one – they *ought* to do otherwise if given the chance, and so the attributive-good of elephant-seals can't be *the* good of elephant seals.

If the objection is simply that there are more ideal methods for achieving ends than those such entities currently employ, and so their current characteristic means aren't *ideal*, then it's already been answered. Attributivist evaluations of goodness aren't intended to capture some *ideal* way of being, and only if we fetishize the means by which entities achieve their ends is it a problem that some behaviours are worse than they could be. Attributivists have no problem conceding that many categorical behaviours aren't ideal, but they are also only circumstantially relevant to evaluation. That they ought to be changed when possible in favour of more effective alternatives is consistent with Attributivism.

Alternatively though, FitzPatrick can be taken as making this argument: how an entity *ought* to act can't be determined by what it *is,* because if given the chance it ought to choose to *be* some other kind of thing. That argument requires a new response. Here it is: a case wherein an elephant-seal ought to act according to the standards of some other kind would have to be a case wherein the elephant-seal has *become* another kind. Meanwhile, whether or not some organism ought to become another kind would depend upon facts about what is good *qua* the organism's original kind. Thus this objection reinforces, rather than invalidates, Attributivism.

When is it that an elephant-seal should cease engaging in traditional elephant-seal behaviours? I have to think it's when they have developed some new faculty that allows them the capacity to act and succeed otherwise, say the faculty of rationality. A *rational* elephant-seal, however, is in an important sense a very different *kind* of animal than an irrational-seal, just as a man who could fly, who had invulnerable skin and who could see through walls would be a very different *kind* of man – a superman, say. Certainly a man who *became* a superman,

or an irrational-seal turned rational, would be held to new standards – being members of a new kind, the standards of the old have lost their authority. That though has no relevance for subjects that have *not* developed these new faculties.

Should we say that a non-super-man, or an irrational-seal, should act as superman or rational-seals should? Certainly not. That which might be good for Superman to do, after all – say to stand in front of a hail of bullets – might be very bad for Adequateman. And similarly, what might be bad for a rational-seal to do – say to settle social disputes through violence – might be just what an irrational-seal ought to do, having no recourse to rational alternatives. If the possibility of supermen and rational-seals has any normative relevance for members of their less capable counterparts, it's in the sense that it would be very good *for* an irrational-seal to become a rational-seal. This sort of evaluation however means little for the authority of irrational-seal standards for irrational-seals; in fact it relies upon them. If it weren't true that being a rational-seal would be a very good way for an *irrational-seal* to achieve *its* ends – good, that is, by the standards of irrational-seal goals – there would be little merit to the claim that the irrational-seal ought to make the change.

If this interpretation of FitzPatrick's objection is to have any merit, it must make the claim that there are some *kinds* that are simply better to be than others, whether or not there are any common interests to appeal to. That's a difficult claim to support. We aren't, I think, in the habit of directly comparing subjects with entirely disparate ends – it would be very odd to say, for example, that humans are *better* than candles, or books are better than clocks, as though the latter is somehow worse for not being the former, and for good reason: candles and books are fine as is. To comparatively evaluate two subjects we require a common basis of comparison. Attributivism not only provides such a basis, but explains the difficulty we have in comparatively evaluating entirely different kinds of subjects. Kind *x* is better to be than kind *y* only if and when being an *x* represents a better way to achieve *y* ends than being a *y* does.

Put another way: what makes it the case that being a rational kind, say a human, would be a good move for an irrational-seal to make? Plausibly, the ends and interests of the irrational-seal itself. Were rationality not an effective way of achieving *its* goals, whence the normative merit? The claim that rationality is *better* for irrational-seals is explained by and *requires* recognizing them as members of a *kind* with ends which rationality would serve, and appealing either to the standards of the irrational-seal itself. The attributivist hardly has to look beyond attributive-goodness to explain the sort of judgments in question here. For FitzPatrick to insist otherwise, to insist that the capacity to comparatively evaluate *kinds* somehow invalidates Attributivism, must *assume* in favour of some non-attributive-goodness that Attributivism fails to track – one whereby rationality is simply and always good, for example. To do so however assumes what needs to be proved if Attributivism is to be criticised along those lines, and hence begs the question against it.

A large part of the difficulty in responding to arguments of the sort FitzPatrick has offered here is in interpreting them. I have attempted to evaluate what I take are the strongest versions of the objection available. There may be other ways to leverage the preceding objection. Still, I leave it here: FitzPatrick's objection either begs the question in favour of non-attributive goodness, or misunderstands the role attributivists attribute to categorical features of kinds. The objection from detrimental teleology has no merit.

§3.2 – Differentiating Genuine and Incidental Functions

FitzPatrick makes a second argument in *Teleology and the Norms of Nature*, connected to the one just discussed. It goes something like this: (i) attributive-goodness theorists derive functions and ends from actions or behaviours – they observe that organisms, for example, act in certain ways, or bring about certain effects, and so decide 'this is what the organism does', and identify its functions and standards accordingly.[66] (ii) Certain actions can be done incidentally in the course of an entity's operations, and, any functions derived thereby would, as such, be incidental functions, and irrelevant to evaluation. Therefore, (iii) attributive-goodness theorists have no way to distinguish incidental actions from functional actions, and so to distinguish what they should and should not be evaluated in terms of.

Micah Lott, in *Have Elephant Seals Refuted Aristotle*, provides a convenient characterisation of FitzPatrick's argument. Assume that, in the course of inventing a new car-engine, an inventor happens to design an engine that emits vibrations which, as a happy accident, are good for the health of nearby dogs. That is, the engine acts in such a way to improve canine-health. Clearly though, thinks FitzPatrick, it would be wrong to identify this veterinary capacity with the *function* of the engine. The engine's function is to efficiently power automobiles, its veterinary benefit is no function of the *engine* at all and is entirely incidental.[67] That sort of misidentification however – identifying the engine with something the function of which is to improve canine-health – is exactly what FitzPatrick thinks the attributive-goodness theorist is prone to, through identifying standards 'after the fact', without reference to some pre-existing design-schema.[68]

---

[66] FitzPatrick speaks less broadly in his work, targeting Foot specifically, and accuses Foot of deciding function based off 'benefit', as in 'this behaviour is beneficial in the life-form of the species, and so indicates a salient function'. My aim here is to represent FitzPatrick's arguments against Foot as arguments against attributive-goodness theorists more generally. I am reformulating FitzPatrick's claims in accordance.

[67] See Lott 2012, 15.

[68] We might also look, as Foot does, to the example of a tree. It is a categorical fact about trees that their leaves rustle in the wind – it would be wrong, however, to assign to trees the function of generating leaves that rustle, this is incidental. Foot's response is to say that this is because the rustling of leaves plays no part in the life of the

In short, the objection from incidental standards argues that attributive-goodness can't distinguish relevant standards from irrelevant standards if such standards are derived *ahistorically*, from the operations of a subject alone. Attributivism is prone to generating false teleological ascriptions – to identifying functions that are not *really* functions – and so incorrect, or implausibly broad, standards of evaluation.[69]

There are two ways of understanding the problem here. (i) Attributivism has no principled way of distinguishing between what a subject does functionally and what it does incidentally. (ii) Attributivism is apt to assign functionality incorrectly – that is, to assign

---

tree, it has no *benefit*. Even in cases where there is benefit however, says FitzPatrick, it may be equally incidental, and representative of no real teleology on the part of the organism in question. See Foot 2001, 33 and FitzPatrick 2000, 10.

[69] There is also some reason to question whether or not FitzPatrick is entirely correct to call the attributive accounts of neo-Aristotelians *ahistorical* at all. Michael Thompson – whose contributions to the development of a neo-Aristotelian conception of life-forms is important enough that some have taken to calling a particular strain of neo-Aristotelianism the *Thompson-Foot account* (see Lott 2012) – is seemingly indifferent to whether or not historical considerations play a role in determining function, and at times seems very much in favour of the notion. He asserts, for example, that Davidson's *swampman* is nothing but a "mere congeries of physical particles" with "no ears … no brain … no head … no skull … not so much as alive", specifically *because* the accidental nature of its creation divorces it from any 'wider context' by which it can be evaluated (Thompson 2008, 60).

At the same time, Thompson is quite vague about what *does* connect one to such a 'wider context'. Not mere replication or reproduction (Ibid 49-50), nor temporal persistence or success ("[neither] two [nor] twenty years in which … my accidental *Dopplegänger* remains much as [I am] … will manage to hook my double up with any *determinate form*" Ibid 62, also 50-51), nor historical selection or design ("Even if [something] were to bring a certain life-form into being 'with a view to' securing an abundance of pink fur along the shores of the Monongahela, this 'purpose' would have no effect on the inner natural teleological description of that form of life" (Ibid 79)). In fact what seems to be Thompson's point is that there is *no* way of arriving at this context through consideration of entities or objects themselves ("we are wrong to think of the concepts of the various life-forms as reached through abstraction from features of their particular bearers", (Ibid 59)). Rather, some life-form concept must be *assumed* and *imposed* in a sort of 'top-down' fashion, before individual actions can even begin to be classified in any way (Thompson 2004, 52).

Is this sort of *pre-supposed* life-form concept *ahistorical* in the way FitzPatrick dislikes? It's not immediately clear that it is. For Thompson, being a member of a 'life-form' seems to involve some sort of causal connection to a wider system (he follows Anscombe here: "Oaks come from acorns, acorns come from oaks; an acorn is thus *as such* generative (of an oak)" (Anscombe 1981c, 87) and "In thinking of something as an acorn, we tie it specifically up with *oaks* … and so … we 'look beyond' the individual lump of stuff" (Thompson 2008, 54)). Foot seems to agree – her analyses are not restricted to the operations of discrete individuals, but to "the life of … individuals that belong to [some] species [over] a certain time" (Foot 2001, 32). However, she also rejects a view of 'species' as "gradually developing … organism[s] [that] stretch for millions of years" (Ibid 29). The point then for the *Thompson-Foot* account seems to be that, while membership in a species involves some degree of causal/historical participation in a sort of biological community, whether or not members of that community are themselves members of the species is decided by how well they accord to pre-supposed or stipulated conditions, not by participation in some empirically classificatory biological process. I suspect this won't appease FitzPatrick – this sort of stipulation still seems to imply the sort of arbitrariness he seems to find distasteful – but it may nevertheless be important to note that historical considerations *do* play a role in neo-Aristotelian life-form evaluations.

functionality to non-functional behaviours. These two variants deserve two responses, which I develop in the next two subsections.

In §3.1.1, against (i), I will argue that the objection naively conflates the idea of being *causally connected to x* with the idea of *doing x*. While Attributivism *does* derive a subject's functions from what it does, it is possible to distinguish between some act or effect *x* as *part of* a subject's operations – i.e., *entailed by*, *necessary for*, or *constituted in* them – and *x* as only causally connected to them. I argue that observing that distinction provides the tools to identify what a subject does only incidentally.

In §3.1.2, against (ii), I argue that the objection relies on an account of function the attributivist has no reason to accept. The belief that Attributivism fails to track *genuine* functions stems from an implicit endorsement of *selected-effect* or *historical* accounts of function. I will argue however that Attributivism employs an *ahistorical* or *systemic-capacity* account of function. It is likely true that Attributivism lives or dies on the plausibility of such accounts, but since such accounts *are* plausible, insisting that Attributivism fails to (exclusively) track genuine function because it fails to (exclusively) track *historical* function is baseless.

§3.2.1 – Doing vs. Causing

Proponents of objection form (i) argue the following. If functions are drawn from what a subject *does*, then functional ascriptions are implausibly broad – subjects can *do* all sorts of things, often just incidentally, and if we allow that all of these *doings* impose normative standards, then normative evaluation becomes absurdly broad. The flaw in that argument is that it conflates two senses of *doing*, which I'll now distinguish.

There is one sense of doing whereby, when we say that *x is doing y*, we mean that *x* is *causing y* to happen. That sense of doing is certainly irrelevant to determinations of function.

We *don't* want to say that *x* is functionally *aimed* at doing whatever *x causes* to come about, because what one causes to happen can be incidental. There is another sense of doing, however, whereby when we say that *x is doing y*, we mean that *x* is explicitly acting to accomplish *y*, i.e., that the end of bringing about *y* is in some way *part of x*'s behaviours. In that case, we *do* want to say that *x* is *aimed at y*. This is, roughly speaking, the distinction between playing catch, and accidentally breaking a window while doing so. One of those things – the latter – is something I may have done, in the sense that I *caused it*, but the former was something I was *actively doing*.

The preceding objection seems to assume that there is no way of distinguishing between these two senses of doing, and so that, if we derive functional aims from what one *does*, we must say that one aims at what one does incidentally. There *is* a way to distinguish these senses however. The clue is in the terms I've used. If we want to know whether or not accomplishing *y* is an *aim* of *x*, we should ask whether accomplishing *y* is *part of x*'s behaviours, such that if *y* is *not* achieved, some corresponding change in *x*'s behaviours will be observed. For example, to keep me from throwing a ball in the air, you would need to affect a change in my behaviours. To keep me from breaking a window, you might not involve me at all, but merely shatterproof the window. Thus we can see that one of those actions, throwing the ball, is something that is *part of* my behaviours, while the other, breaking the window, is only incidentally involved. Micah Lott provides an example very much along those lines.

Lott (2012) imagines a tree, the branches of which grow so high and thick that flocks of birds nest within them. How do we know if that trait is a functional end of the tree – such that the tree would be defective if its branches were insufficient for birds to nest within – or merely incidental? Lott says the answer depends on whether any other behaviours in the tree depend upon supporting nesting birds. If it turns out that nothing else the tree does hangs upon nesting birds, then the trait is incidental. Alternatively, if, say, the roots of the tree are fed by

the droppings of the birds, which in turn allows for the maintenance of the branches in which the birds nest, the test tells us we have identified a functional end within the tree's system, precisely because we have identified a system of *reciprocal dependence*. Both supporting bird nesting and taking nourishment through roots are functional ends of the tree, as neither can be eliminated without impacting the other, within a complex system of mutual dependency.

Lott and the *reciprocal dependency test* capture the sort of general principle I argued for in the preceding. Roughly speaking we should say that *x* is functionally involved in *doing y* only when *y* itself is so closely integrated with *x*'s behaviours that the cessation of *y* will involve a corresponding change in *x*.

Consider how the above allows the attributivist to resolve the problem of the engine/veterinary-device, provided in the previous section. Here is the engine, whirring away, providing locomotive force and emitting vibrations that invigorate nearby dogs. Are the engine's behaviours such that taking in provided fuel and transforming it into locomotive force is *part* of them in the established sense – inseparable without affecting change within the system? As it happens, yes: the various operations of the object are constitutive of the end in question. As long as they are functioning as they are, fuel provided *will* be transformed into force, and failure to do so will invariably accompany some breakdown in the system of operations. The end cannot come apart from the means; the capacity to transform fuel is constituted by the operations of the object being considered, it is *part of* those operations, and so we have an engine.

Similarly then, we should ask: are the operations of the engine such that they similarly entail the improvement of canine health? As it happens, I'm inclined to say no. It's very easy to imagine that the engine might cease or fail to improve canine health without its operations being impacted by the lack of the effect in any way. The dogs might be taken away, or their biology perhaps slightly altered such that the engine's vibrations no longer benefit them, and

the engine's operations would continue unperturbed. And so it turns out that attributivists can determine, in a principled way, the function of the engine – it *is* to transform fuel to force, and it *isn't* to invigorate dogs.[70]

Surprisingly, few attributivists have argued along the lines I have presented here. Yet precedence *is* found, largely among philosophers of biology and function. Davies, for example, argues that functionality can only be ascribed to a trait or behaviour *within* organised, hierarchical systems of such, relative to how a given trait contributes to the maintenance of the whole. A heart, for example, both distributes blood within the cardiovascular system, and creates vibrations in the sternum, but only the former behaviour is a function of the heart, as the latter makes no contribution to the capacities of any system (Davies 2001, 77-79).[71] That reinforces the idea that functional ends of organisms should be restricted to those operations which are *part of*, i.e., causally efficacious within, systems of operations.

With the above distinction in hand, I'll say that attributivists do have a way to differentiate functional ends from incidental effects, via the *reciprocal dependency test*. The test works, and so defuses variant (i) of the objection at hand.

---

[70] It might seem worth objecting here that the engine only operates as an engine under the condition that fuel is provided, but how do we know that it is *supposed* to receive fuel at all? I think this objection is confused – I don't think there's any sense to asking under what conditions some object is *supposed* to operate prior to establishing some pre-existing operations or dispositions in the light of which such judgments are to be made. Many philosophers seem to agree. See Thompson, who argues that the 'normal', 'ordinary' or 'standard' conditions for an organism are 'presupposed' by a prior understanding of a how a given life-form behaves, (Thompson 2008, 70-72). Similar views are espoused by Richard Levins and Richard Lewontin, who contend that what constitutes an organism's environment depends essentially on the life of the organism itself, and Buller, Walsh and Ariew, who all assert that characterisation of environments as hostile to or part of an organism's operations comes *subsequent to* an understanding of the organism (Levins 1968, Levins and Lewontin 1985, Lewontin 1991, Walsh and Ariew 1996, 508-509, Buller 1998, 511). Even noted selected-effect theorist Ruth Millikan argues that the 'normal' operating conditions of an organism are drawn from those conditions which must be mentioned in a description of the organism as performing its functional operations (Millikan 1984, 33). Millikan of course disagrees with attributivists over how such functional behaviours are derived, but in this case the point remains the same.

[71] In this vein, Davies continues a tradition of identifying organisms with complex systems, and interpreting their functional behaviours in terms of the role those behaviours play in maintaining or developing the system. See also, as already referenced, Thompson (1995, 2004, 2008), or Boyd's' 'homeostatic property cluster' analysis of organisms (Boyd 1999a, 1999b).

§3.2.2 – Ahistorical Functions

The previous section established that Attributivism has the capacity to draw a principled line between incidental and functional actions. Objection variant (ii) remains. That objection isn't that Attributivism can't draw a principled line between functional and non-functional actions, but that Attributivism identifies the *wrong* actions as functional. Specifically, philosophers like FitzPatrick and Odenbaugh argue that there is a *correct* functional analysis that we can discover when we look to the *history* of a life-form, and discover the reasons why certain behaviours were selected for. They argue that Attributivism assigns functions to organisms that don't correspond to the historical picture, and so aren't *really* functions. In this section I argue that objection variant (ii) privileges historical function analyses without justification. No independent argument for doing so is given, and the dialectic in philosophy of function doesn't decisively support doing so. I argue that Attributivism can plausibly employ *ahistorical* function analyses instead, and that to insist otherwise begs the question. An in depth evaluation of the accounts of function available goes beyond the scope of this thesis, but I will end this section by briefly sketching the core ideas behind the account Attributivism should be committed to, and the reasons why it isn't an implausible one.

FitzPatrick argues against Attributivism by providing cases where, he thinks, Attributivism fails to track what is clearly the *correct* functional analysis. One such case involves the behaviours of specific birds. Swifts reduce the size of their egg clutches in times of scarcity. Suppose we want to know what function this serves, and correspondingly how swifts should be evaluated. FitzPatrick thinks there are two possibilities: (i) this behaviour has the function of contributing to the survival of the species, allowing more resources to be available to the group, and thus swifts are cooperative animals and should be evaluated thusly. (ii) This behaviour has the function of *maximizing personal reproduction* and reducing the cost

to the individual in its operations; it serves a personal or selfish end and so the swift is a selfish animal, and should be thusly evaluated.

FitzPatrick thinks either of these stories is plausible, and further thinks that, by accepting *ahistorical* function analyses, Attributivism is committed to both of these interpretations. He may be correct – reciprocal dependency will justify both interpretations, *if* it turns out that failing to act cooperatively and failing to manage reproductive costs would both impede the maintenance of the swift's operations. Yet FitzPatrick argues that, regardless, only *one* of these function ascriptions can be correct. He thinks that's because when we look to the history of the life-form, we will find that the *reason* the trait of clutch-reduction is present owes to only one of the two plausible evolutionary accounts – it is *because* clutch reduction reduces the burden on the individual and so promotes survival, and not because of group pressure, that swifts today reduce clutch size. So what the swift is *really* doing when it reduces clutch size is acting selfishly, and they are *not* cooperative in this respect (FitzPatrick 2000, 194-199). According to FitzPatrick then, Attributivism is committed to functional analyses that, given the historical evidence, are *wrong*, and so is apt to impose incorrect evaluative standards.

FitzPatrick can be parsed as saying that Attributivism doesn't accurately track the *real* function of organisms. According to Attributivism, a swift's behaviours may be functionally cooperative *and* functionally self-serving. But even if they do have principled, discriminating reasons for thinking so, all that proves is that their principles are *wrong*, because only one of those behaviours is *really* functional. Specifically, FitzPatrick thinks only the behaviour that is present due to historical selection processes is *really* functional. But why? It isn't really explained – rather FitzPatrick just seems to take it as obviously true. Nor, surprisingly, is he the only one to do so. Jay Odenbaugh makes a similar argument, that ahistorical function analyses can't identify natural normative properties, because ahistorical theories don't provide

proper functions, functions that a subject *ought* to perform. But he thinks so specifically because he defines a *proper* function as one that is grounded in historical selection processes – he thinks that a member of kind *x* only functionally *ought* to perform function y if previous *x*s also did so – without offering any further reason to believe so (Odenbaugh 2017, 1043). In the absence of such a reason though, there's no reason to accept the argument.

What variant (ii) of the objection has devolved into is pure ideological disagreement. Philosophers like FitzPatrick and Odenbaugh assume the exclusive truth of *etiological* or *selected-effects* theories of function, where the functional end of a trait or behaviour is exclusively that "*for the sake of which* the trait is manifested" (FitzPatrick 2000, 198), or, as per Philip Kitcher, "the function of *S* is what *S* is designed to do" (Kitcher 1993, 380).[72] Attributivists though obviously reject that assumption. As we've seen, they determine the function of an action *ahistorically*, in terms of its role in the ongoing maintenance of the organism.

For variant (ii) to be a valid objection, it must be argued that attributivists *aren't* entitled to their view of functionality. And to make *that* argument at this point, one would need to attack ahistorical analyses themselves as incoherent or otherwise implausible. They aren't, as I'll argue below, and so attributivists have at least equal claim to the validity of their functional analyses as do their opponents. Although this is unfortunately not the place to engage in an extended foray into the philosophy of function, I think that even a brief look at that particular philosophical landscape will confirm that the sort of ahistorical account attributivists require is

[72] For modern accounts of selected-effect theories, see Millikan 1984, 1989a, 1989b and 2002, Neander 1991a and 1991b, and Sullivan-Bisset 2016. See Buller 1998 for a review of modern function theories generally, and Davies 2000 and 2001 for critiques of selected-effect accounts. Interestingly, FitzPatrick doesn't take himself to be advocating for a selected-effect or etiological account of function. He objects to common etiological accounts for their 'atomistic reduction' of functional teleological facts to causal history facts. For my purposes here though the difference is unimportant: FitzPatrick *does* see etiological accounts as being 'on the right track' for acknowledging the importance of natural selection. See FitzPatrick 2000, 9-10.

secure. To be as thorough as possible then, I'll now close with a *brief* review of the relevant material.

Attributivism should be seen as employing what's known in the philosophy of function as *systemic-capacity* analyses, advocated for originally by Robert Cummins.[73] The basic idea of systemic-capacity theory is to ascribe functions to traits and behaviours in terms of the capacities of systems they constitute. A systemic-capacity theorist begins by observing some phenomenon – say the keeping of time – then posits the existence of a system that accounts for it – say a clock – and assigns functions to aspects of the system – say cogs and gears – in terms of how they contribute to its overall capacity to do so (Cummins 1975, 1977, 272 and 1983, 31). That should seem very similar to what we've already seen of the attributivist process, re: assigning functions to traits in terms of their role in the maintenance of the organism. Moreover, systemic-capacity theories share the commitments of attributivists like Thompson, who say that there is no way of empirically distinguishing between genuine functional traits and other effects, without presupposing a wider context or candidate system within which analysis should take place (Cummins 1975, 751-756).[74] Attributivism and the commitments of systemic-capacity theory go hand in hand.

Regarding plausibility, the position of several recent authors has been that not only do systemic-capacity theories remain coherent alternatives to selected-effect theories, but even that the latter must incorporate aspects of the former.[75] Rather than canvas all those sources

[73] See Cummins 1975, 1977 and 1983, Prior 1985, Amundson and Lauder 1994, and Davies 2000 and 2001.
[74] Specifically, for Cummins, this candidate system is presupposed by the interests we take in certain phenomena as researchers, but if there's an argument for why the sort of presupposition attributivists advocate – whereby candidate systems are identified in response to something *being done* – shouldn't also be valid, it's unclear. FitzPatrick seems to offer something like that sort of argument – he says that neo-Aristotelians are focused on *goods* and *needs*, and seems to think that these are the sort of thing to be presupposed. Little more is said however, and it isn't clear that attributivists should agree: *needs*, it's plausible to say, are only derived after systems and their objective interests are presupposed. See FitzPatrick 2000, 8.
[75] See Buller 1997 and 1998, 506-511, Walsh and Ariew 1996, 508-509, and Davies's 2000 article, tellingly titled: *Why Selected Functions are Systemic Capacity Functions.*

however, it will be more expedient to focus on why some philosophers have *questioned* systemic-capacity accounts of function, and to explain why their arguments are unconvincing.

Arguments against systemic-capacity theories have generally focused on a supposed inability of such theories to accommodate *malfunction*. If a system's function is just whatever it actually *does*, then how can it ever *fail* to function? If function is predicated on current behaviour, argue ahistoric-function sceptics, then there is no sense to suggesting that the current behaviour of a system ought to be other than what it is. Davies summarises nicely: "[a malfunctioning] T is no longer a member of the category of Ts, *if* the category is defined by reference to some systemic capacity" (Davies 2001, 199).

Historically many etiological function theorists have been convinced by arguments like the above.[76] Yet their appeal has waned, due both to attempts at rescuing malfunction, and to observations that etiological accounts suffer the same deficiency. Davies, yet again, argues the latter point:

> "if functional types are defined in terms of historical success, then tokens that lack the defining property due to defect … disease or damage, are excluded from the functional category. Historically based malfunctions, in consequence, are impossible" (Davies 2000, 19).

As to the former, I can't possibly canvas here all the proposed methods of rescuing malfunction and thus normativity. I will say though that many philosophers have found it plausible to assign functions to systems that lack the dispositions we would normally derive those functions from, by referencing homologous systems in counterfactual scenarios.[77] One might, for example, decide that a heart that doesn't pump blood is acting defectively, by observing how homologous

[76] See Millikan 1989b, 299, and Millikan 1993, 32: "a crippling defect in any definition that looks for function in current dispositions … is that such dispositions cannot ground the notion of malfunction".
[77] See Amundson and Lauder (1994) or Davies – "if [a] category is defined … in terms of homologies … [a malfunctioning] token may retain membership in that type" (Davies 2001, 199).

cardiovascular systems are supported by hearts which *do* pump blood, and extrapolating the aim of that particular organ *within such systems* accordingly. Even some etiological function theorists have taken similar approaches – although they point explicitly to *hereditable* physical traits shared between candidate defective system *a* and homologous functioning system *b* to justify interpretations of malfunction, the sentiment seems largely the same. They interpret a system as *attempting* to instantiate some capacity in virtue of possessing certain traits in common with systems that do.[78] In any case, it seems plausible that ahistorical function theories *can* accommodate malfunction, or are at least no *less* capable than their historically-oriented rivals.

Despite how briefly I have had to summarise the function landscape, I hope I have at least made apparent that there is no obvious reason to believe that the function analyses Attributivism commits to are any less capable of establishing genuine normative functions than alternatives. Given then that attributivists have plausible reasons for taking their functional analyses to track genuine functional operations, and that their method of analysis is coherent and plausible, it's safe to deny the objection to the contrary.

To summarise, critics have argued that attributivist commitments to *ahistorical* accounts of function renders them either (i) completely unable to non-arbitrarily differentiate between genuine functions and incidental effects, or (ii) unable to make the *correct* distinction. In response I presented how attributivists can distinguish between behaviours that are *part of* a system's operations, as opposed to merely causally connected to them, and how the resultant *reciprocal dependency test* allows attributivists a principled method of demarcation. I argued also that a further objection, that the reciprocal dependency test fails to draw the *right* line between genuine and incidental functions, relies on an unwarranted insistence that only etiological, selection-based theories of function can ground *proper* functions. I argued that

[78] See Sullivan-Bisset 2017.

attributivists have at least equal justification for favouring their approach to functional analyses as do their opponents. The objections levelled by FitzPatrick and others are useful in illuminating just what sort of functional analyses attributivists are committed to, but they don't identify a problem. Attributivism is secure as long as systemic-capacity accounts are plausible, and to all appearances, they are.

§3.3 – TELEOLOGICAL EVIL

The objections I have discussed so far have been predicated on a common complaint, that attributivists, and particularly neo-Aristotelians, rely on a teleological conception of species that "makes questionable sense in the context of modern science", and that the prospect of drawing morality from modern biology is "more or less uncontroversially forlorn".[79] Those objections targeted attributive-goodness at the core, rejecting the plausibility of Attributivism outright.[80] I've argued that those objections *aren't* as damning as they once appeared, and that Attributivism is entitled to its teleological suppositions. There remains, however, room to object to modern attributive-goodness projects even after *accepting* their view of biological teleology. The most prominent of this sort of objection I call the objection from teleological evil.

In the simplest terms, the objection is this: attributivists propose to ground traditionally moral virtues like justice and faithfulness in natural teleology, but the same teleology can ground – and seemingly justify – moral vices. Attributivism operates by locating teleological standards in organisms, and identifying goodness with the satisfaction of such standards. To satisfy one's functional standards is to be good. Critics reject that claim. Elijah Millgram objects, in a critical review of Thompson's *Life and Action*, that to believe such is to maintain an unjustifiably – even blindly – optimistic interpretation of natural teleology – something he calls the *Polyanna problem*.[81] In actuality, satisfying one's functional standards will often produce *bad* results, both from external and internal perspectives.[82]

---

[79] See for example Hull 1998, Kitcher 1999 and, for the source of the quotes, James Lenman's 2006 review of neo-Aristotelianism in his SEP article, Moral Naturalism.
[80] Although FitzPatrick's accounts are the most clear and detailed, he is echoed by others: Adams, as already mentioned, and Joseph Millum in 'Natural Goodness and Natural Evil' to name another.
[81] See Millgram 2009, 561-562.
[82] See also Scott Woodcock – "Foot's proposal … sanctions prescriptive claims that are clearly objectionable" (Woodcock 2006, 445).

Millgram is not alone in developing that sort of objection. Numerous exercises in biological philosophy have, over time, pointed out that biological systems often tend to develop as much through destructive processes as anything else. Hursthouse herself points out that categorical characteristics of many animals involve behaviours of questionable virtue. Cuckoos categorically exploit other animals; male polar bears abandon their young and male cheetahs neglect their mates.[83] Likewise, female sharks devour their young and male lions slaughter foreign cubs then assault their distressed, yet physiologically aroused, mothers. Animals that act in such ways act in accordance with categorical standards, yet it would be strange to call that *good*. Nor is this sort of observation restricted solely to *non-human* animals.

Thornhill and Palmer (2000), Hrdy (1999) and Frank (1985) have likewise all argued that *human* nature endorses a range of vicious values and behaviours, including rape, infanticide and political injustice. Millgram appeals to such studies to argue that straightforward natural teleology is as apt to endorse vice as it is virtue. Chrisoula Andreou does the same in 'Getting On in a Varied World' (Andreou 2006), although Andreou's claim has an additional dimension. On top of the claim that natural teleology conflicts with traditional ideas of moral action, she argues that natural teleological standards are apt to be *internally* contradictory – to promote the development of traits in organisms that are bad even by their own relative standards. Andreou makes much of the phenomenon of *polyphenism*, whereby organisms can have numerous genetically determined developmental 'paths' that can be triggered by early environmental cues. Andreou claims that there might be 'mixed naturally sound types' for any given species. What this means, in short, is that the categorical

[83] See Hursthouse 1999, 220-221. See also James Lenman 2005, 45. Lenman echoes Hursthouse's observation, and is likewise critical of 'natural goodness' approaches to normative realism. Of course, Lenman doesn't merely think that attributivists wrongly identify bad actions as good, but that we shouldn't treat these sort of observations as normative evaluations at all, but merely as exercises in classification. Nevertheless he seems to endorse the idea that, if we *are* to evaluate these behaviours, we should oppose neo-Aristotelian conclusions, which endorse behaviours that have no claim on being *good*, and decry behaviours that are certainly not *bad*. See Lenman 2005, 46.

characteristics of a species might entail divergent teleologies, or different and contrasting developmental paths. Andreou thinks this produces paradoxical situations, wherein an organism's own teleology both endorses and condemns the same traits, as they are successful and virtuous by one sound developmental path, and defective and vicious according to another.[84]

Andreou employs many examples which I'll largely omit here. Briefly she thinks that, as some insects might be 'programmed' to develop near-sightedly or far-sightedly in response to environmental cues, with neither path being *defective*, so too might humans go with traditionally moral traits. She thinks it conceivable that human teleology could endorse the development of sociopathic or cruel behaviour, in the right circumstances, while being neutral on what circumstances are appropriate. This isn't merely to say that human virtues may consist in behaviours that are immoral by objective standards, but to say that human teleology, containing multiple sound developmental possibilities, can *itself*, simultaneously, endorse opposing developmental schemas, and so both endorse and condemn the same traits in accordance with the whims of environment. Human teleology may recognize selfishness as a vice – according to the soundness of a schema wherein charity is a virtue – and also as part of a sound schema, and thus virtuous. Just as a naturally sound birds may act both maternally and unmaternally at different times, so too might a naturally sound human see as virtues both justice and injustice, or honesty and deceit.[85] This is taken to present an internal incoherency – something cannot be simultaneously bad and good, a vice and a virtue.

In the end, whether they rely on objective or relative standards, these authors grant the view of attributive-goodness theorists – that natural teleologies can be identified – but deny that natural teleology delivers the sort of virtuous conclusions that are sought. Natural

[84] See Andreou 2006, 65-72
[85] See Lack 1968 if you're interested in birds.

teleology endorses behaviours that are, it seems, simply *unconscionable*, and that we surely shouldn't countenance.

§3.3.1 – Natural Indifference

I've outlined two variations of the Teleological Evil objection. The first – call it the *Millgram* or *Natural Evil objection* – claims that natural teleologies endorse immoral behaviours. The second – call it the *Andreou* or *Polyphenism objection* – argues that natural teleologies can be internally incoherent, advocating *and* condemning the same behaviours simultaneously, and being thus unsuitable for grounding moral evaluations. I intend to focus primarily here on the latter objection. I argue that the polyphenism objection relies on an insufficiently relativized understanding of what polyphenic teleology promotes, and is resultantly mistaken in claiming that polyphenic development results in contradictory natural norms. First though, I'll explain why I *won't* be focusing extensively on the Natural Evil objection.

The Natural Evil objection is predicated implicitly on acceptance of reflective equilibrium. That is, it advocates against Attributivism on the grounds that Attributivism entails consequences that clash with strongly held moral intuitions. It implies thereby that theories of the normative, if they are adequate, should corroborate our considered and persistent moral intuitions, and produce conclusions that cohere with them. I don't take the Natural Evil objection to be a particularly significant objection *against Attributivism*, because I don't take attributivists to be aimed at producing reflective equilibrium. *Some* are, certainly, but at its core Attributivism is inherently revisionary – it got its start in the rejection of prevailing moral ideologies, and attempts to begin not from intuitive starting points, but from observable facts about psychological and biological dispositions. That such evidence may demand revising our moral intuitions is less an objection to Attributivism than a feature of it. I am more interested in objections that attributivists themselves can't reject on ideological grounds, objections like

the polyphenism objection, which target the logical coherency of Attributivism, and so pose a problem even for those willing to accept significant revisions in their moral commitments.

That isn't to say that there is definitely no merit to objections like the Natural Evil objection. It may be that requiring us to revise our views on topics like murder or rape is just too much to ask. There may be an abductive war to be fought there. Yet even if attributivists must respect reflective equilibrium to some extent, it remains unclear *which* intuitions they should respect. They might choose moral intuitions. They might alternatively respect intuitions in favour of specific principled methodological commitments, and believe that if, say, those commitments lead to the *bizarre* conclusion that murder is virtuous, that would be evidence that we have radically misunderstood what murder is, and how it relates to what's worth pursuing. Nor is it obvious that attributivists *do* have to respect reflective equilibrium. Prominent philosophers like Brandt and Hare have objected to that principle on 'garbage in, garbage out' grounds – arguing that coherence among intuitions is meaningless without some impartial, objective reason for taking those intuitions as correct in the first place.[86] I tend to agree. For all the above reasons, I take the polyphenism objection to be the weightier and more worthwhile objection, and will turn now to it.[87]

In the polyphenism objection, Andreou imagines an organism with teleological dispositions that reflect multiple possible environments, such that it is disposed to develop very different and mutually exclusive attributes given certain environmental cues. To draw out the analogy with human moral traits, we can imagine that *humans* are an organism that operates in just this way: when given, say, a supportive family environment it would be entirely

[86] See Hare 1973, Brandt 1979, and, for the 'garbage in, garbage out' formulation, Jones 2005, p.74.

[87] As many readers have pointed out though, if one is going to reject moral intuitions as the basis for evaluating moral theories, one will need to propose some *alternative* basis for evaluation. My own view is that the best way to evaluate moral theories is by evaluating normativity itself, as to what metaphysical premises must be true for it to be the case that failure to Φ constitutes some *all things considered* failure on the part of the agent, and then testing moral theories for adequacy on those grounds. The subsequent chapters of this thesis proceed on just those grounds.

appropriate for a human to develop traits of benevolence, honesty, cooperation, empathy … all those traits so commonly considered *virtuous*. At the same time, however, it would be entirely appropriate for a human belonging to a hostile social environment to develop traits of deceit, aggression, sociopathy and selfishness. What this means – says the *polyphenism objection* – is that human teleology stands to call *both* honesty and deceit, or empathy and sociopathy, virtues – traits humans *ought* to develop in order to flourish. The problem though is that recognizing any of these traits as virtues implies recognition of the opposite trait as vices – traits humans ought *not* to develop. We can't, it seem, have it both ways, and so human teleology – or biological teleology in general – can't ground a coherent account of virtue, or of what traits we should or should not pursue.

I think that the *polyphenism objection* is mistaken in what it takes polyphenic teleology to endorse and to decry. The proponents of this objection are no doubt right to point out that biological development is or can be a complex and subtle process – it is *fine-grained*, prompting different developmental paths in response to different environments. Puzzlingly though, these critics don't develop an equally fine grained account of biological teleology, vice or virtue. While developmental goals and needs are taken to be relative to environmental queues, virtues – which are supposed to be drawn in response to the needs and ends of the organism – are treated somehow as indifferent to environment or circumstance. Although it is never suggested that, for example, an organism should develop as *both* near-sighted and far-sighted, or that a human should be *simultaneously* empathetic and sociopathic, we're expected to take human teleology as endorsing empathy and sociopathy as *simultaneous* virtues, as suggesting that *both* ought to be pursued, and thus contradicting itself. This seems wrong – rather, just as we take polyphenic development to endorse *either* near-sightedness or far-sightedness, polyphenic teleology should be taken to endorse *either* honesty *or* dishonesty, empathy *or* sociopathy. This is to say polyphenic teleology endorses *divergence* – at least

originally – under which an organism ought to be *either* one way or another, depending on its environment, but never *both* at once. What we get from natural teleology is virtue relativized to conditions, not virtue full stop. This no longer seems inherently contradictory though – is there still a problem here?

If there does seem to be, I imagine it's for something like the following reason: even if we take polyphenic teleology to endorse *either* of two mutually exclusive traits, this still amounts to the tacit implication that both are *good*, and so that their negations are *bad.* Yet each trait is the negation of the other, and so we must assert at the same time that each trait is both good and bad – an incoherent position that entails that both developmental paths ought and ought not to be pursued simultaneously. There's no more sense to this arm of the objection than there was to the previous however. Just as virtues are relativized to circumstances, so too should be vices. The thought that the negation of a circumstantial virtue must be a vice full-stop is easily refutable. A virtuous agent may eat heartily when, say, preparing for strenuous activity, but that's hardly to say fasting is therefore always and only vicious. Rather it might *also* be virtuous for them to fast or go hungry at other times, say when trying to lose weight, or when protesting the injustices of British rule in India. The point is that the viciousness or *badness* of a given trait is likely to depend on details of the environment or subject, and in cases of polyphenic development, it only makes sense to label any of the potentially viable behaviours as bad *after* development has diverged, and one path has been selected. Again though there is no conflict here, as at this point it makes little sense to say that both traits are equally viable. Once empathy has been selected for by the environment, sociopathy is no longer an option. Rather than being a potentially virtuous development capable of grounding the viciousness of empathy, sociopathy is simply out of the picture.

Just as polyphenic development occurs in stages, so do the corresponding evaluations of virtue and vice. In a polyphenic organism *pre-divergence*, while it is true to say that multiple

mutually-exclusive traits may represent virtuous behaviours – in that they promise to deliver the flourishing of the organism – it is wrong to say that the incompatibility of the traits presents a contradiction. This is because it is never the case that these virtues are intended to be realized *simultaneously*, or relative to the same circumstances, and prior to one path being selected, there is no cause to label either one as running counter to the interests of the organism. *Post-divergence* there is still no contradiction, as at this point the developmental paths have been settled. The original teleology has been overridden, and it's no longer true to claim that the alternative development traits are still viable options – they are no longer virtues, only vices, as they retain no connection to the organism's flourishing.

To conclude, in this section I have given a brief response to two varieties of the 'teleological evil' objection. I rather summarily – though I hope not uncharitably – dismissed the Natural Evil objection, as it is founded upon commitments that attributivists might coherently reject. I have instead focused on the 'polyphenism objection', or the claim that biological teleology is inherently contradictory. I have argued that the view of virtue and teleology employed in this objection is insufficiently fine-grained, and that once we properly relativize the virtues and vices of organisms to their teleological aims in their current stage of development, the apparent contradiction disappears.

§3.4 – THE RATIONAL KIND

In an insightful paper – 'Why be a Good Human Being?' – Micah Lott develops what I see as the strongest of the recent objections to modern versions of attributive goodness, which begins with what he calls the *Authority of Nature Challenge* (Lott 2014). In his paper, Lott demonstrates a keen and sympathetic understanding of neo-Aristotelian theories, and of a potentially critical flaw. He makes much of the role of rationality in moral and practical reasoning; he grants a common thought that acting morally is supposed to be *rational* – that ethical, normative facts should provide *reasons* for acting. That fact, he claims, poses a problem for Attributivism.[88] An irrational animal or some such may have its reasons exhausted by facts about its biological kind – it is a different story for rational creatures:

> Our rationality enables us to step back and ask whether we *should* live in the way that is 'naturally good' for us. And this question can only be properly answered with a *reason* – i.e., with a point about why it is good to live one way or another. An answer that appeals simply to our 'nature' is of the wrong sort, since it just re-asserts that this *is* the way things are with us, rather than giving an answer to the question of *why* we should embrace this way or depart from it when attempting to live and act well. As an answer to [that question], an appeal to our 'nature' will only be relevant if it is supported by [a further appeal to reasons – reasons why our nature is *good*]. Thus [natural goodness] cannot account for the rational authority of morality (Lott 2014, 766-767).

In essence, Lott repeats what, by now, should seem like a familiar question. Much like the sentiment expressed in earlier sections on normative motivation, Lott asks why – even granting a sort of natural normativity for species – should we *care* about being good human beings? More specifically, what reason is there for us, as *rational beings*, to care about the natural goodness of *humans* as a species?

[88] Lott once again focuses on neo-Aristotelianism specifically, Foot and Hursthouse primarily. I am repurposing his arguments slightly to address Attributivism more generally.

Lott claims that being rational provides us with a new faculty, the ability to 'step back' from our natural or biological situation, and evaluate whether or not this is, in fact, a *good* way to be. He thinks modern attributive-goodness theories have a problem in answering this without appealing to reasons that go beyond the internal reasons of natural kinds.[89] Most of this, as I say, should seem familiar in the context of the earlier *normative problem of motivation*, but Lott doesn't stop there.

Lott endeavours to develop, on behalf of the neo-Aristotelian, what he takes to be the most promising response to his own challenge; it is a response that he thinks demonstrates a particular difficulty. In what he calls *the Practical Reason Response*, Lott focuses on the role that practical rationality plays in both Foot and Hursthouse's accounts of natural normativity. Both Foot and Hursthouse appeal to the role of practical rationality in human life. They think human standards are significantly more varied and complex than those of other life-forms, because we are *in essence* rational creatures.[90] Practical rationality suffuses human life, it is integral to every aspect of human activity. That is why *good* human acts can involve what would seem naturally defective in other animals – wilfully forgoing food, for example, in the name of some social cause. And that is why, suggests Lott, we have reason, even after 'stepping back', to care about human nature. We *can't* 'step back' and consider human nature apart from what we have reason to do, because human nature *is* responsiveness to what we

[89] Lott graciously provides a number of other authors who think similarly. McDowell asks why some *individual* wolf who finds itself able to survive *without* a pack should care that wolves *in general* need to behave cooperatively – "Reason does not just open our eyes to our nature, as members of the animal species we belong to; it also enables us … to step back from it, in a way that puts its bearing on our practical problems into question" (McDowell 1998, 172). See also FitzPatrick – "departure [from biological function is] *systematic* in the human case, by virtue of the rational capcities" – Finlay – "it seems coherent … to question whether I have sufficient reason to perform my natural human functions well" – and Copp and Sobel – "why should the constituents of natural goodness for members of my species (or 'life-form') determine what counts as morally good for me?" (FitzPatrick 2000, 319, Finlay 2007 and Copp and Sobel 2004). There is also some similarity an argument in Woodcock – wherein natural goodness is said to require 'tacit appeal' to an 'independent ethical standpoint' to be justified – and to Korsgaard's arguments on 'substantive realism' and 'the normative question' – the requirement of some rational justification for the moral force of natural demands (Woodcock 2006, 445, and Korsgaard 1996).

[90] See Hursthouse 1999, 220 – "Our way of going on … is a rational way. A 'rational way' is any way that we can rightly see … as something we have reason to do" – and Foot 2001, 66 – "to speak of a good person is to speak of an individual not in respect of his body … but as concerns his rational will".

have reason to do. Whatever human nature tells us to do, it will be what we have reason to do, and so as rational creatures we must care about human nature.

Lott thinks this is the strongest response the neo-Aristotelian has to the *authority of nature challenge*. He also thinks it's a trap. This is because in pursuing the defence of essential rationality, the attributive-goodness theorist is treating 'human kind' as only a stand-in for 'rational being'. At this point it seems that there is little weight left to the idea of 'the human' as a natural kind itself – claims about human characteristics are secondary to claims about reasons, and don't speak to what those reasons are. The worth of human characteristics has become contingent on the ambiguous standards of 'the rational being'. It's entirely plausible, goes the argument, that there is nothing naturalistic left here, and so any attributive-goodness theorist that follows this line of reasoning – as Lott thinks they should – has abandoned naturalism, and with it much of their theoretical motivation.

Although Lott addresses neo-Aristotelianism exclusively, let me rephrase his objection in general attributive terms. In essence, Lott's challenge is that if there is any *kind* that has moral weight, it is no natural kind, defined by reductively natural standards, but only the *Rational Kind*, defined only by the standard 'responsiveness to reasons', whatever those may be. This tells us little about what we have reason to do, and sacrifices the naturalist commitments of general attributive-goodness theories, and so poses a significant problem – the strongest, I think, of the modern criticisms here reviewed.

§3.4.1 – Rationality About What?

Lott's argument above rests on four points. 1. Attributivists view Humans as essentially rational. 2. Rationality often recognizes reasons to depart from the traditional naturalistic goods of organisms and to embrace what would otherwise be defective. 3. From 2, reasons plausibly come apart from naturalistic attributive goodness. 4. Thus the normativity of 'rational

kinds' is not determined by any naturalistic facts about attributive goodness, and Attributivism is mistaken. Lott is certainly correct about points 1 and 2. Attributivists have no reason to accept points 3 or 4.

Lott's argument only holds if we accept that there *are* reasons that are not grounded in kinds and their attributive goodness. But attributivists needn't accept anything of the sort. Attributivists can plausibly argue that *all* reasons are derived from kind standards, and that responsiveness to reasons just is responsiveness to kind facts. Doing so defuses Lott's objection. Simply because we *can* ask whether or not kind facts give us reasons to promote them doesn't imply that the answer is no. Nor does the fact that we can ask whether or not there are *further* reasons to promote kind facts means either that there are, nor that there needs to be.

Lott is correct to say that the 'rational authority' of natural goodness relies on there being reason to promote natural goodness, but that doesn't preclude 'natural goodness' itself being the source of that reason. Lott's argument is only a problem for attributivists if the reasons rational beings are taken to respond to are not themselves cached out in terms of attributive goodness. The only particular reason to think that such must be the case – to think that the reasons Humans respond to as rational beings come apart from Human goodness – owes to the apparent fact that rationality often gives us reasons to deny certain organismic needs, and to act in ways that would be defective in non-rational creatures. But, as I'll now argue, the attributivist can explain that observation, and so defuse the motivation behind Lott's objection.

Why think that being responsive to reasons involves being responsive to more than just the goodness of one's kind? Because our reasons often seem to involve rejecting what is *good* by the standards of the human organism – excluding the standards of rationality – and embracing what is characteristically *bad* by those same standards. The human *organism* needs

to eat, to cooperate with its immediate community, and to reproduce. These however are behaviours which, as *rational* beings, we often shirk. An actor who starves himself to better portray his character, a protestor who vocally rebels against the fascist standards of her society, or a woman who decides against having children to pursue her career, is not obviously acting defectively. We can often decide *against* conforming to many characteristically Human standards, because we observe that to do so would be *bad*, or at least that it would be *better* not to. A non-rational animal meanwhile, one that obviously *is* bound to the norms of its kind, doesn't have that option. A deer which fails to flee at signs of danger is defective, even if fleeing leads it into a trap, and failing to flee would have been better for it. A deer ought to flee, even if it would be better for it to do otherwise. Similar claims can't be made for Humans as regards their characteristic behaviours. Human often have good reason *not* to partake in traditionally human activities – child-bearing for example – and can act well by doing so. That, I take it, is why Lott thinks that the reasons Humans respond to are not the same sort of naturalistic facts.

The problem with drawing Lott's conclusions from the above observations is that to accept that we have reason to reject certain characteristic behaviours at certain times isn't to accept that we have reason to reject *all* of them, or that those reasons must come from beyond the characteristic goods and necessities of our kind. Reasons to reject one characteristic Human good might be held because of how doing so will promote some *other* Human good. In fact that seems very plausible, and is certainly what attributivists should assert. The artist who starves for his craft sacrifices some degree of physical health, but does so for at least commensurate gains in social standing, satisfaction or psychological comfort. The revolutionary who rejects the fascism of her immediate society does so in the name of advancing the health of the wider Human community. The woman who decides not to reproduce does so to advance her fiscal stability – with all the creature comforts that provides

– or her social standing, or to focus on other relationships. All these are instances of sacrificing one Human good to better promote some others. Nothing so far said implies reasons grounded in something outside the good of one's kind. There is no objection to Attributivism here.

To summarise, Lott is correct to think that being essentially rational leaves Humans at the normative mercy of reasons. That is no problem for Attributivism though if those reasons themselves are determined by facts about kinds. And that of course is exactly what attributivists believe. Moreover attributivists can explain away the obvious observation that casts doubt on that belief – that practical rationality often plausibly justifies Humans in disregarding certain characteristic Human goods – in a way that retains the coherence of their view. Lott writes at times as though being *an* essentially rational kind equates to being a member of *the* Rational Kind, a unique kind of being whose function is to respond to some special set of pure Reasons. But the attributivist of course means no such thing. Humans are not just rational, full stop. They are rational about *being Human*, or rational *qua* Human. Rationality does indeed mean that Human goods are more varied and complicated than those of non-rational beings. Not though because Humans are subject to some special class of considerations, but because rationality allows Humans to evaluate, weigh, compare and act upon the full range of their goods in ways that non-rational beings cannot. Of course one might still question, upon using one's rationality to step back and ask why we have reason to do what is in our attributive goodness, the attributivist assertion that facts about one's attributive goodness actually are the sort of facts that can ground one's reasons. That though is a different question, one which has already been touched upon in earlier sections, and one which will be more fully answered in later chapters. What's important here is that the concept of essentially rational kinds poses no problem for Attributivism.

§3.5 – MOVING ON

Even if we can charitably say that early attributivists were clear about the basic principles on which Attributivism is built, what this chapter has demonstrated is that, since those early days, attributivists have been anything but clear. More often concerned with using natural norms to vindicate *other* normative theories – by grounding the *virtues* of virtue theories, for example – they have largely left the core principles of Attributivism itself to stagnate. Their failure to develop Attributivism explicitly in conjunction with the biological sciences and philosophy of function has resulted in confusion. Philosophers have been unable to see how Attributivism can plausibly navigate a landscape of callous and cruel natural processes, identify biological functions, or make sense of individual adaptations which diverge from the characteristic traits of one's species. In this chapter I have explained how Attributivism can do so. No argument levelled against Attributivism so far has been successful.

In §3.1 I responded to the argument that Attributivism requires organisms to act in ways that are detrimental to individual welfare, or are otherwise contrary to one's good. I argued that the objection confuses the sort of categorical features Attributivism endorses, namely categorical *necessities*, and that behaviours are only relevant to evaluation under Attributivism insofar as they are conducive to or required for the maintenance of the organism. I argued also that while it is true that even some beneficial categorical traits ought to be rejected, attributivists can explain that truth as grounded in *further* categorical needs.

In §3.2 I responded to the argument that Attributivism is unable to track genuine functional behaviours, or to differentiate them from incidental effects. I argued that attributivists can and do differentiate functional and incidental behaviours and effects through the *Reciprocal Dependency Test*. A behaviour is functional when it is *part of* a complex, hierarchical system of reciprocally dependent behaviours, such that to remove the behaviour would impact the continued maintenance of the system at large. I argued that any further

objections are based upon a commitment to *historical* functional analyses, while attributivists are instead plausibly committed to *ahistorical*, *systemic-capacity* style analyses.

§3.3 addressed the argument that Attributivism entails clearly objectionable results, both intuitively and internally, in such a way as to render it internally contradictory. I argued that intuitive objections are unpersuasive given Attributivism is, at heart, a revisionary doctrine, and prepared to reject the doctrine of reflective equilibrium upon which such objections rest. Meanwhile arguments of internal contradiction rest on an insufficiently fine-grained understanding of the relationship between organism development and behavioural standards.

Finally §3.4 responded to the argument that the capacity in rational animals to justifiably act in ways that would be defective in non-rational animals implies that appropriate Human behaviour is essentially dictated by reasons responsiveness, and not by attributive goodness. I argued that although Humans *are* essentially rational, that amounts to no objection if, as attributivists can and should maintain, reasons are reducible to facts about kinds. If rationality seems to justify sometimes acting defectively, it is only because rationality allows Humans to recognise how certain Human goods can be achieved through the sacrifice of others.

In the course of responding to the above objections I have, I hope, made clear many details about Attributivism that have gone unsaid for too long. With those details made clear, it is apparent that Attributivism can respond to every objection so far levied against it.

4 – The Marks of Normativity

Across the previous two chapters, I have explained Attributivism. An important takeaway from those chapters is that although Attributivism originates in a particular analysis of 'goodness', its primary concern is not a linguistic one. Attributivism is a theory of normativity; it attempts to explain a connection between the *property* of attributive goodness, and normativity, "the whole class of subject matters which are broadly … those traditionally studied by ethical theory: the *right, just, good, rational, correct* and so on" (Schroeder 2005, 3). Attributivism argues that facts about attributive goodness ground facts about what we ought and ought not to do. We are directed by Attributivism to do that which promotes our attributive goodness, and not to do that which impedes it. More than that though, Attributivism is a theory of the *fundamentally normative.* Attributive goodness isn't derivatively normative, in the way that, say, we ought to donate to charity *because* of some further duty, like promoting overall happiness. Attributivism presents attributive goodness as a fundamentally normative property, one that is capable of substantiating normativity even in the absence of any other normative properties. What that claim entails will be the focus of both this and the subsequent chapter.

To evaluate Attributivism is to evaluate the claim that attributive goodness is fundamentally normative, i.e., that whatever normativity *is*, attributive goodness is that which accounts for it, on the most fundamental level. But what *is* normativity? What is it all about, and what is it to 'account for it'? What is it that a fundamentally normative property *does*, in virtue of which it grounds normativity? What is it to be fundamentally normative? Before we

can begin to evaluate *any* specific property *qua* normativity, we need an answer to those questions.[91]

To start with, at least one thing we can all agree on is that normativity is about what *should*, or *ought* to be, or how we *ought* to act. We might disagree on whether or not it's a question of facts or attitudes, or on whether or not normative facts can be true, but fact or attitude, true or false, it's a question of what ought to be. With that as our starting point then, what it is for a property to be normative is for it to be such as to, somehow, make it the case that *certain* actions ought or ought not to be done. For a property to be *fundamentally* normative, meanwhile, is for it to be normative without that fact being grounded in any *other* normative property. Formally,

> **Normativity.** For any *P*, *P* is normative if the existence of *P* entails that for some agent *x* and action *Φ*, *x* ought to *Φ*.
>
> **Fundamentality**. For any *P*, *P* is fundamentally normative if *P* satisfies **Normativity** and there is no *P** such that *P** also satisfies **Normativity** and *P** explains why *P* satisfies **Normativity**.

For example, for attributive goodness to be normative, it must be the case that (i) attributive goodness explains the fact that some agent ought to take some action and (ii) there is no further normative fact that explains (i). i.e., it is not the case that attributive goodness explains that *x* ought to *Φ because* attributive goodness is also *good*, or because we have a duty to uphold attributive goodness, etc. The aim of this chapter then is to develop a specific account of just what a property must do to satisfy those criteria.

[91] Schroeder makes a similar and salient point: "in order to convince us that any reductive view [of normativity] *does* manage to capture everything … about the normative", reductivists owe us – in advance – details of what they take the "central truths" of normativity to be (Schroeder 2005, 8).

Two ways of approaching that task suggest themselves. The first is to look at properties themselves. The method here is to propose a 'test' of the form 'for any property *x*, *x* is fundamentally normative *iff y*'. Here *y* takes the place of some feature of the property, or some way the property itself interacts with agents, in virtue of which standards of behaviour are derived. As a, woefully poor, example, we might suggest that 'for any property *x*, *x* is fundamentally normative *iff x* is *good*', and thence derive the standard that 'one ought to do what is good'. Failure of such a test follows when it can be demonstrated that the conditions outlined in *y* fail to guarantee, alone, normative claims. This may be for any number of reasons, but, examples would include the proposed conditions being (a) dependent upon some more basic normative facts, (b) in some way internally contradictory or allowing for the derivation of contradictory commitments, of the sort 'one ought to Φ and it is not the case that one ought to Φ', or (c) able to be used equally to require or forbid any action whatsoever, and thus generate no specific normative content.

The second, alternative approach is suggested by the above. One thing that seems clear is that any fundamentally normative property must somehow correspond to a normative standard of behaviour. Attribute goodness, for example, is fundamentally normative *iff* there exists a standard of the form 'do what is attributively good', and that standard is fundamentally normative, i.e., normative in a way that relies upon no further normative facts. In that case, the standard can stand in heuristically for the property itself, and we can formulate tests in the following way: 'for any standard of behaviour *x*, *x* is fundamentally normative *iff y*'. Failure of the test would follow in the same way as above.

Throughout this chapter it will be at times appropriate to formulate tests in terms of properties, and at times in terms of standards. I will treat them interchangeably.

Finally, as the topic of normativity is one with a rich and contentious history, it is expected that many philosophers will already favour certain views, explicitly or implicitly. In

the interest of clarity I will begin this chapter by reviewing three popular approaches to explaining normativity. I will explain why I don't think them sufficient, before arguing for the pair of conditions I *do* think necessary and sufficient to establish fundamental normativity.

In §4.1 I evaluate 'reactive attitude' accounts. These accounts argue roughly that a standard of behaviour is normative if and when society disapproves of its violation. I argue that either reactive attitudes depend upon further normative facts, or entail implausible and ad hoc commitments.

In §4.2 I evaluate the view I call *Naïve Instrumentalism*, a theory which argues that normativity is exhausted by end-conducive norms, and that standards are normative when following them is conducive to achieving one's ends. I argue that any theory of normativity that doesn't allow for comparative rankings of ends is one in which every action comes out right, and thus one in which there is no real normativity at all.

§4.3, 'Reasons', evaluates the 'reasons first' theory of normativity, wherein normativity is analysed exclusively and non-reductively in terms of reasons. I argue that without further reducing normativity, 'reasons first' theorists are unable to explain the difference between normative reasons and non-normative reasons, in a way that allows reasons to provide an informative analysis of normativity.

§4.4, 'Authority', transitions from explaining the weaknesses in other theories, to presenting the first of two necessary elements of any fundamentally normative property. Any fundamentally normative property must be able to answer the 'why should I care?' question – that is, the question as to why one should care about any particular normative claims. Authority is the capacity through which that question is answered and, I argue, must be instantiated in a particular sort of necessary connection between any supposed normative property, and the motivations – both actual and ideal – of agents.

In §4.5, Action Regulation, I present the second capacity required of fundamentally normative properties, the capacity to regulate action. My argument is simple – in any world wherein there are no actions that come out *wrong* as well as *right*, there is no normativity *worth talking about*. Any normative property therefore must allow for the separation of actions into at least two categories, those *ruled in*, and those *ruled out*. I review several methods by which that might be accomplished.

§4.6 concludes. The discussion so far is summarised, and I segue into the subsequent chapter, wherein I demonstrate what is entailed in meeting these requirements, and how Attributivism proposes to do so.

§4.1 – REACTIVE ATTITUDES

Why are we motivated to consider normativity? Where do thoughts of normativity come from? A plausible-seeming hypothesis might suggest that they arise in virtue of our participation in society – from our interactions with other people, the attitudes that agents hold towards certain actions, and our reactions towards those who fail to act as we feel is appropriate. We might even think that finding something normative goes hand in hand with endorsing actions that promote it, and condemning those which conflict, and so that normativity is in some way embedded in those attitudes themselves. A popular family of theories of normativity is predicated upon those sorts of beliefs. Call these *reactive-attitude* accounts.[92] The focus of this section is on rejecting these theories, insofar as they propose to explain the constitutive features of normativity. In this introductory subsection I explain a bit about how reactive-attitude accounts work in practice. I argue that from that practice, two distinct theories of normativity emerge – those that are predicated on *appropriate* reactive attitudes, and those predicated on *actual* reactive attitudes. In §4.1.1 I argue that the first is trivially true. In §4.1.2 I argue that the second entails implausible consequences.

In practice, reactive-attitude theories amount to something like the following: an action is normatively *wrong* iff it is *blameworthy*.[93] Meanwhile, an action is blameworthy *iff*, as Mark van Roojen paraphrases Gibbard, "it fails to meet standards of action the intentional or negligent violation of which in a normal state of mind would be sufficient for finding the agent prima facie blameworthy" (Roojen 2013, 2.5). That is to say, an act is blameworthy *iff* it is the sort of act found blameworthy by agents in a normal state of mind, when taken by agents in a

[92] Archetypical reactive attitude theories are developed by philosophers like P.F. Strawson (1962), Alan Gibbard (1990, 2003) and Simon Blackburn (1998). Others, like R.J. Wallace (1994), have developed accounts that see reactive attitudes importantly involved in moral practice, without being necessarily basic *to* the practice. I won't be dealing with any single, specific account here.

[93] There are a host of reactive attitudes that might be employed here. Approval, for example, is also a reactive attitude we hold towards actions. In this section I am using the attitude of *blame* as a stand in for the whole host of attitudes we might hold.

normal state of mind. More generally, a standard of behaviour is normative *iff* we blame people for violating it. As an example, attributive goodness would be a normative property if, for example, when we recognise an agent as promoting their attributive goodness, we approve of their behaviour *on that account*, and would find them blameworthy in some sense were they attributively deficient.

It is appropriate here to ask whether there might be any restrictions on the circumstances in which such attitudes are held, if they are to be taken as corresponding to a normative standard. For example, is a standard normative just in case *any* agent whatsoever blames others for violating it? Or is it only so when *certain* agents do, or in *certain* circumstances. In the above quote from van Roojen, for example, the attitudes of agents in a *normal state of mind* are appealed to. I think that particular clause slips in a little more easily than is justifiable. In fact I think it causes significant problems for reactive-attitude theorists. Namely that they must either take *normalcy* to be itself normatively privileged, or have no reason for privileging it at all. I will explain shortly why that is a problem. In any event, at least two sorts of reactive-attitude theories must be considered. That a standard is normative (i) *iff* we *appropriately* blame agents for its violation, or (ii) *iff* we *actually* blame agents for its violation. Neither interpretation is plausible.

§4.1.1 – Appropriate Attitudes

It seems true that many reactive attitudes can be held inappropriately. I may, for example, inappropriately blame someone for something they had no control over. I may also blame someone for something inappropriately due to defect on *my* part – because I am drunk, and scold my friends for stopping me from drinking more, or because I am depressed, and hold it against others that they are happy. It's that sort of observation that motivates the following.

Reactive Attitudes (Appropriate). For any standard of behaviour *x*, *x* is fundamentally normative *iff* it is appropriate to blame agents for violating *x*.[94]

Simon Blackburn's quasi-realist approach seems to endorse something along those lines, where normative judgments of the sort '*y* is bad' are taken to express attitudes of censure towards *y*, yet where those attitudes can be held *correctly* or *incorrectly*. Correctly, for example, towards instances of child abuse, and incorrectly towards instances of child support. It may not be entirely clear *why* child abuse is that which is appropriate to censure. Perhaps because that's just what normal minded members of society censure, or perhaps because child abuse just *is* the sort of thing that a normal minded agent *would* censure.[95] Nevertheless, what *identifies* child abuse as a violation of a normative standard is, on this sort of account, that it is appropriate to blame child abusers.

It isn't plausible however to say that child abuse is a violation of a normative standard *in virtue of* the appropriateness of blaming abusers. The problem with thinking so, and with **Reactive Attitudes (Appropriate)** in general, is that it builds normativity into both sides of a supposedly reductive equation, and so defines normativity, uninformatively, in normative terms. What the above says is that *x* is a normative standard *iff* certain attitudes are held *appropriately* in response to violations of *x*. Appropriateness though is a normative concept. So what **Reactive Attitudes (Appropriate)** amounts to is that standard *x* is normative if and when it is normatively fitting to hold certain attitudes towards violations of *x*. Likewise, as the

[94] Here and elsewhere I refer to the reactive attitude of *blame*, for the sake of simplicity. It should be noted that blame is only one reactive attitude that one could hold, to establish normativity on this account. *Approving* of agents who *uphold* a standard would likewise imply normativity of that standard.

[95] Blackburn himself seems to favour the latter view, claiming that what makes child abuse blameworthy are facts about child abuse, and not about attitudes. See Blackburn 2006, particularly pp.146, 148-9. It may be correct to say that Blackburn is more concerned with explaining facts about moral *practice*, and explaining the psychological force of moral judgments, than he is with explaining the grounding features of normativity. It may also seem plausible however to argue that the normative relevance of facts about child abuse is grounded in the appropriateness of holding specific attitudes towards those facts. I take it as plausible then that Blackburn-type views can coincide with the **Reactive Attitude (Appropriate)** thesis.

biconditional flows both ways, if we are normatively obliged to hold certain attitudes towards violations of *x* then *x* is normative. Nobody denies *that* though, it amounts to a vacuous definition of normative standards as standards which are normative. That has told us nothing about what actually generates that normativity.

So it being appropriate to blame agents for violating a standard can't be the mark of such a standard's fundamental normativity. Normativity that is explained by the appropriateness of reactive attitudes can only be derivative, dependent upon some further normative property to explain the appropriateness of the attitudes itself. With the appropriateness of reactive attitudes off the table then, where is the reactive attitude theorist to turn?

§4.1.2 – Actual Attitudes

What makes a standard normative can't be that it would be *appropriate* to hold reactive attitudes towards its violation. If we want to retain a place for reactive attitudes in the analysis of normativity, and we can't employ *appropriately* held attitudes, all that remains are the attitudes that are *actually* held. The only place for a reactive-attitude theorist to go, then, is to say that what makes a standard normative is that members of society *actually* hold reactive attitudes towards its violation. Thus,

> **Reactive Attitudes (Actual).** For any standard of behaviour *x*, *x* is fundamentally normative *iff* there exists an actual social practice of blaming agents for violating *x*.

The problem with the above is ultimately that it makes normativity too easy. As I'll now argue, **Reactive Attitudes (Actual)** reduces all normative standards to generic norms of practice – wherever there exists an actual social practice of blaming for violations of a standard, then we

have genuine normativity. But norms of practice are abundant – they exist everywhere someone asserts a rule, or imposes a standard. It is implausible to hold agents to every standard someone, somewhere, will blame them for violating. Yet without the ability to normatively privilege one practice over another for employing the *right* attitudes – which was sacrificed in §4.1.1 – that's just what we must do. The result is an incoherent account of normativity, according to which any action stands to be one that both ought and ought not to be done, in equal measure.

To take a step back and put all this into perspective, it will be helpful to consider for a moment how normativity and attitudes like blame typically seem to interact. There's no doubt blame and normative concepts like *wrongness* are closely connected. We blame people for doing wrong, and to think an act is blameworthy is generally to think that it's wrong. Brad Hooker writes that "Most of us believe that … blameworthiness [in some domain] is very closely linked to … wrongness [in that domain]. We might even … say that acts … are … blameworthy if and only if … wrong" (Hooker 2000, 73). The natural way to read Hooker here is as saying that most of us think facts about wrongness are antecedent to blameworthiness – something is worthy of blame *because* it is wrong. A biconditional flows both ways though, and the position of the reactive-attitude theorist is that we should take Hooker's observation to imply the opposite. Something is wrong *because* it is blameworthy. For an act to be blameworthy, meanwhile, is just to be the sort of thing society actually blames agents for doing. Their view here is that a standard of action is normative *because* of its relation to our attitudes.

That approach seems endorsed by Gibbard, and particularly by Strawson, who argues in 'Freedom and Resentment' that facts about moral responsibility derive from an internally coherent practice of holding responsible, which "neither calls for nor permits an external 'rational' justification" (1962, 23). Once we have, that is, a coherent understanding of when, why and how agents actually assign blame, what actually *is* wrong – i.e., *is blameworthy* – is

determined accordingly. Thus a standard is normative insofar as it actually is the case that agents blame others for violating it.

At this point, the above must be the view of the reactive-account theorist. Yet it is patently implausible that we should think that agents genuinely ought to Φ just in case someone will blame them for not Φ-ing. Racists may blame their children for entering mixed-race relationships; nonetheless those children have been given no reason to comply, and they genuinely shouldn't. Norms of practice exist everywhere – the rules of games, the arcane standards of etiquette, the right way to drink your whisky, the norms of antisemitism. Anything you do is likely to run afoul of *some* practice, and thus be impermissible, if all we require for normativity is that some social practice will blame you for acting. Likewise anything you do is likely to be *approved of* by some other social practice, and thus be permissible. The result is an incoherent quagmire of norms, impossible to navigate.

At this point the reactive-attitude theorist must offer qualifications, to attempt to prioritise amongst social practices and the standards they respect. Three tactics seem available. They might object that not *every* social practice is meant to bear upon one's actions, only practices one takes part in. Alternatively, they might idealise, and argue that the prioritised social practice is the one an idealised agent would take part in. Finally, and relatedly, they might appeal to *normalcy*, and argue that prioritised practices are those that agents in *normal* states of mind take part in.[96] I lump all these tactics together, as they all fall prey to the same refutation. Either each privileged category is *normatively* privileged, in which case some

[96] Gibbard, for example, appeals to normalcy in *Wise Choices, Apt Feelings*. He posits authority upon the standards of the 'normal' state of mind, where normal is presumably read non-normatively, but rather in the sense of the typical member of society. He argues optimistically from the supposed function of reactive attitudes to claim that individuals will inevitably hold favourable attitudes towards social agreement – that is, their *actual* attitudes will turn out to be sympathetic to the larger system of norms. Later work in *Thinking How to Live* sees our attitudes as expressing the acceptance of plans, with a sort of 'master set' of attitudes being determined by the maximal set of compatible plans, representing a sort of hyper-plan (2003, 54). I tend to think that this strategy is not a promising one – I can't see how it avoids the problem I've laid out. Either the 'normal' state of mind, or the hypothetical attitude towards the hyper-plan, trumps the agent's *actual* state of mind or attitudes in the moment in virtue of some further normative fact, or it fails to explain their supposed authority.

further source of normativity is required to explain that privilege, or the prioritisation is unconvincingly ad hoc. Answers are required to all of these questions: why, if we're abnormal, should we care about what's normal? What makes any given account of idealisation any more important than the non-ideal? Why does my choosing to partake of a given practice make that practice more authoritative than others? Likewise, why does the authority of that practice not just end when I decide to do other than what it prescribes? Isn't my 'society of one' just as legitimate as any other, thus making normative anything at all I approve of in the moment? The answer to any of those questions can't apply to the attitudes agents actually hold, otherwise we're right back where we started. If it appeals to any other normative property though, the reactive-attitude project is lost.

Without appealing to more basic normative properties, **Reactive Attitudes (Actual)** is held hostage to the attitudes involved in *any* social practice. The result is an incoherent normativity, one that sees agents pulled every which way by the authority of endless conflicting standards. So it's not plausible that wrongness derives *from and because* of actually held reactive attitudes. If it is also implausible to claim that normativity stems from *appropriately* held attitudes, then reactive attitudes cannot account for fundamental normativity.

In summary, the practice of blaming for violations of a standard cannot be what accounts for normativity. It is difficult to conceive of a way to express the connection between such attitudes and normativity without relying on some further, non-reactive-attitude-dependent sense of wrongness, appropriateness or normativity. The reasons in favour of reactive attitude accounts are less convincing than the reasons against it, and I won't consider them further here.[97]

[97] I lie. This section is somewhat uncharitable in that it leaves out what many expressivists take to be a major motivation, the idea that we can't count as honestly endorsing or agreeing with a proposition unless we express non-cognitive attitudes towards it. Thus holding such attitudes must be constitutive of finding something normative, or even rational, and since cognitive beliefs don't entail such attitudes, cognitivist theories of normativity are mistaken. I want to acknowledge this argument, although I can only say briefly here that I disagree with the contention on two grounds: first, there is plausibly a difference between what is required for finding or

§4.2 – NAÏVE INSTRUMENTALISM

This section presents and argues against what I'll call *Naïve Instrumentalism*, a view that claims normativity is constituted entirely by means-end conduciveness within *Norm-Implying Frameworks*, a term defined below. I reject this view on two counts, the first being that it fails to explain the connection between normative judgments and motivation, and the second being that naïve instrumentalism actually amounts to an error-theory of normative reasons. In rejecting naïve instrumentalism, I draw the conclusion that normative properties must be non-instrumental properties that inform the prioritization of some ends over others.

There's no question that normativity is intimately tied up with practical reasoning. The paradigmatic example of a normative judgment is a judgment about what to do, specifically what we *ought* to do. The paradigmatic example of an ought judgment is, in turn, a judgment about what to do *instrumentally*, given some particular end, or within the context of some *rule* or another. That is to say, a judgment that one *ought* to do *x in order to y*. These are *practical* judgments, concerned with adapting means to ends. Thus if I ask in respect to construction, 'what tool should I use for this job?', or in a game of chess, 'what move ought I to make here?', or at a dinner party, 'which is the right spoon for soup?' in all cases I am asking a normative question.

Although uncontroversial itself, the above observation often leads to a controversial conclusion, that this sort of practical reasoning in the light of hypothetical standards is all there is to normativity. This is the view I call *Naïve Instrumentalism*; it says simply that there is nothing more to normativity than the norms of convention, rules, and means-end conduciveness

holding something normative, and what it is for something to *be* normative. Secondly, I disagree that having a cognitive, propositional belief doesn't also entail having a minimal sort of attitude towards the object of the belief sufficient to satisfy the expressivist intuition. If I genuinely believe something – say that Norwegian folk music is superior to Wagnerian opera – I will be disposed to take attitudes that favour Norwegian folk music over Wagnerian opera if pressed, otherwise we have discovered that I don't actually believe what I've claimed. Philosophers like Scanlon and Skorupski seem to agree with, and to employ in their explanations of the connection between reasons and motivation. More on that, however, in §4.3.

therein. This is the sort of normativity Derek Baker calls *Formal Normativity*, "the normativity displayed by any standard one can meet or fall short of" (Baker 2017). Informally, *naïve instrumentalism* is the view that normativity is exhaustively constituted in the set of all formal norms, such as the rules of etiquette, chess and the law. Somewhat more formally, it says that normativity is grounded and located solely by and within the context of *normative-frameworks*, or *NF*s.

> **Normative Framework.** A normative-framework, or NF, is any system of ends and/or standards, means, and rules, within which practical deliberation can occur. Any *x* is an *NF iff x* is composed minimally of
>
> (i) an end or formal standard which can be realised or frustrated with varying degrees of success,
>
> (ii) rules or laws which must be followed if (i) is to be realised
>
> (iii) actions required or forbidden by (ii).

The rules of chess, accordingly, are a NF: one can deliberate about moves in chess precisely *because* there is an end that can only be achieved through taking certain actions, actions picked out and made permissible/impermissible, right/wrong, good/bad, etc., by a set of rules. If there were no particular *aim* involved in playing chess, and no rules dictating *how* to achieve it, there could be no deliberation about whether or not any particular move were *correct*. The law is a NF, involving a formal standard, the standard of the law-abiding citizen, and a series of rules according to which individuals frustrate or realise that standard through taking, or refraining from, certain actions. Etiquette is a similar NF. On the other hand, having the end of doing *either* A *or* not-A provides no framework – one can't deliberate about what to do in virtue of

that goal, because *anything* will achieve it, and so nothing is any more or less appropriate in light of it.[98]

Naïve Instrumentalism maintains that the existence of these frameworks is what allows for normativity, through providing norms with respect to which one can act. Moreover and more importantly, it maintains that normativity can't exist independently of them. As a result, they maintain there can be no such thing as an obligation, a reason, or any other normative concept or property beyond those that are concerned with meeting a given standard or end. The naïve instrumentalist thus believes that what one should do is contextually determined by the frameworks they fall under – say the *NF* of chess for a chess-player, or the norms of a country's laws for a citizen of that country. Thus, the the following proposal.

> **End-Conduciveness.** For any standard of behaviour *x*, *x* is fundamentally normative *iff* adhering to *x* promotes end-conducive action within a normative framework.

End-Conduciveness doesn't seem *prima facie* implausible. The reason why a chess player ought not, all things being equal, to trade her queen for a pawn, is because to do so is unconducive to the end of playing, or winning, at chess. If I am building a house, I ought to include a roof, because doing so is required to successfully build a house. A philosopher ought to reason well, because doing so is conducive to the ends of Philosophy.

Arguing transparently is likewise good, and ought to be done, relative to the *NF* of Philosophy. Interestingly though, arguing transparently may also be bad, and ought *not* to be done, relative say to the *NF* of political fearmongering. That raises a host of further questions:

[98] My formulation of Naïve Instrumentalism may seem to be running together two different ideas: instrumental norms, relative to practical requirements for achieving ends, and formal norms, relative to any rule or standard one can fall short of. I purposefully don't observe this supposed distinction, as it seems to me rules or standards of behaviour will always follow from, or entail, the existence of an end they promote, as the rules of etiquette follow from or imply the end of upholding etiquette, and function to pick out the actions instrumental to achieving that end. If there is a distinction, it isn't one I think worth pursuing here.

how do we decide between conflicting *NF*s? Are some frameworks more or less relevant than others for specific agents, and how are they made so? Can any *NF*s be more *normatively* relevant than others, or is n*aïve instrumentalism* indifferent to which ends agents pursue?

Those questions will be discussed throughout this section. In the meantime though, what is important to take away from *this* section is that, according to *naïve instrumentalism*, acts are only right or wrong, good or bad, required or forbidden, relative to the rules of some normative framework, and insofar as they are end-conducive or unconducive therein. That's where standards of behaviour come from, that's what it is for them to be normative, and that's what normativity consists in.

§4.2.1 – The Appeal

Why accept *Naïve Instrumentalism*? The basic idea is incredibly intuitive – so much so that generations of philosophers have observed it. That idea is that we can deliberate about actions *only* insofar as they are viewed as means to an end. Aristotle supported such an idea: "we deliberate not about ends but about what contributes to ends" (*NE* 1112b11-12). So, too – on a face-value reading – did Hume: "reason is, and ought only to be the slave of the passions" (Hume 1989, II:iii:3 415). On reasons, Velleman argues that a reason is essentially "a consideration that justifies" and that something can only be justified in the context of "a *jus*, or norm of correctness" (Velleman 2000, 15). William Frankena (1976, 120), Dan Brock (1977, 75) and David Copp (1997, 101-103) have all made similar arguments, to say both that deliberation and justification – whether of actions or ends themselves – can only occur *internally* within a system of norms, relative, that is, to some standpoint or point-of-view, and that because of this no particular framework can have any greater claim to 'finality' or 'supremacy' than any other. To claim priority, they argue, would require yet a further framework to justify accepting said priority, and that very dependence would undermine any

such claim. Stephen Finlay too has recently argued for parity between morally and instrumentally normative languages; he argues that terms like 'good' and 'ought' operate explicitly within end-relative frameworks, and track probabilistic relations therein (Finlay 2014, ch.1). Henry Richardson sums all this up nicely as the *scope obstacle* to reasoning *about* ends – the idea that evaluations of ends 'overstep' the scope of deliberation (Richardson 1994, 13-18).

So the appeal of *Naïve Instrumentalism* is built on what seems to be a ready truth, that when we deliberate, when we ask what we *ought* to do, or what is *right* to do, we do so always with regard to some framework of standards or ends. Of course that claim isn't *universally* accepted. Stephen Darwall, for example, claims that we often and intelligibly wonder what we ought to do *simpliciter*, and seek *unqualified reasons*, *sans phrase* (Darwall 1990, 258). If I were, in a whimsical sort of mood, to look to the sky and wonder "what should I do?" I would be – thinks Darwall – intelligibly asking what ends I should have, not what I should do in terms of some framework I'm currently located within. Darwall then seems to reject *Naïve Instrumentalism* for failing to fully account for the scope of the normative questions we can ask.

I'm unsure of that argument – it seems even that if I *do* ask such a question, I might be doing so with some abstract end in mind, say that of 'achieving what's best for me', or with a view to some abstract deontic standard, divine law or the like. All the same, the possibility of 'unqualified reasons' is one that resonates with many philosophers. The thought that normativity requires – if we are to 'take it seriously' (Enoch 2010, 111)[99] – some sort of decisive 'normative bedrock' that cannot be undermined (Parfit 2006, 377), or that is justified in-itself (Radzik 2000, 637), or that resists the demand for justification altogether (Silverstein 2008, 7) is one held by a multitude. I tend to agree. Here's why.

[99] See also Enoch 2011, particularly chapter 2.

§4.2.2 – The Importance of Practical Reason

Consider the following: "any system which proposes multiple, potentially competing ends will need some ranking of those ends as better or worse, if it is to provide definitive advice in complex situations" (Baker, 2017). This is evidently true. If I find myself possessed of multiple competing ends and sufficient motivation to pursue any of them, given the right push, I require one of them to be normatively privileged if there is to be any truth to the idea that there is anything in particular I *ought* to do amongst those things I *am* motivated towards. Do I gracefully concede the chess match to my fortunate opponent, or do I leap across the table and throttle them? If the best that my theory of the normative can ever say is to do one thing or any other, as long as I'm motivated to do it, then what we have is no normativity at all, but rather a theory that tells us to do whatever it is we end up doing. To resolve this sort of dilemma, a normative theory needs to be able to tell us what our *ultimate ends* are, or ought to be. That though is just the sort of thing *naïve instrumentalism* can't do.

Warren Quinn argued a salient point in 'Putting Rationality in Its Place'.

> "Practical thought [deploys] a master set of noninstrumental evaluative notions: that of [the non-instrumentally good or bad]. Practical reason is … the faculty that applies these fundamental evaluative concepts. If there is no truth to be found in their application, then there is no point to practical reason and no such thing as practical rationality" (Quinn 1993, 233).

Quinn makes this claim in the course of arguing that caring solely for end-conduciveness undermines the importance we attribute to the faculty of practical rationality. Foot nicely paraphrases Quinn as asking 'what would be so important about practical rationality if all it

concerned was reasoning from means to ends, whatsoever those ends might be?' (Foot 2001, 63).

There are two ways to take Quinn's statement. He can most directly be interpreted as arguing that if practical rationality is neutral to ends themselves, and compels us merely to do whatever best achieves whatever end happens to motivate us, then we have no good reason for *valuing* the results of deliberation, since we can hold ends that are manifestly absurd, cruel or self-destructive.[100] While no doubt a compelling observation, that result is different from there being, as Quinn asserts, 'no such thing as practical rationality'. That suggests a second interpretation.

That second interpretation is the one I think most relevant. It says that insofar as justification for action stems from end-conduciveness, without care for *which* ends are pursued beyond a minimal claim to motivation, there can be no authoritative critique of action, provided it actually promotes the end that motivated and condones it. This is a particular problem given that it seems *every* action is taken with an eye to *some* end – *at least* the minimal, implicit, motivating end of *taking that action* – which will positively situate the action within as valid a normative-framework as any other.[101] That leaves no possibility of reasoning *wrongly*, since *every* action will be justified by *some* end, even if just the very minimal end of taking that action. As such, *naïve instrumentalism* leaves every agent immune from criticism, provided

[100] See particularly Quinn's 'radioman' example, also in Quinn 1993.

[101] I might rightly be accused here of ignoring a category of action that seems to be taken without an eye to any end, and yet be normatively relevant, namely *negligent* actions. As an example, if in the course of driving carelessly I strike a pedestrian, I've done something it seems I ought not to have. Although I didn't choose to do so, and was in a sense unconscious of acting to strike the pedestrian, the situation was under my control, I *could* have avoided it if I'd been more cautious, so *prima facie* it is something I can be held accountable for. I don't think that sort of case can redeem *naïve instrumentalism* however. The normative status of negligent, unintentional action must ultimately depend upon facts about intentional action, i.e., that I ought to have *chosen* differently than I *did*, or ought to choose something specific now. That though, under *naïve instrumentalism,* is false. Whatever I *did* do, as long as it was motivated by *some* end, even just the end of *doing it*, and likewise whatever I *end up doing*, I was/will be normatively justified, regardless of what extra consequences have unintentionally been brought about. Once the dialogue returns to talk of ends, everything I have said remains applicable.

that their actions actually promote the ends towards which they are taken, leaving us with no meaningful understanding of rationality, or normativity.

Under *naïve instrumentalism*, any normative-framework, or set of normative-frameworks, justifies as well as any other. If these frameworks are the sources of their own authority, then *naïve instrumentalism* has no resources with which to argue against that conclusion.[102] Yet since every action is taken with an eye to at least *some* end though, *every* action is justified by *some* framework. That is what *Naïve Instrumentalism* entails. Any theory of normativity though with the result that everything we do is normatively equivalent to anything else we might do is a theory that is indistinguishable from one which posits no normativity at all. Thus *naïve instrumentalism* fails as a theory of normativity. The possibility of normativity, and the importance we attribute to practical rationality, presupposes that some ends are more important than others. An essential part of investigating normativity, then, is in trying to determine what that importance consists in.

§4.2.3 – Progression

Have we reached an impasse? The arguments I've laid out, in the vein of Smith and Quinn, suggest *Naïve Instrumentalism* is at odds with core commitments of normativity. Yet other arguments already discussed make a compelling case that practical rationality – our faculty for recognising and employing reasons and justification for action – can only be undertaken within

[102] There may seem to be a salient objection I haven't considered. It argues that priority amongst frameworks can be established by considering the relevant *strengths* of the ends a given agent holds. I don't spend any time on that objection however, because it doesn't defend *naïve instrumentalism. Naïve instrumentalism* is *not* a view that takes *ends* themselves to be fundamentally normative, or that says what one ought to do first and foremost is realise one's ends. Rather, the *norms* produced by those ends are normative – *if* one ought to realise one's ends, it's because one *ought* to obey the end-conducive norms they produce, and not the other way around. As a result, *naïve instrumentalism* is not able to appeal to, for example, the strengths or centrality of particular ends within our motivational sets in order to do things like claim certain ends are more normatively important, or weighty, or otherwise *ought* to be pursued over others. To appeal to the strengths of ends would require an additional substantive claim along the lines of that what we ought to do is to realise our strongest desires. That is a claim *naïve instrumentalism* doesn't and can't defend.

a normative-framework. NFs and their ends, it seems, *get reasoning going*. How do we resolve the tension?

I propose to accept the central claim of *naïve instrumentalism*. Ends, and their frameworks, 'embed the normative', as Richardson would have it (1994, 50); they serve as the base of our deliberative architecture, and ground the chain of reason. We should reject, however, the idea that all ends stand to have equal place in the structure of deliberation. That contentious principle within *naïve instrumentalism* comes about as a result of the thought that, in Copp's words, any privileged end would have to be *normatively privileged*, i.e., given supremacy by yet another normative-framework, and justified in the same way actions are justified, as an end we *ought* to hold (1997, 101-103). Yet plausibly we *choose* actions, while we *have* ends, and the structure here is not obviously analogous. For one, we tend to think we could always have chosen *to do* something else, while there may be ends that we *can't* ignore, fundamental ends from which others derive contingently, or that are fundamental to deliberation itself, and which, insofar as we deliberate, we *must* ultimately be deliberating towards. Deliberation about action does indeed start from practical questions, but deliberation about action doesn't answer *why*, or to what *end*, we ask those questions in the first place. Questions about ends themselves are arguably not about *what to do*, but about what *is*.

This tactic, imposing the structure normativity requires by supplying unavoidable and inherent deliberative structure, seems to me the more promising one. Many seem to agree. Darwall believes end-priority can be justified without reference to other systems of internal norms, and he is joined by philosophers like Parfit, Korsgaard, Silverstein, all of whom argue that normativity involves some end, broadly speaking, that is *privileged* in some way so as to end justificatory regress. A 'source of normativity' that we *have* in a way that grounds deliberation, and which defies the demand for further justification, or which can't have its

relevance further questioned, or about which we can't sensibly ask 'why?'[103] These views attach normativity to features that we *cannot help but be ruled by*, if we are to engage in deliberation *at all*. Candidate views include any constitutivist or constructivist view – where certain norms are constitutive parts of rationality itself – attributivist views – where facts about *what we are* constitute our practical commitments and the essential topic of deliberation – and any realist theory of reasons – where reasons are *what we deliberate about*.

To summarise, any theory of normativity that is entirely internal to independent hypothetical frameworks will not get far. The mere existence of such frameworks will never substantiate alone reason to be guided by their imperatives. Further, if justification is entirely internal to independent frameworks, in the absence of any unifying feature, each framework is a sufficient source of normative authority, an island unto itself with no requirement for the referencing or consideration of any other demand. A meaningful theory of normativity requires both a connection to motivation, and the capacity to impose an authoritative, *rationally and practically unavoidable* structure upon deliberation, allowing for authoritative prioritisation of normative-frameworks. I'll return to that thought in §4.4.

[103] See Darwall 1990, Korsgaard 2008, 163-4, Parfit 2006 377-9, Silverstein 2008, 2012.

§4.3 – REASONS

According to a predominant school of thought, ethics is the science of reasons. It's the study of whether reasons exist, what they are, how they obligate. What we ought to do is that which we have most reason to do. That is something many philosophers take very seriously. The result is a view whereby for something to be normative is for it to stand in a particular sort of relation to reasons – we ought to Φ *iff* we have reason to Φ, or more precisely if reasons count in favour of Φing. That produces what, following Derek Parfit, we might call the *reasons-implying* formulation for normative properties.[104]

> **Reasons-Implying.** For any property x, x is fundamentally normative *iff* having x is necessarily reason-implying.

Under the above, goodness, for example, is a fundamentally normative property *iff* any time something is *good* that gives us reason to act in certain ways, say to promote that thing. That's a very plausible view; indeed I think it's true. Notice though that the above says nothing about the *direction* of the relation between what we ought to do and reasons. A natural way of reading the above would be to assume that goodness is reason-implying because of facts about goodness itself – that what we have reason to do owes to facts about goodness. An alternative view however sees the relation run the opposite way – goodness is reason-implying simply because it corresponds to what we have reason to do. On that view, reasons are the fundamental ground of normativity – they are *sui generis*, facts about reasons neither depend upon nor reduce to any other facts, and they alone determine what they count in favour of. Thus,

[104] See Parfit 2011, chapter 1, §1.

> **Reasons-First.** For any property x, x is fundamentally normative *iff* having *x* is necessarily reason-implying; what is reason-implying depends on facts about reasons that are independent from and irreducible to any other facts.

That is, I think, not true. In the rest of this section I'll say a bit more about reasons and reason-relations themselves, by way of explaining why some philosophers take **Reasons-First** to be attractive. In the next section I'll explain why it isn't.

According to **Reasons-First**, normativity consists in the relation between reasons and properties. Goodness is normative if and when it stands in a relation to reasons, whereby when something is good that implies the existence of reasons to act in some way towards it. That *reason-relation* is the fundamental normative unit, and is grounded in reasons themselves, which determine the nature of the relation. So what is a reason? Broadly put, a reason here is just whatever counts in favour of something. That it is lonely on the mountaintop is a reason for my going to the mountaintop, because it, alongside my desire for solitude, counts in favour of my hiking up the mountain. My having a headache is a reason for me to take painkillers, because it counts in favour of taking painkillers. Likewise going to the mountaintop is *good*, and my headache is *bad*, specifically because features of mountaintops and headaches give me reasons to act, by counting in favour of certain actions. Normative properties like goodness and badness derive from the reason-giving features of their bearers.

It's relatively easy then, I think, to see why some philosophers want to rest normativity on reasons. 'Counting in favour of something' is just what normative facts must do. Tautologically if something is normative it counts in favour of us acting in some way. Reasons almost by definition just are the sort of things normativity is concerned with. So it's natural to think that they are the core normative concern, and it would be wonderfully parsimonious if reasons just were the end of the matter.

Still, we should want a little more to be said. What exactly is a reason-relation? How do we know when a fact actually does count in favour of something? How do we differentiate reasons from non-reasons? Related to these questions, some philosophers attempt to more precisely formulate reasons-relations. Scanlon, for example, argues that a reason is just any fact *p* that stands in a particular relation, namely R(*p, x, c, a*), a relation R which holds between a fact *p*, agent *x*, condition *c* and act or attitude *a*, where the place of the agent within the conditions makes the fact *p* a reason for *x* to *a*. Skorupski posits a similar but more complex reason-relation, R(π_i, *t, d, x,* ψ), where a fact π_i is a reason in virtue of standing in relation to an agent *x* and an act ψ, at a specific time *t* and, crucially, to a degree of strength *d*.[105]

In some ways those accounts of reasons-relations are vague and unilluminating – they don't really provide any means for identifying when something is a reason, or what it is specifically to count in favour of something. That's intentional. **Reasons-First** theorists believe that there is nothing to say about what *makes* a reason. That poses a problem.

Reductivists about reasons can offer specific analyses of what it is to 'count in favour' of something. A Humean, for example, might describe reasons as facts the truth of which explain how an agent's *Φ*ing will promote her desire(s).[106] For **Reasons-First** theorists though, that reasons 'count in favour' of things isn't meant as any sort of *analysis* of reasons. It's more a sort of paraphrase. To say that something 'counts in favour' is just another way of saying it's a reason. There's nothing more to be said about what 'in virtue of which' reasons count in favour of things; reasons *are* facts that count in favour of things, and they are *sui generis*.

There are many kinds of reasons though. In fact, *everything* happens for a reason – even things we often think we had *no reason* to do. Therein lies the problem.

[105] See Scanlon 2014, 31, 111-112 and Skorupski 2010, 38.
[106] See Schroeder 2007, 59.

§4.3.1 – Non-Normative Reasons

If we accept **Reasons-First** we should expect investigation into normativity to go something like as follows. I want to know whether or not *goodness* is a normative property. I ask then whether or not something's being good implies *reasons* – if, for example, what makes it good gives me any reason to act. I happen to know, for example, that 'In a Lonely Place' is a good film, in virtue of the fact that Humphrey Bogart gives a nuanced performance, that the plot is tense, and that the characters play well with each other. Since those features all seem to give me reasons to see the film, I conclude that goodness – at least in this case – is a normative property.

Yet there are many kinds of reasons. Normative reasons certainly, but many others. There are explanatory reasons, motivating reasons, instrumental reasons, etc. In general, any facts that *make sense* of why something happened, or that explain why doing something *would* make sense, we call reasons. For example, that humans have particular automatic nervous responses to stimuli is the *reason* why Tom jumps when startled, and a reason why it *makes sense* for Tom to do so. Arguably too it's a fact that *counts in favour* of Tom jumping – if we wonder whether or not it's appropriate for Tom to jump when startled, facts about his automatic nervous functions and how difficult they are to resist will feature significantly in that assessment. For sake of ease we'll call all those sorts of reasons *explanatory reasons*.

Generally though we *don't* want to say that explanatory reasons are normative reasons. There are intuitive reasons for that hesitance: that there is an *explanation* of why Fred punched his wife doesn't mean that he *ought* to have punched his wife. We should also be hesitant for the sake of coherence: if explanatory reasons can ground normativity, there can never fail to be sufficient normative reason for any action, since there is always sufficient *explanatory* reason for every action.

So even if the goodness-making features of 'In a Lonely Place' give me reasons to see the film, we need to ask what *kind* of reasons. If Humphrey Bogart's performance merely *explains* why my seeing the film would *make sense* – say because I just happen to respond to nuanced performances – then here the film's goodness is *not* normative. Yet explanatory reasons *are* reasons. If we want to say that nevertheless they don't ground normativity, and other reasons *do*, then we will need some further explanation of that fact. An explanation that *isn't* simply that they are reasons. In that case though, normativity can no longer be analysed exclusively in terms of reasons, but rather in terms of whatever makes up that further explanation.

The obvious response to the above is to deny that explanatory reasons are *reasons* at all, in the normative sense that we mean. That's exactly the response many philosophers give. Parfit, for example, argues explicitly that *incorrect* beliefs we have are not *reasons* for acting, specifically because they don't *count in favour* of acting in any way, even though those beliefs are certainly *explanatory reasons* for action.[107] At times it seems as though that is intended as a sort of knock-down refutation to the sort of argument I've offered, but it's not that easy. *Why* don't those explanatory reasons *count in favour* of acting? Or, more to the point, why *do* 'genuine' reasons *count in favour* when explanatory reasons don't?

The answer to that question can't be 'because they are reasons' in the general sense that we're familiar with, because explanatory reasons *are* reasons in that sense. Instead the answer must be because they are *normative* reasons – that is to say, they count in favour in a special, *normative* way. The **Reasons-First** theorist must assert that only *normative* reasons are reasons, and that normativity is to be understood in exclusively *their* terms. Now we're in trouble though. What made **Reasons-First** initially attractive is that we have some idea of what a reason *is*, separate from just 'a normative (grounding) fact'. That's what made **Reasons-**

[107] See Parfit 2011, pp.34-35.

First an informative analysis of normativity. Now though, instead of analysing normativity in terms of reasons, reasons themselves are understood in terms of normativity. What it is to be a reason is now just to be a normative fact, i.e., a fact that counts normatively in favour of doing something, or that entails normative truths. That gets things the wrong way 'round for **Reasons-First** theorists. Instead of proceeding to normativity from reasons, what we're now left doing is proceeding to reasons by way of normativity – once we know that *x normatively* counts in favour of *y*, we're entitled to call *x* a reason. The result is that reasons no longer offer any kind of *analysis* of normativity. Instead the term 'reason' has become just a label we apply to whatever fact turns out to account for normativity.

The result of the above is that **Reasons-First** can offer no kind of insight into normativity. Understanding normativity in terms of normative reasons is circular and trivial. Even if we accept that reasons are *sui generis* facts that normatively count in favour, all we've now asserted is that normativity is *sui generis*. That's no *analysis* of normativity, but rather a denial that normativity can be analysed. So I conclude that denying the status of explanatory reasons as reasons, and asserting that only *normative* reasons are reasons, does nothing to save **Reasons-First**. If we want to analyse normativity in terms of reasons, we will need some further explanation as to what a normative reason *is*, an explanation that isn't just 'something that normatively counts in favour'. We need an explanation of what makes normative reasons *normative*. A further reduction is called for.

To summarise, **Reasons-First** presents the attractive thesis that normativity should be analysed in terms of reasons. Yet there are many kinds of reasons, normative and non-normative. To differentiate normative reasons from non-normative reasons, a further explanation of what *makes* normative reasons normative is required. *That* explanation, however, will prove the key to explaining normativity, and so normativity cannot rest on reasons alone.

§4.4 – AUTHORITY

> *Obligation derives from the command of [a] legitimate authority … and it is this authority which gives normativity to moral claims* (Korsgaard 1996, 18-20)
>
> *Morality presents itself as something with ubiquitous and inescapable authority* (Joyce 2008, 104), *it is not something that a person may escape* (Ibid. 42)
>
> *Moral requirements apply to you even if you ignore them and try to renounce every concern whatever* (Wiggins 1995, 208)
>
> *Moral obligation is inescapable* (Williams 1985, 177)

Stephen Darwall paraphrased David Falk as asking the following question. "How … can rational agents be bound by morality in a way which is inescapable and which gives them rationally conclusive motives?" (Darwall 1989, 208). That's a very important question, because I take it that genuine, all-things-considered normativity *does* bind agents in just such a way. In fact, I think that's definitive of fundamental normativity.

Throughout the previous sections of this chapter I have, at length, discussed several theories which *fail* to explain what makes a property a fundamentally normative one. This section, §4.4, and the following §4.5 will now argue for my account of what *does*. Across these two sections I will develop a two-fold account of the constitutive features of fundamental normativity, such that any property *x* will make it the case that Φing ought to be done, all things considered, if and only if *x* instantiates those features. In this section I'll argue that for any property *x* to be fundamentally normative, *x* must have *authority*, or be *authoritative*. I will moreover argue for the following.

Authority. For any property *x*, *x* is fundamentally normative *iff* an idealised agent would derive overriding motivations to act from facts about *x*.

§4.4.1, *Platitudes*, reviews the general concept of authority. §4.4.2, *The Constitution of Authority*, employs clear cases of a lack of authority to argue that authority requires a specific sort of connection to agents, through which normative obligations can exert normative influence upon them, and explores naïve models for achieving such. §4.4.3, *Why* Do *I Care?* refines the criteria developed in the previous section, and ties normative influence to influence over motivational states. §4.4.4, *Motivational Authority*, summarises the preceding, and presents the formula for authority, linking it to the motivational states of ideal agents. §4.4.5 responds to anticipated objections. §4.4.6 concludes.

§4.4.1 – Platitudes

Intuitively, normative standards break up into two types. First are what I'll call *norms of practice*. Broadly speaking these are constituted by rules within specific domains. The rules of games, or of etiquette, or of aesthetics are all examples of norms of practice. These are norms that say one ought to lay a certain card, or dress a certain way, or eat a certain way, because those are the *rules*, at least of some domain. On the other hand, we have what I'll call *ethical norms*. These are the sorts of norms that have to do with reasons, or perhaps with goodness, or justice, or duty, or virtue, etc. These are norms that say one ought to do what one has reason to do, or to do what is good and not do what is bad, and so on. It's common to think that there is something different about these two categories. That thought, I argue, owes to the perception in the latter of *authority*. Exploring what seems different about those two categories will help to get a grip on what that is.

A common feature of norms of practice is that, generally speaking, they can be rationally ignored. You might, according to some external arbiter, play a game wrong, or use the wrong fork, or wear the wrong clothes. As a rule though, you needn't care about those infractions. The mere fact that some practice has, as a rule, that one should always keep the bottom button of one's waistcoat unbuttoned, implies no practical incoherence on the part of an agent who decides they prefer to do otherwise. Only if you have some additional reason to *want* to care – say because you want to be accepted by a society with particular aesthetic or etiquette rules – do norms of practice *matter*. Even then, you might, at any point, *cease* to care, or decide that your time is better spent on other endeavours, and so rob those norms of their importance. Norms of practice are *non-authoritative.* If you aren't interested in them, they aren't important.

Very often however, we make normative judgments that don't seem to be of the above sort. These are, paradigmatically, judgments about what we have reason to do, or perhaps what is good to do, or right to do, etc. Unlike with norms of practice, what seems definitive of norms grounded in *these* sorts of considerations is that agents are *not* entitled to disregard them. As discussed above, deciding to button the bottom of one's waistcoat in contravention of Eton's aesthetic demands can be the act of a perfectly rational agent. Shirking the demands of *reasons*, however, or perhaps of goodness, or duty, is different. If it is bad and unjust and contrary to normative reasons to murder a man on a whim, then it doesn't matter how I happen to feel about it. I really ought not to do it, regardless of what I think. If I do it nevertheless, that is evidence of some important deficiency in me, and not just in an abstract sense. Failing to do what I have most reason to do, in whatever way we cache that out, implies something *practically* defective in me, in my capacity to move from perception of fact to action, or to reason about what ends to pursue.

The difference between those two types of standards is that the latter have *authority*, while the former do not. Authority is what explains the impermissibility of disregarding certain standards. That difference is also a hallmark of the fundamentally normative. Norms of practice are not fundamentally normative, specifically because they can be rationally disregarded. They don't genuinely determine what we ought to do, because at any point we *have an out*. A simple shift in perspective eliminates their relevance to our practical rationality, and their relationship to our motivations. Fundamentally normative standards, therefore, must be so at least partially in virtue of being unable to be so disregarded. They have, or are grounded in properties that have, *authority*. If we genuinely ought to do what is good, for example, it *must* be irrational to disregard goodness.[108] Fundamentally normative properties *determine* our practical obligations, and it is through their authority that they do so.

It is difficult to see how any property could be so unavoidably relevant. Authority, therefore, sets an extraordinarily high standard for a property to meet, if it is to be fundamentally normative. Some philosophers even point to the need for it as evidence that fundamentally normative properties can't exist.[109] The rest of this chapter then will investigate and argue for what authority is constituted in, and what any property must do to possess it.

§4.4.2 – The Constitution of Authority

To pinpoint what authority consists in, it will be helpful to consider cases in which it seems *absent*. If we can identify what non-authoritative properties *fail* to provide, in virtue of which we identify them as non-authoritative, then we know what an authoritative property must supply. In this section I argue that a property is picked out as non-authoritative when it provides

[108] Unless, perhaps, it is disregarded in favour of some *other* fundamentally normative property.
[109] See, for example, Smith - "morality … may well be incoherent, for what is required … is a strange sort of fact … a fact whose recognition *necessarily* impacts upon our desires" (Smith 1994, 11) – or Joyce – "if *x* morally ought to *φ*, then *x* has a reason for *φ*ing regardless of whether *φ*ing serves his desires or furthers his interests … But there is no sense to be made of such reasons … Therefore, *x* is never under a moral obligation." (Joyce 2008, 42). Likewise, famously, Mackie in *Inventing Right and Wrong*, 1977.

no answer to the 'why should I care' question – a question rightly raised by any indifferent agent upon being told that they ought to Φ. I argue that a property fails to provide such an answer when there is no *connection* between that property and the agents that it purports to govern, so that there is no normative force behind the derived standards. I then explore what sort of connection, and what sort of force, might do.

There are many obviously non-authoritative properties and standards. The rules of chess or standards of aesthetics from the previous section are examples. They, though, are *so* implausibly authoritative that they don't help narrow down exactly *why*. There are more plausible-seeming properties that, I think, also come out to be non-authoritative. Beginning with one will help narrow down what's missing more precisely.

Take goodness. Moreover, take a specific, reductive analysis of goodness, whereby goodness consists in the maximisation of agent-neutral happiness. *Carolinus* believes that goodness, understood in that way, is fundamentally normative, and informs *Ommadon* of such. Carolinus tells Ommadon that what Ommadon ought to do is maximise agent-neutral happiness, and that, in fact, there is a specific action that will do so. All Ommadon needs to do is *completely anonymously*, never telling anyone, burn the £10 note in his pocket. If Ommadon does so, an entire planet of people across the galaxy, whose existence will *never* bear in any way upon any of Ommadon's subsequent affairs, will be raised from abject misery into prosperity and happiness. That will maximise agent-neutral happiness, and so, Carolinus argues, Ommadon ought to do so.

Ommadon reviews Carolinus' research, and finds it checks out. That planet full of people exists, they are thoroughly miserable and, through an almost unbelievable chain of events, Ommadon burning his £10 will improve their happiness almost beyond measure. Convinced of those facts, Ommadon nevertheless says to Carolinus, '*so what*?' Ommadon

doesn't *want* to burn his money, he likes his money, and he doesn't care about those far-away people. 'What is it to me that they're miserable?' he asks. '*Why should I care?*' [110]

Let's assume for the moment that Carolinus can't answer that question. What could he say? He can't simply repeat that many lives will be improved at a negligible cost – that's exactly what Ommadon doesn't see a reason to care about. Nor can he just reiterate that it would be *good* to do that – the same objection applies. Likewise, he can't just insist that good things *are* what Ommadon should care about – Ommadon wants to know *why*, and even if he accepts the stipulation, he can ask instead why agent-neutral happiness is *good* then, i.e., why he should care about *it*. The answer to Ommadon's question can't be one about which he can coherently ask the same question. If no adequate answer to Ommadon's question is forthcoming though, then agent-neutral happiness is in trouble. For any property to be authoritative, and thus fundamentally normative, it must provide and answer to the 'why should I care' question.[111]

So what's missing in the agent-neutral happiness example such that said happiness fails the 'why should I care' question? Here's a partial answer. The example provided is deficient

[110] The 'why should I care' question has been the focus of a number of philosophers. Korsgaard's 'normative question' (Korsgaard 1996, ch.1) and Mark Johnston's 'So what?' objection (Johnston 1989, 157-8) make the same observations. See also Wedgewood 2007, 70-1.

[111] Shamik Dasgupta, in 'Non-naturalism and Normative Authority', agrees. He argues – specifically against non-naturalist realist accounts of normative properties – that there are two elements to any theory that identifies normativity with some particular property. "First, there is an ontological claim that in addition to all the natural properties and relations out there, there is also a … property P. And second, there is a [claim that P ought to be promoted]. … Grant for the sake of argument the ontological claim that there is this … property P that some things have and others lack. The question is why it should be promoted. What is it about P in virtue of which we should promote things that have it? Why shouldn't we promote things with some other property instead?" (Dasgupta 2017, 6)

Dasgupta's argument says that even if we grant the stipulation that *z* explains *ought*, such that when we have *z* we have normativity and that whatever *is* normative we will call *z*, we still need an explanation as to why any specific property *p* deserves to be called *z*. We need to say *what it is* about *p* in virtue of which we *really ought* to promote it. Dasgupta wields this argument against theories that argue that *sui generis* Reasons are Authoritative because Reasons just are the sort of thing it's rational to respond to. He argues that even if we accept that being rational = responding to Reasons, when one makes the additional claim that R is a reason, they must still explain what it is in virtue of which R *is that* which it's rational to respond to. One can't simply respond that R *is* what it's rational respond to *because* it's a Reason; that's not 'playing fair' (Idem. 8-9). Likewise, *x* can tell *y* all she likes that what *y* should do is promote agent-neutral happiness, but *x* can't explain that by appealing to the idea that agent-neutral happiness just *is good*, i.e., just *is* what *y should* promote. *X* needs to explain what it is in virtue of which her toy property *really is* what *y* should promote, how it answers the 'why should I care' question.

because maximising agent-neutral happiness will *make no difference* to Ommadon. It will make no difference to him, because the toy normative property, agent-neutral happiness, has no influence over anything in Ommadon's life. Nothing in his life depends upon upholding the derived standard. If success or failure in promoting any property *x* will *make no difference* in any aspect of an agent's life, then that absence of impact in itself will always support their sceptical indifference. No difference can be made, meanwhile, in the absence of some causally efficacious connection between *x* and the agent, such that adhering to or violating the relevant standard derived from *x* will cause, or *prevent* the causing, of some specific effect upon the agent. The authority of *x*, therefore, must be constituted in such a connection.

Of course not just any effect will do. Imagine a world in which it happens to be a brute fact that when one fails to abide by the dictates of etiquette and promote adherence to social standards, one inevitably feels a small and neutral pressure upon their shin. This is a change in the agent's affairs, but one the agent may surely still claim indifference about. A particular *kind* of effect is called for, a *normative* effect, one that it would be practically incoherent not to care about. Now though it seems we're right back where we started, asking what kind of thing it would be practically incoherent not to care about.

I don't think we are where we started however. I've argued that authority requires a connection to an agent, which is causally efficacious in a way it would be practically incoherent for the agent not to care about, i.e., in a way that will answer the 'why should I care' question. To proceed, we should take a closer look at what it means to be practically incoherent, and what the 'why should I care' question is really asking.

One way for it to be irrational not to care about *x* is for it to be the case that *x* is required in order to achieve the things one actually *does* care most about. Relatedly, there are at least two different ways of reading the question, 'why should I care about *x*?' One is as asking for further normative justification, to answer the 'should' part of the question. Another, however,

is as synonymous with the question 'so what? What does *x* matter to me?' *That* question is a statement of indifference, suggesting that it isn't practically incoherent not to care about *x* *unless* an answer can be given as to why *x actually* matters to me – why it bears upon something I *do* care about. I think the latter reading is, generally, the correct one. Reading it that way, moreover, suggests what sort of connection authority must be grounded in – a connection to the motivational sets of agents. In the following two sections, I will argue for that conclusion.

§4.4.3 – Why Do *I Care?*

I start with a strong proposal: when someone asks 'why should I care about *x*', what they are *really* asking is 'why *do* I care about *x*?' By that of course I don't mean 'why do I explicitly value *x*' – I don't – but rather 'how does *x* bear upon something I do care about?' I can see no way around arriving at that conclusion, assuming that 'why should I care?' is intended as a question which could have an answer. This is, more or less, simply because if the fact that I don't care is enough to generate an explanatory challenge – as in the agent-neutral happiness example – then further appeal to things I don't care about will never answer that challenge.[112]

Further, it turns out that the 'why should I care' question *should* be taken as predicated on indifference – it's not merely that I see no *reason* to care, or don't see *why I ought to* care, but that I don't see that I *do* care about what is called a reason, or what is taken to explain ought. I see no reason that actually speaks to anything I care about, and so I am justified in shrugging my shoulders and ignoring the alleged ought. Trying to answer the fact that I don't care about *x* with further appeal to things that I don't care about is an endless quest. So if you

[112] It should be noted here that the way I am using 'care about' is not synonymous with 'explicitly value'. There are many things we can care about, without realizing it. I might not realise how much I care about a united European society, until it is taken away. That isn't to say that I didn't care about European harmony beforehand, but rather that my care for it was implicit – unrecognized, but influential nonetheless.

take your normative property to be authoritative, what you need to demonstrate to me is why I *do* care about your normative property.[113]

What's more, further expanded the question actually looks to be asking 'why do I care *most?*' since if the normative authority of *x* is predicated upon actual concerns, a greater concern for *y* would justify disregard for *x* in many situations. If I *do* care about etiquette, but care much more about self-preservation, I seem justified in some degree of rudeness if it saves my life. Thus in the case that I actually care more about *y* than *x*, it's *not* the case that it would be irrational not to act on *x*; I could very rationally disregard *x*, in favour of *y.* So, in that case, *x* would not be authoritative. What you need to demonstrate to me then is two-fold. In order for property *x* to be authoritative, it must (i) bear upon some concern *c* that I actually care about and (ii) it must be irrational for me to care about some other concern, *c**, *more* than I do *c*.

This sounds like a tall order so far. Even if I accept that in order to be authoritative a property must bear upon something we actually care about, it may seem difficult to believe that there is anything I care about in such a demanding way, such that I must care for it over and above anything else. I think though part of that difficulty is due to how loosely we speak about *caring*. I think that we can come to terms with that demand, by considering more analytically what it is to *care* about something.

My view here is straightforward. To care about something is, broadly speaking, to be *autonomously motivated* by it, positively or negatively. By that I mean that to care about *x* is (i) to be motivated by *x*, and (ii) such that I would *choose* to continue to be motivated in such a way, upon critical reflection, given the choice, and *ceteris paribus*. That may seem opaque, but the reasoning behind it is plain if we consider the difference between *caring* about something and being *merely* motivated by it. I might be motivated by many things I would, on

[113] Both the necessity for grounding normative properties in things we *actually* care about is echoed by a number of philosophers, many already mentioned. See Korsgaard (1996), Williams (1979), Smith (1994) or Joyce (2008).

reflection, choose not to be motivated by – by the need for drugs, for example, while addicted. On the other hand, I am motivated to make my wife happy, and, given the choice, I would choose to continue being motivated so, because I *care* for my wife, and I *care* about her happiness, in a way that I do not care about, for example, heroin. That's what it *is* to care about *x*; it's to be motivated by *x* in a way that would survive upon informed and critical reflection on my motivations, at least in part.[114]

Thus I argue that when one asks 'why should I care about *x*?', what one really asks is 'how does *x* bear upon my motivations, and why would those motivations remain upon informed, critical reflection?'. I will explain now how understanding the phenomenon of caring in this way helps to understand authority.

§4.4.4 – Motivational Authority

The shift from the nebulous language of caring to the language of motivation provides the key to moving forward. A property is able to answer the 'why should I care question' – that is, the question of why one should care about the reasons to act that derive from the property – when an influential connection can be demonstrated between the property and the motivational states of agents to act. I argue that an account of authority in terms of overriding motivations to act best accounts for all the conditions so far discussed. So, finally, I propose that the following is true.

[114] There may be much more to the phenomenon of caring. Shoemaker for example, in 'Caring, Identification, and Agency', points out that a feature of caring about *x* is being emotionally *open* or vulnerable to facts about *x*: i.e., to care about my wife is, generally, to be disposed to be made happy by improvements in her affairs, and to be made sad by the inverse (Shoemaker 2003, 91-92). That seems right, yet for my purposes here it is, I think, irrelevant – the crucial feature is that which I've now argued for. Such a feature – an interconnectivity with deeply held and enduring motivations – is likewise supported by Shoemaker himself, alongside others like Frankfurt and Watson. See Frankfurt's 'Freedom of the Will and the Concept of a Person' (Frankfurt 1988) or Watson's 'Free Agency', (Watson 1982) or 'Free Action and Free Will' (Watson 1987).

> **Authority.** For any property x, x is fundamentally normative *iff* an idealised agent would derive overriding motivations to act from facts about x.[115]

Of course, specifically theories, like Attributivism, will additionally require that *specific* overriding motivations be derived, not just any motivations. Nevertheless, those details may be filled in by whatever specific theory is under consideration. The formula I've proposed is that which must minimally be satisfied by any theory that posits a fundamentally normative property. Left by itself however, the above definition should raise concerns, due to its unexplained appeal to idealisation. Thus,

> **Ideal Agent.** The idealised agent y is a version of y who is (i) fully informed about their motivational states and their circumstances, and (ii) motivationally coherent, such that their actions cohere with their informed judgments. [116]

The intention behind **Ideal Agent** is merely to remove any internal impediments to the efficacy of the actual agent's motivations. Importantly, no *new* dispositions are assumed; the ideal is merely that which is ideally situated to achieve the aims of the actual.

---

[115] I find it interesting that some views on authority do *not* take this shape. Andrew Seppielli, for one, has recently written in 'Consequentialism and the Evaluation of Action qua Action' about authority as a sort of ranking property attained by systems of normative evaluations that take less 'for granted' than others. I think though it's a significant problem with Sepielli's account that it lacks the sort of authority I have just outlined, and is thus oddly toothless, offering perhaps a coherent account of the authority of certain norms over *other* norms, but providing no explanation of the authority of normativity itself. See Sepielli 2019. See also Thomson's intuition that the 'more general' is somehow *weightier* than the 'less general' (Thomson 2008. 213).

[116] Many ethicists develop accounts of ideal agency. Mine is in some ways more demanding than is common. Most accounts demand that the agent be fully informed about their circumstances; mine demands they also be fully informed about their own motivations. Likewise most accounts demand coherence, in such a way that the ideal agent has no motivations which fail to cohere with one another. I am sceptical of the need for such coherence; instead I demand only that ideal agents are practically rational, suffering no cognitive errors which would keep them from acting on their motivations, whatever those might be. For more accounts of idealization, see Williams 1986, Lewis 2000, Firth 1952, or, particularly, Railton 1986, where an idealized agent is one who is "fully and vividly informed about himself and his circumstances and entirely free of cognitive error or lapses of judgment" (Railton 1986, 16).

Let's look at an example to bring all the above together. Suppose that pleasure, or the property of being pleased, was proposed as being fundamentally normative. As I've argued, for that to be the case pleasure would need to be authoritative. What that means, I've argued, is that it must be able to answer the 'why should I care' question, and provide a reason why it would be practically incoherent not to act in a way that promotes being pleased. For *that* to be the case, I've argued, being pleased must bear upon some actual concern of agents, and do so in such a way that it would be incoherent to care about something else *more* than being pleased. And, with our new understanding of caring, what *that* means is that agents must be motivated towards being pleased, in such a way that their ideal selves – fully informed and motivationally coherent – would be overridingly motivated to promote pleasure, and to *not* take actions that would conflict with or impede its promotion.

Although that is a demanding requirement, arguments to support it might be of the following sort. For example, we might argue that the drive to attain pleasure is a basic, non-instrumental motivation *in virtue of which* other motivations are held; if we only desire food, or companionship, or security, *because* they bring pleasure, then motivation towards pleasure in implicit in those other concerns, and promoting them at the *cost* of pleasure would be incoherent. Alternatively, we might argue that our desires are generally *conditional* on the promotion of pleasure. In that case, it would be incoherent to pursue a desire unless doing so would be pleasurable, in the same way that, if I desire to go to the pub only on the condition that it's quiet, it would be incoherent to promote going to the pub if the pub were *not* quiet. In that case, all our *actual* motivations would be influenced by pleasure, and it would be incoherent to promote any of the former at the cost of the latter.

Although those scenarios seem far-fetched in the case of pleasure specifically, it isn't *prima facie* implausible that *some* property could serve in such a fashion, and so my account of authority isn't, I think, *overly* demanding. Any theory of normativity must provide an

argument for how its candidate fundamentally normative properties connect to motivations in the way I have laid out.

There is support for the sort of theory I've now argued for. Although not many philosophers have argued for **Authority** in exactly the way I have, many *have* argued that normativity depends upon motivation and/or idealisation in a way that's *something like* what I've described.[117] Here's a short list.

> "A conclusive reason would be … unavoidably stronger than all opposing motives (Falk 1986, 39). [T]his is meant by a 'dictate of reason': an impulse or will to action … which derives a special forcibleness from [the fact that] no further … 'reason' would change or dislodge it (*idem* 190). [People are] under obligations when … they have … a specially compelling or deterring motive for doing or not doing them" (*idem* 184).

> "[A person is obligated in the case that] when some way of acting is [wrong] and reflection reveals that fact, the person finds that she must reject that way of acting, and act in another way" (Korsgaard 2010, 4.3.7).

> "[Morality requires] a strange sort of fact … a fact whose recognition *necessarily* impacts upon our desires" (Smith 1994, 11).

> "If a person really does have a reason to do something, then it must be something he could be made to care about, something that could move him to act … a reason is, by nature, the sort of thing such that if a person has one, it cannot leave him cold" (Heathwood 2011, 96).

[117] Judgment Internalists, Reasons Internalists, and even Error Theorists in particular all tend towards thinking that some kind of intimate, internal connection between normative facts and motivation is required for normativity. The strangeness of that requirement is what leads people like Joyce or Mackie to reject the truth of normative claims – see Joyce (2008) or Mackie (1977). See, among many others, Smith (1994), Bjorklund et al (2011) or Suikkanen (2018), for particularly cogent arguments for Judgment Internalism. For explanations of Reasons Internalism see Williams (1979) or Finlay (2009).

> "*y* is a substantive reason for/against … *x iff* we are disposed stably to take it to be so under conditions of increasing information and critical reflection" (Johnston 1989, 162).

All the philosophers above link normativity to necessary and particularly significant influence over the motivations and desires of agents.[118] Likewise they argue that influence must be recognised by those agents under increasingly *idealised* conditions, or upon informed reflection. Even more philosophers have come to recognise the need for the relevant motivations to be ones we cannot coherently disregard, or that have a particularly priority over other motivations.[119] Given that extensive precedent, and the arguments I have put forward, I conclude that being authoritative in the particular, motivating way I have laid out is required for any property to be fundamentally normative. Any such property must supply an answer to the 'why should I care' question, by appealing to an idealised and overriding motivation that would emerge given informed progression through the agent's actual motivational set.

§4.4.5 – Objections

There are numerous compelling objections against the sort of account I've presented. Specifically, there is significant reason to think any account of normative authority that relies on a relationship to the motivations of agents is *implausible,* largely due to the plausible objection that just because we *are* motivated to do something, even ideally, doesn't mean we

[118] We mustn't forget, of course, Bernard Williams, perhaps the most influential advocate for grounding normativity in the motivational sets of agents, and to whom I am particularly sympathetic. See Williams' 'Internal and External Reasons' 1979, and particularly Finlay's 2009 interpretation in 'The Obscurity of Internal Reasons', whereby Williams is taken to argue that a fact is a (normative) reason for an agent if it explains an action they *would* have taken – or have been somewhat motivated towards taking – in an ideal state corrected for error and ignorance. See also Finlay and Schroeder's 2017 SEP article 'Reasons for Action: Internal vs. External', particularly 2.1.2.

[119] See particularly constitutivist or constructivist accounts which ground normativity in motives that are constitutive of rationality, and so inform any further motivation held by rational agents. See Katsafanas (2013, 2018), Korsgaard (1999, 2008, 2009), Smith (2013), Street (2008, 2010, 2012) or Velleman (2000, 2009).

should be. Two families of objections in particular spring to mind. I will present, and then respond to, each in turn.

O1. *Prima facie*, our motivations often fail to track what we 'obviously' ought or ought not to do. It seems possible that an agent's motivations might be such then that their ideal self will likewise fail to track these 'obvious' moral truths. Parfit, for example, imagines a scenario wherein an agent cannot be brought to care about pain they will suffer on future Tuesdays, despite 'obviously' having reason to avoid such pain. Williams, likewise, envisioned an abusive partner who can't be brought to care about the welfare of their spouse, despite having 'obvious' reasons to do so.[120] The key idea in all the above cases is that, *prima facie*, certain standards of action are normative, even though it seems true that ideal agents might fail to be motivated accordingly, and so my view of authority is wrong, since it grounds normativity in those same motivations.

There are two ways of responding to **O1**. The first way I will call **R1** and involves denying the coherence of the supposedly counter-intuitive examples provided. The second, **R2**, involves accepting their plausibility, but denying that they constitute an objection. I'll go through both of these strategies now.

R1. We should grant that agents *actually* do at least sometimes develop counter-intuitive motivations. However, **Authority** is predicated on the motivations of the **Ideal Agent**, not the actual one, and it doesn't seem obvious that **Ideal Agent** would share those motivations. There

[120] Parfit 1984, 124, and Williams 1989. Less extreme examples are also easy to come by. Imagine someone who understands theoretically that they have most reason to eat a sandwich, that eating a sandwich is in their best interest, that it will satisfy their desires, and that there is no reason *not* to eat a sandwich, and who yet develops no correspondingly predominant motivation to eat a sandwich. The motivations of that agent seem to fail to track what they ought to do, yet it's plausible that such might be the case, and so it seems like what they *are* motivated by can't determine what they *ought* to do. For an example of this sort of *akrasia*, see Silverstein 2017.

are, as I see it, two reasons to think the provided examples are incoherent *qua* **Ideal Agent**. Those reasons emerge when we ask *why* an agent might ever lack motivation in the ways just discussed. It seems likely that any time those examples actually *are* plausible – that is, any time an agent can realistically be imagined as failing to be motivated by what they 'obviously' should be motivated by – that plausibility will be explained by either (i) the agent involved suffering from psychological malady or (ii) the agent involved lacking information. As **Ideal Agent** heads off both of those possibilities however, there is no reason to believe that the motivations of an idealised agent ever would fail to track those 'obvious' moral truths. By way of illustration, consider the following explanation of agents failing to be properly motivated.

(i) Fred sits at home, hungry. It strikes him that the best way of resolving that hunger would be to get up and make a sandwich. He knows he has all the ingredients to do so at hand, that doing so would take but a moment, and that he even likes sandwiches. Still, Fred doesn't move to go make a sandwich. Instead, Fred waits for his partner Wilma to come by, at which point – as has becomes his custom – he demands that she make him food, and hurls invectives and abuse at her, even becoming violent, when she argues, or fails to comply fast enough. He knows that he shouldn't, that Wilma doesn't deserve his abuse, and that he is being cruel. He can't, however, bring himself to care.

Fred seems like a person we might plausibly encounter. What else we might plausibly expect to find however, is an accompanying *explanation* of Fred's behaviour. A likely such explanation is some sort of psychological malady. The picture of Fred doesn't seem like the picture of a *happy* person. Rather, Fred seems like someone suffering from pathological anger issues, or perhaps depression. Although Fred knows what is in his interest, every time he would act upon it, he

instead is waylaid by rage or malaise. Crucially though, people who suffer from these ailments aren't *devoid* of what we would think of as the proper motivations. In fact, having those motivations but being for whatever reason unable to act on them best explains the pathology of those conditions.[121] That though makes this a case of motivational incoherence. The appropriate motivations are there, and a fully informed Fred would judge that he ought not to abuse his wife, but Fred's ability to act is impeded by a psychological malady that prevents him from acting as he judges he should. The **Ideal Agent** is, by stipulation, free of such conflicts however, and so that problem is resolved.

(ii) Parfit's Future-Agony-Indifferent agent presents a different sort of case than considered above. It seems theoretically possible that such an agent could be indifferent to her future-agony while being fully psychologically capable. Still I think any plausible case of such an agent would have to be explained by a lack of information.

Grant first that the agent does, generally, have motivation to avoid being subject to pain. If told, absent any temporal data, that she is going to suffer pain if she *Φ*s, she will avoid *Φ*ing. What could explain her indifference when told it will be a Tuesday on which she is in pain? Something must have changed in her understanding of the situation. The only conditions necessary to activate her avoidant motivation generally though are that she understand (a) there will be pain and (b) that it will happen to her. Assuming that there are no further exculpatory facts about Tuesdays to rationalise the agent's change in motivation, if she nevertheless fails to be motivated to avoid future-Tuesday-pain, she must then, in

[121] See Law 2009, Roberts 2001, Lewis, G. 2006 or Ratcliffe 2014, particularly ch.7.

some way, fail to understand that (a*) the pain on Tuesday will be *pain*, or (b*) that the agent on Tuesday is really *her*. Accepting for the moment that those are both true however makes the explanation of her indifference a lack of – or misunderstanding of – the information. **Ideal Agent** though has all the relevant information, and so would not be so counter-intuitively indifferent.[122]

The cases above show that the supposedly plausible cases of ideal agents who fail to have the motives they prima facie ought to have aren't as plausible as they seem. They're often incoherent. Most examples – believable examples – of agents with deficient motivations will typically be explained by lack of information or capability. Thus, by stipulation, those agents can't be *ideal* agents, i.e., fully informed and capable ones, and so such examples are not objections to my view.

R2. We might be unconvinced that *all* cases of agents failing to have the appropriate motivations can be explained by psychological conflict or lack of motivation. Perhaps it is possible for an agent to be fully informed, fully capable, and simply unconcerned with anything we generally feel that he ought to be concerned with. If any plausible cases of such deficient motivations still exist however, the appropriate reaction is to bite the bullet and accept their

[122] Smith makes a similar argument that may be salient here. In 'Desires, Values, Reasons, and the Dualism of Practical Reason', Smith asserts the following. "If someone has an intrinsic desire that p, and an intrinsic desire that q, and an intrinsic desire that r, and if the objects of the desires that p and q and r cannot be distinguished from each other and from the object of the desire that s without making an arbitrary distinction, then she has an instrinsic desire that s" (Smith 2009, 120). Smith's argument is another way of refuting the coherence of Parfit's example. If Parfit's agent has a desire to avoid pain on Monday, and there is no way to differentiate between being in pain on Monday and being in pain on Tuesday save for making an arbitrary distinction based on the day of the week, then it must be understood that the agent *also* has a desire to avoid pain on Tuesday. Parfit's example, as such, is incoherent. I am attracted to Smith's argument, but I don't pursue it further here, for two reasons. (i) It's unclear to me what should be taken as arbitrary and non-arbitrary as regards making distinctions – if the motivations of Parfit's agent *can* be said to concern not just being in pain, but being in pain on certain days, then it's unclear in what sense the day of the week is arbitrary, re: their motivations. (ii) It seems possible that the motivations of an agent *might* make arbitrary distinctions based on days of the week. If so, it's unclear why arbitrariness should count against them, normatively speaking.

possibility, without accepting that they constitute an objection to the account of **Authority** I have laid out. Most specifically, I argue that any such cases will be predicated on cases so abstracted from the norm that they sacrifice their intuitiveness.

Parfit would object to my treatment of his future-agony-indifferent agent above. He foresaw the sort of argument I gave, and insisted that his agent is to be understood as having no "false beliefs about personal identity," and as agreeing "that it will be just as much him who will be suffering on Tuesday" (Parfit 1984, 124). I've already argued that the idea of an **Ideal Agent** with those views *and* future-Tuesday-indifference is incoherent. Even if we accept the possibility of Parfit's stipulation though, I think it is less of an objection than it seems.

Sharon Street has argued that examples like Parfit's prey on a superficial intuitive similarity to plausible cases, a similarity that can't be born out. She thinks if we really examine those examples, we will find the agents so alien, so far removed from every day circumstances, that the intuitiveness of the idea that they are failing to do that which they obviously should do dissipates.[123] I agree. We think Parfit's agent is acting poorly because we are prompted to imagine them as being just like us, with the minor tweak of not caring about future-Tuesday-agony. *We* would be wrong if we didn't care, so *they* must be too. But Parfit's agent *isn't* just like us, it can't be. We *do* care about future-Tuesday-agony, we *must*; it's part of caring about our own suffering at all. To explain the indifference of Parfit's agent, we must imagine an agent so alien that it is no longer obvious *what* they ought to do.

To conclude my response to **O1** then, either cases of deficient motivations are explained by a *lack* of idealisation, *or* the supposedly deficient agents are such alien creatures that intuitions about what they ought to do break down, and it ceases to be clear that their motivations *are* deficient. Perhaps it is possible for there to be a fully informed and capable agent whose motivations regarding pain really do fluctuate, seemingly arbitrarily, with the days

[123] See Street, 2009, 'In Defense of Future Tuesday Indifference'.

of the week. If it's true though that that agent perceives no facts incorrectly, is fully capable of acting on those facts, and yet still has *no* motivations that can be appealed to in order to make them act on those facts, then that creature is so different from us that it isn't clear what it ought to do. Different types of creatures have different reasons. That, though, is okay, and not an objection to my view.

O2. A second family of objection focuses not on the unintuitiveness of motivation-based accounts of normativity, but on the role of idealisation. My account of **Authority** relies on the motivations of **Ideal Agent**s to answer the 'why should I care?' question. It seems reasonable though to ask 'why should I care' about the Ideal Agent? Appeals to Ideal Agents seem just to appeal to ways we *would* act if we *were* in different situations. Why should we care about how we *would* act in different circumstances? We're not *in* those circumstances! The gist of the objection is that there is no explanation why the facts that motivate the ideal version of you should also motivate the actual you.

David Enoch makes that sort of objection in 'Why Idealize?' He thinks that appeals to ideal agency are only justified if the ideal agent can be construed as tracking some independent normative fact. That's because he takes it that the 'natural answer' to 'why idealize' is to say that idealisation best allows us to track *independent* facts, in the way an *ideal* watch best tracks independent facts about time. It's only because there *are* independent facts that we care about what the idealised watch says. If there weren't independent time facts, the readings of the so-called ideal watch would be no more significant than any other watch. So, thinks Enoch, we can't ground normativity in the motivations of ideal agents themselves: just like in the case of the watch, unless those motivations track some further normative fact, so-called idealised

motivations would be no more significant than actual motivations.[124] The motivations of our ideal selves, therefore, can't *themselves* be part of what determines what we ought to do.

R3. **O2** is a compelling objection. If the motivations of the **Ideal Agent** track *further* normative facts, then her motivations aren't themselves part of the grounds of normativity. If they don't, though, then why should we care about them? Well the most straightforward answer is that we should care because we *do* care. The **Ideal Agent** is just the agent that is ideally situated to respond to what we *do* care about – it is poised to fulfil our *actual* motivations, whether or not we realize we have them. And that demonstrates the flaw in arguments like Enoch's against idealisation.

Enoch's arguments are aimed at accounts that see ideal motivations as part of the fundamental grounds of normativity. That's not the sort of stance we should take – not exactly. The role of **Ideal Agent** is derivative, not fundamental. The **Ideal Agent** is important because it reliably tracks our *actual* motivations, which we can fail to *actually* have a clear picture of. It is that *actual* motivational set in which normativity is grounded. In this case though, **Ideal Agent** actually does function in just the sort of way that Enoch thinks is coherent. In his own terms, we reference **Ideal Agent** *qua* normativity in just the way we reference an idealised watch *qua* telling the time.[125] We respect its determinations not for their own sake, but because it is situated to track that which is actually relevant.

I've argued that ultimately the answer 'I unavoidably do care' is an answer to 'why should I care', and we *do* unavoidably care about what the ideal agent would do, because the ideal agent just does what we care about. To question whether we should care about the ideal agent's active motivations is akin to questioning whether a mathematician seeking answers to

[124] See Enoch 2005, particularly section II.
[125] Enoch 2005, 762.

a mathematical puzzle should care about the knowledge of a version of herself who knows the answer – the knowledge of her ideal agent is exactly what she *is* seeking. Being guided by the sort of person that best knows how to get what I want *is* the best way to get what I want, and I can't help but care about *that*. The way that my view employs idealisation is not vulnerable to the preceding objection.

§4.4.6 – Conclusion

Normative authority is what provides the answer to the 'why should I care?' question. In this section I have argued for that conclusion, and provided a conceptual structure for the constitution of Authority as a causally efficacious connection to motivational sets. The result is the following:

> **Authority.** For any property *x*, *x* is fundamentally normative *iff* an idealised agent would derive overriding motivations to act from facts about *x*.

Where

> **Ideal Agent.** An idealised agent is one who is (i) fully informed about their motivational states and their circumstances, and (ii) motivationally coherent, such that their actions cohere with their informed judgments.

I've responded to a number of significant objections to making normativity dependent on motivations of even idealised agents. I conclude finally that the account I have provided is coherent. Moreover I think it is the *only* sort of account that can respond to the 'Why should I care?' question, and thus is a necessary condition on normativity itself.

§4.5 – ACTION REGULATION

The previous section has argued that any property is fundamentally normative *only if* it is authoritative. I may even go so far as to say a property is normative *if and only if* it is so. It turns out though that even should that be the case, it isn't enough to establish any sort of *meaningful* normativity. At this point we may have normativity of a sort, but not of a sort *worth talking about*. The normativity established by normative authority may come out to be trivial, or impossible *not* to achieve. It may turn out, for example, that *any* action satisfies the standards derived from the candidate property. Any theory of normativity however that results in every action being normatively equivalent is practically indistinguishable from one which posits no normativity at all. Every action is left on the table. In order for normativity to be worth talking about then, it needs to play an additional and significant role in regulating the set of possible actions available to agents so that some possible actions are ruled out or excluded from the set of permissible acts. Thus a fundamentally normative property must satisfy the following.

> **Regulation.** For any property *x*, *x* is fundamentally normative *only if x* provides a standard of action that some, but not all, possible actions fail to satisfy.

It must, in short, be *act regulating*.[126] If, for example, *goodness* were to be a normative property, it would have to be the case that we could derive a standard of action from that property – i.e., *be good* – that some potential actions would violate, or otherwise fail to satisfy.

[126] Many ethicists argue that normativity must be action *guiding*. My argument here is slightly different. The capacity to be action *guiding* is typically linked to epistemology – a property or standards is action *guiding* if there is some way we can *know* or be moved by the perception of it in such a way as to take *specific* actions, and not take others. Act *regulation* is a metaphysical condition. It must be the case that, *if* we knew the normative facts, we *would* discover that some actions are ruled out and others ruled in. *How* we come to know those facts is a question I don't deal with here.

That is a very weak condition. All theories of normativity will take themselves as providing some structure according to which certain actions are ruled out – we wouldn't be talking about them otherwise. Many ethicists will demand a stronger view, according to which significant and *specific* actions are regulated. Nevertheless a weak condition is all that's required here – any ethicist that demands a stronger view will of necessity accept this weaker condition, and more to the point it is my view that more theories fail to meet this constraint than is appreciated, or at least fail to do so while still respecting the authority constraint. In particular I will ultimately argue that *Attributivism* falls short as regards **Regulation** when taken in conjunction with **Authority**. So although the Act Regulation condition may seem obvious, I argue for it here explicitly so that we might appreciate both how Act Regulation can be achieved, and how it might fail to be.

§4.5.1 argues that the existence of normativity is compatible with a world in which *everything comes out right*, or a normatively trivial world. §4.5.2 and §4.5.3 assess potential methods of regulation, with §4.5.2 assessing a deontic or overall ranking of actions whereby normative property *x* is either instantiated or not, and §4.5.3 assessing a multi-factor ranking by which *x* may be more or less instantiated. §4.5.4 concludes.

§4.5.1 – Worlds without Regulation

The need for a property which satisfies **Regulation** is best illustrated through imagining a world in which there exists a property which is authoritative but not regulatory. Consider the following.

God's Country is a world with exactly one normatively authoritative property: *Being Loved by God.* This is, in God's Country, a real property – God exists, it has love to give, and said love is authoritative. God's love is unavoidably motivating – ideal agents will always be motivated to do *only* things that God loves. Thus the standard 'do only that which God loves'

seems to be a normative one. God's love is indeed a property which makes it the case that we ought to take certain types of action.

God's Country however has yet another interesting feature. The God of God's Country is a liberal and magnanimous Being; God loves everything in God's Country equally, such that there is *no* action that fails to be loved by God to the maximum extent. The result is that whatsoever an agent in God's Country chooses to do, they've done the right thing; in God's Country, *everything comes out right*.

Now this isn't to say that God's love isn't normative, exactly. God's Love *is* legitimately authoritative, and that authority supports the claim that agents *really ought* to do only that which God loves. If there *were* actions God didn't love, they *would* certainly be wrong to do them. No, the problem isn't that God's love isn't normative, but that it fails to regulate action, to rule anything out, and so God's Favour doesn't entail any sort of normativity *worth talking about*. If everything will come out right regardless, there's no reason to care about normativity. But here in *our* world, outside of God's Country, we *do* care about normativity, or think we should. So if we're right to do so, there had better be some capacity on the part of any candidate normative property to regulate amongst our *actual available* options for action.

The problem, in short, is straightforward, and two-fold. (i) Non-regulatory normativity is trivial. (ii) Non-regulatory normativity fails to do that which makes us think it's important in the first place. Either way, a non-regulatory normativity *isn't worth talking about*.[127]

[127] I take myself to be preaching intuitive orthodoxy here, that a normative property which doesn't *carve nature at the joints* and cause actions to be delineated as appropriate or inappropriate, good or bad, ruled in or ruled out, etc., is in some sense *no good*. If that conclusion isn't as obvious as I think however, it's worth noting that many philosophers seem to agree. Judith Thomson, for example, seems to define normativity in terms of just such a division. She writes, "For it to be the case that A ought to V is for it to be the case that there is a directive kind K such that: (α) A is a K, and (β) if a K doesn't V than it is a defective K" (Thomson 2008, 212). That is to say, it's only because failing to V is ruled out by some normative standard that we ought to V – if there were no *defect*, no ruling out, there would be no *ought* at all. Philosophers like Jackson and Smith in 'The Implementation Problem for Deontology' (Jackson & Smith 2016), or Scott Woodcock with his *indeterminacy objection* to neo-Aristotelianism (Woodcock 2015) make similar points. Although they are among the ranks of philosophers who argue that an adequacy condition on theories of normativity is that they provide clear methods for arriving at

§4.5.2 – Methods of Regulation: Incompatible Actions

I have argued for **Regulation**, i.e., that the ability to classify actions as ruled in or ruled out is required of any fundamentally normative property. It's one thing however to know what's required, and quite another to know how to achieve that requirement. As one aim of this thesis is ultimately to judge whether or not Attributive Goodness *does* achieve that requirement, we'd be best off with a clear picture in mind of exactly what doing and *failing* to do so would look like. §4.5.2 and §4.5.3 will propose two different ways a normative property might *rule out* or otherwise regulate actions.

There are at least two different ways of approaching act regulation, discounting hybrid accounts that sample from both. One way is to posit a normative property the standards of which are satisfied in a strict 'on/off' fashion. On such an account, any two acts which satisfy the standards derived from the property will do so to an equal degree. Imagine, for example, we take 'maximising happiness' to be a normative property – an act ought to be done when it has the property of maximising happiness. Since maximisation allows for no sense of degree, if an action maximises happiness it will do so either completely, or not at all. Actions thereby will be either equivalently ruled in – in virtue of maximising happiness – or equivalently ruled out in virtue of not doing so. Although what counts as maximising happiness may vary from situation to situation, in any particular situation an act either strictly will do so, or strictly won't. In short, some acts 'have it' – where 'it' is whatever it takes to count as promoting the normative property – and some don't.

Deontic theories fit most easily into that sort of structure. Deontological thinking sees actions sorted into two categories, those that uphold our duties (or are compatible with doing so), and those that don't. As Jackson and Smith put it, "deontology traffics in "on-off"

conclusions about what to do, their arguments presuppose the idea that there *are* right and wrong ways of doing things – it is precisely because there *are* actions that must be avoided that a clear rule for avoiding them in required. Thus I take it as implicit in any theory which argues that normativity must be action *guiding*, that normativity must also be metaphysically delineating in the regulatory way I have argued for.

properties of actual and possible actions, properties that are either possessed or not possessed: an action is or is not the breaking of a promise, the telling of a lie, the punishing of someone innocent, etc." (Jackson & Smith 2016, 290). This sort of regulation appears regularly even in cases of more mundane, instrumental normativity. Familiar by now, the rules of etiquette or games like chess regulate action in this manner: either one has used the right spoon for soup or one hasn't, either one has made an acceptable move by the rules of the game, or one hasn't. There is no 'almost', you get no credit for *nearly* picking the right spoon, or for moving the bishop in *mostly* the right direction. Either your action follows the rules of the game, or it doesn't.[128]

Prima facie the sort of regulation outlined above seems not only plausible, but paradigmatic. Most ethical theories at least *seem* to involve a sharp divide *somewhere*. Perhaps there are many degrees of goodness, for example, but presumably there will be *some* point wherein some act is just no good at all. Nevertheless the implications of that 'have vs. have-not' approach often seem implausible or impractical. Strict and unyielding deontological theories lead us to intuitive quagmires – do we fail in our duty not to lie when questioned by murderers about our friend's whereabouts, or do we fail in our duty to protect our friend? We're left to deal with a nasty sort of indeterminacy when choosing between independent and mutually incompatible duties.[129] What's more, while most ethical theories will seem to employ some sort of hard division between the forbidden and the permissible, few will fail to also

[128] Perhaps games are a more complicated case than I'm giving them credit for. In some cases, after all, it seems that the *right* move is the one that *most* accomplishes the goal of the game, whereas others that don't are technically permissible but normatively *wrong* – it may not be forbidden for me to move my bishop, but if I could checkmate my opponent by moving my knight, moving my bishop would be the *wrong* move. Nevertheless, when a move *is* definitively ruled out in these systems, it is ruled out in virtue of straight-forwardly *not instantiating* the normative property, not in virtue of not instantiating it *enough*, or being outweighed by the capacity of some alternative move to do so.

[129] Shyam Nair discusses a similar picture in 'How Do Reasons Accrue', regarding what we're left with if we accept a Rossian view of brute independent duties that leaves us no method for arbitrating between them. Similarly Lord and Maguire discuss the 'deep normative dilemmas' strict, non-weighted theories of the normative face, in 'An Opinionated Guide to the Weight of Reasons'. See Nair 2016, 70-71, and Lord and Maguire 2016, 6.

employ some sort of weighing or ranking system to further regulate action within the permissible set. Although a deontological system which allows for some variety of *weight* amongst duties might go a long way towards solving the indeterminacy problem and satisfying weighting intuitions, it's worth exploring weight itself as an alternative method of regulation.

§4.5.3 – Methods of Regulation: Weight, Accrual and Hybrid Theories

What follows from the previous section is that whenever something is ruled out, it is ruled out to the same degree of strength. That can seem implausible. It is, after all, difficult to swallow that my promise to meet you for tea should, without qualification, exert normative pressure equal to my duty to avoid slaughtering innocent people. Likewise it is difficult to believe that someone who fails to donate to the poor is just as in-the-wrong as somehow who routinely slaughters the poor. It is thus increasingly recognized that theories of normativity require some sort of machinery for weighing individual obligations, in order at least to allow for obligations of varying importance. It's also increasingly thought that strict divisions between the obligatory and the forbidden can be explained in terms of the weight and/or accumulation of token normative properties.[130] So it's reasonable to focus on this sort of weighting capacity of normative properties as representing an alternative way of regulating actions.

The overall idea is simple enough, and at least equally as intuitive as its binary counterpart. Even if we accept that there is a strict divide between the sort of things we ought and ought not to do, nothing seems more natural than that *among* those things that we ought to do, some are more important than others. Action, it seems, can be regulated even amongst those acts that genuinely satisfy normative standards, favouring those actions which do so *more*, or those with greater weight.

[130] Lord and Maguire argue as much: "any decent theory needs to have at least one *weighted* notion in its normative toolkit", and "It is common to think that the non-strict notions will explain the strict. It is also increasingly common to think that the weighted notions will explain the non-weighted." (Lord and Maguire 2016, 3-4).

The operations of reasons as a basic normative unit offer a paradigmatic example of this sort of regulation. We have, of course, reasons to do and to not do certain things. But it's relatively uncontroversial that we have *more* reason to do some things than others, or that some reasons are weightier or stronger or more pressure-exerting than others. So if we take our normative obligations to be determined by reasons, we take our actions to be regulated less through being sorted into binary categories of forbidden or required, but by being arranged into a hierarchical structure where some actions are *more* justified than others. Even if there is *some* reason to do absolutely *everything*, we can rule certain actions out if the reasons supporting them are *less weighty* than those supporting other actions.

Of course, actually explicating the nature of normative weight is tricky. It seems only recently that serious attempts have been made to account for the mechanics behind the existence or accrual of normative weight. Nor is it clear whether weights and/or weightier-than relations should be taken as basic or reducible. W. D. Ross in 1930's *The Right and the Good* advocated for a complicated deontological theory involving several different duties of distinct, intrinsic weights, all of which can be satisfied to degrees. One's overall duty in any given situation on this theory is to take the action which satisfies the individual duties to the weightiest degree, while there is nothing further to be said about *why* the individual duties have the weights that they do.[131] So Ross, and one would assume any other non-reductive realist, takes weight to be a brute feature of normativity. Alternatively we might understand normative weight in a reductive sense, where the weight of a particular obligation will depend upon how the particular elements from which normativity derives combine and accrue. If what we are to

[131] Specifically, Ross advocated for the duties of fidelity, reparation, gratitude, justice, beneficence, self-improvement and non-maleficence. Ross, an intuitionist, provides no explicit method for arriving at a conclusion as to what action will 'best' satisfy the demands of the combined duties, but we might imagine that if, say, non-maleficence intrinsically outweighs fidelity, and if one action satisfies non-maleficence to a high degree and another fidelity to an equal degree, non-maleficence should be prioritised. However, if one action satisfies non-maleficence to a low degree and another fidelity to a high degree, satisfying fidelity may be called for. The lack of an explanation as to the accrual of overall weight, however, is a problem for the theory. See Ross 1930.

do is determined by, say, how much pleasure an act brings about, then the weight of an obligation will depend upon how pleasure is measured, weighed and combined.

I mention the above only to further illustrate the wide range of approaches to regulation via weight and accrual. It is not important here to resolve the question of whether or not normative weight is brute. Nevertheless, what this discussion has made clear is both that weight and accrual of normative properties is a plausible and indeed intuitive method for regulating action, and also the ways in which it differs as a method from the binary approach of the previous section. Where a binary normativity allows only for actions to be completely acceptable or completely unacceptable, a normative property which admits of degrees of instantiation allows for a wider range of evaluations, for actions to be more or less right or wrong. Where a binary method of regulation paints the world in black and white, seeing actions as ruled strictly in or out, a method that admits of degrees sees shades of grey, ruling out actions in virtue of them being *less* ruled *in*. Finally, where a binary method excels in simplicity it suffers in implausibility; where a method of degrees excels in intuitive plausibility, it suffers in obscureness of method.

Of course it's entirely likely that many theories of the normative will turn out to be hybrids of the two regulatory methods. In fact that seems to be the case more often than not. Almost every theory imposes *some* strict cut-off point, even while they allow for degrees among the otherwise permissible. For example, any consequentialist theory that does *not* focus specifically on the *maximisation* of some property will be likely to strictly rule out anything that actively *impedes* the promotion of their toy property, yet allow for evaluation by degrees amongst those actions that *do* promote it.[132] Likewise virtue theories will impose a strict cut off on actions which blatantly impede the promotion of the virtues, yet will evaluate appropriate

[132] Such views are called *scalar* views. They have been defended most recently by Alastair Norcross. See for example 'The Scalar Approach to Utilitarianism' (Norcross 2006).

actions as *more* or *less* virtuous. We should expect to see some elements of both methods of regulation in any plausible theory of the normative – nevertheless it is worthwhile to have an appreciation of the various methods that are and may be employed.

§4.5.4 – Conclusion

This section has made the argument for the second of the minimal two hallmarks of normativity. Call it,

> **Regulation.** For any property *x*, *x* is fundamentally normative *only if x* provides a standard of action that some, but not all, possible actions fail to satisfy.

I have not here considered any particular *objections* to this hallmark – at least partially because I take myself here to be preaching orthodoxy, and making explicit something that is inherent to the very idea of normativity, something that nobody disagrees with. It's a tragic truth however that even the best of us are prone to discover, at times, that we've been advocating something that has run afoul of beliefs we hold, even firmly, and only come to appreciate or realize this too late. Making the need for regulation explicit will become shortly very important, when I turn to how Attributivism fails to meet the demands of normative status.

§4.6 – THE HALLMARKS OF NORMATIVITY, FIN.

Across this chapter I have investigated the necessary and sufficient conditions of normativity. I reviewed several influential theories and found them wanting. Across §4.1-§4.3 I rejected Reactive Attitude accounts, Naïve Instrumentalism, and Reasons-First theories of normativity as offering either derivative normativity or incoherent normativity. Reactive Attitudes either rely on the normative appropriateness of certain attitudes, or are patently implausible. Naïve Instrumentalism ends up with a normativity that condones any action taken. Reasons-First theories require some further method of distinguishing between *normative reasons* and *non-normative reasons*, that can't rely upon reasons-relations themselves.

Across §4.4 and §4.5 I argued for two *necessary* conditions on normativity. §4.4 presented **Authority** – a capacity to ensure that genuinely normative standards are inescapably motivating – while §4.5 argued for **Regulation** – a capacity on behalf of a normative property to rule in and out certain actions.

A few closing notes before moving on. First, in **Authority** and **Regulation** I have offered *necessary* conditions. A property is normative *only if* it possesses those features. I make no claim to those being *sufficient* conditions – although I expect they are – but only that a property which *doesn't* possess them is not a normative one.

Second, following from the above, it should be noted that Authority and Regulation do not ensure *fundamental* normativity, despite that being the ultimate focus. Some property may be – circumstantially or in specific conditions – *derivatively* Authoritative and Regulatory. It may, for example, be genuinely Authoritative that I complete this thesis, without completing theses being fundamentally normative. Perhaps it goes without saying, but I will say it anyway: any property which possesses Authority and Regulation, yet can be shown to be dependent for that possession upon some further normative property, is not itself *fundamentally normative*. A fundamentally normative property must be *non-derivatively* authoritative and regulatory.

Authority and regulation must be able to be derived from the truth of the relevant facts *alone*, i.e., *if* the relevant facts about the property obtain, *then* authority and regulation follow.

That concludes my foray into pure normativity. I return now to Attributive Goodness.

5 – Attributive Goodness and Fundamental Normativity

The topic of this dissertation is whether or not attributive goodness is a fundamentally normative property. Thus we return to Attributivism, in order to evaluate whether it can accommodate the requirements of normativity now established. In chapters 6 and 7 I will argue that Attributivism *fails* to do so. However, to have a clear understanding of where a theory goes wrong, one needs to understand how the theory proposes to work in the first place. What the last few chapters have done is *lay out* all the pieces – we have spread before us the components of Attributivism, and the schematic of normativity. *This* chapter will put those pieces together, to explain how Attributivism proposes to meet the requirements of normativity, so that we might better understand when and how it all goes wrong.

§5.1, *Attributivism*, reminds us of Attributivism's core commitments. §5.2, *Attributive Goodness and Authority*, lays out what Attributivism is committed to by **Authority**, and §5.3, *Authoritative Attributivism* explains how attributivists propose to satisfy those requirements. §5.4, *Attributive Goodness and Regulation*, and §5.5, *Regulatory Attributivism*, repeat that process as regards **Regulation**. §5.6 concludes.

§5.1 – ATTRIBUTIVISM

Recall that Attributivism is the name given to any theory which takes attributive goodness to be fundamentally normative. That is, Attributivism takes normativity to be grounded in attributive goodness; that Φing is attributively good for *x* makes it the case that *x* ought to Φ. Attributive Goodness, in turn, is goodness relative to standards derived from the functions and teleological ends of kinds. A good *x* is one that does what an *x* is supposed to do: a good *clock* is one that *keeps time,* a good knife is one that cuts well, and a good human is one that does whatever it is humans categorically do.

In summary, this is the claim of Attributivism. For any agent *x*, there exists a kind *K* such that (i) *x* belongs to *K* and (ii) *x* ought to do only what is good *qua K*. That is the core component of Attributivism. Beyond that, Attributivism has three key commitments. Those are commitments to cognitivism, realism, and naturalism. I'll explain below.

Attributivism is a cognitivist theory. When we express a judgment about attributive goodness, we are taken to be expressing a proposition, describing facts about the world in a truth-apt way. If we evaluate something as attributively good or bad, according to Attributivism we aren't merely expressing an attitude towards that thing; rather we are asserting the existence of certain attitude-independent states of affairs. We mean, when we say that *x* is attributively good, that *x* is behaving or existing in a certain way, and that those facts are independent of any subjective mental states. Likewise when we say that *x ought* to be attributively good, we are asserting that certain features of the world make that the case, as opposed to expressing a preference or attitude. That *x* ought to be attributively good isn't taken to depend upon any subject mental states or attitudes.[133]

[133] A note should be made here that although Attributivism's claims aren't taken to *depend* upon mental states, they are intimately tied up with mental states. As subsequent sections will demonstrate, Attributivism relies upon there being an intimate connection between facts about one's kind and facts about one's desires: it is because certain desires *follow* from facts about one's kind that attributive goodness is plausibly authoritative. It remains true however that Attributivism's claims don't aren't grounded in any mental states – that *x* ought to be attributively good doesn't *depend* upon any specific mental states, but it *is* taken to *entail* certain mental states.

Attributivism is also a realist theory. As with the above, when we judge that *x* ought to be attributively good, we are making a proposition judgment, a judgment about the way the world *is*. More than that however, Attributivism claims that we are making a judgment that is, at least sometimes, actually *true*.

Finally, Attributivism is a naturalist theory. Attributivism analyses normative facts in terms of attributive goodness. Attributive goodness, meanwhile, is a natural property – it is instantiated when certain facts about the natural functions of kinds and the behaviours of members of those kinds are the case. Attributivism is not taken to entail the existence of any non-natural or irreducibly normative properties.

Some examples will help to summarise. Take it for granted that a human agent ought to keep promises. Attributivism claims that what explains why a human ought to keep promises is that doing so is attributively good for humans. What is attributively good for humans, meanwhile, or for any member of any kind, is that which accomplishes the natural, teleological ends of that kind. Keeping promises, thusly, is understood as contributing to the satisfaction of *human* teleology, or as playing a functional role in the maintenance of the human organism. That is precisely why a human ought to do so. More broadly speaking, as Allyn Fives summarises, Attributivism claims that "what a person should do all things considered is that which he or she should do because human good hangs on it" (Fives 2008, 172). Human good, meanwhile, is natural, objective, and real – it is derived from and understood in terms of actually existing, natural, teleological ends and functional dispositions belonging to members of the human kind. I won't retread the details of how Attributivism identifies the ends of organisms, or how it expects to solve traditional moral problems – chapters 2 and 3 have covered that at length. The focus here and now is to see how the components of the theory just outlined are supposed to fit into the normative schematic provided in the previous chapter.

§5.2 – ATTRIBUTIVE GOODNESS AND AUTHORITY

In the previous chapter I presented two adequacy conditions a property must meet for it to be a fundamentally normative property – i.e., one which, in its distribution over agents and entities alone, grounds what agents ought to do. Those adequacy conditions are **Authority** and **Regulation**. Over the next several sections, I will explain what those conditions mean for Attributivism.

As a reminder,

> **Authority.** For any property *x*, *x* is fundamentally normative *iff* an idealised agent would derive overriding motivations to act from facts about *x*.

The property we are interested in, insofar as we are interested in Attributivism, is attributive goodness, the property realised by a subject when it achieves the natural, teleological ends of its kind. Thus, what **Authority** requires of Attributivism is that ideal agents derive overriding motivations from facts about attributive goodness. Not just *any* motivations however. Attributivism makes a specific claim: agents ought to do only what is good *qua* their kind. Thus, for Attributivism to be true it must be the case specifically that ideal agents do/would derive overriding motivations from facts about attributive goodness to act only in ways that are good *qua* their kind.

Moreover, since attributive goodness must be non-derivatively authoritative to be fundamentally normative, it must be such that the truth of the relevant facts alone is enough to guarantee authority. Thus Attributivism must argue that *whenever* one is a member of a kind, one's ideal self would be overridingly motivated to act only in ways that are good *qua* that kind.

For sake of ease we can put the above in terms of a single case example, the *human* case. If Attributivism is true, and **Authority** is true, as I've argued, then what must be the case is that human ideal agents would be overridingly motivated only to do things that are good *qua* human. If, therefore, it is bad *qua* human to bring about the death of an innocent person, then ideal human agents would be overridingly motivated not to do so. If Φing is an action that would result in agent *x* bringing about the death of an innocent person, then *x*'s ideal self would not Φ. I take it for the sake of charity that this extends also to acts which would *contribute* to bringing such an action about. If, correspondingly, ideal human agent *y* knew that Φing would contribute to bringing about the death of an innocent, or aid another agent in doing so, then *y* would likewise be overridingly motivated not to Φ. In short, if Φ is bad *qua* human, then the ideal human won't do things that bring about Φ. That must be the claim of Attributivism.

To evaluate Attributivism then is to evaluate it as regards the following claim. Any ideal agent will be overridingly motivated to act only in ways that are good *qua* their kind.

What the attributivist needs to do now is twofold. First they need to provide a picture of how they take attributive goodness to link up with motivational states in the requisite way. Second, they need to provide an argument as to why their picture is a *plausible* one – why we should *believe* that attributive goodness and motivation link up in the way attributivists claim. They're relatively prepared to do so, as I'll now explain.

§5.3 – AUTHORITATIVE ATTRIBUTIVISM

In this section I'll argue that the method Attributivism does and must employ to explain the authority of attributive goodness is to argue that agents, ideal or otherwise, can't help but be motivated towards attributive goodness, due to facts about the phenomenon of desiring. In brief, they argue that desire *follows* judgments about attributive goodness, in a way that makes motivation and motivation towards attributive goodness in some way inseparable. Specifically, they argue for a substantive or constitutive aim to desire itself, and thus that *all* desires are desires of a specific form, constitutionally aimed at attributive goodness.

Quotes from early attributivists like Anscombe and Geach hint at the above tactic. Anscombe writes,

> There is some sort of necessary connection between what you think you need, and what you want … it is possible *not* to want something you judge you need … but it is not possible never to want *anything* you judge you need. This [is a fact] about the phenomenon of *wanting* (Anscombe 2005c, 178).

Likewise, according to Geach,

> other things being equal, a man who wants an *A* will choose an *A* that he thinks good … It can [thus] be shown that an action's being a good or bad human action is itself something that touches the agent's desires … what a man cannot fail to be choosing is his manner of acting; so to call a manner of acting good or bad cannot but serve to guide action. As Aristotle says, acting well … is a man's aim simpliciter … and qua man (Geach 1956, 38-40).

Anscombe makes a claim that the 'phenomenon of wanting' is, in itself, tied up with receptivity to judgments about needs. Judgments about needs, *ceteris paribus* result in motivation. That constitutes something like an explanation of how Attributivism accommodates authority, if we

attribute to her two underlying assumptions. 1. Needs are *Aristotelian necessities*, i.e., what one needs is properly cached out in terms of what one needs *qua* one's kind. 2. **Ideal Agent** makes true judgments, such that ideal agents will judge that they need what they need *qua* their kind, and thus, typically, want at least something that they need *qua* their kind. I think those assumptions are accurately assumed on her behalf, and so that Anscombe is arguing that ideal agents will, as a general rule, be motivated by judgments about what they need *qua* their kind.

Geach meanwhile argues that an agent choosing among *x*'s, *qua x*, will be motivated to choose an *x* he thinks is good *qua x*. If we assume again on his behalf that **Ideal Agent** makes true judgments, and so will judge rightly what is good *qua x*, his argument entails that ideal agents will be motivated to choose an *x* that is good *qua x* when motivated to choose among *x*'s, *qua x*. He also argues, crucially, that agents, ideal or otherwise, are *always* in the business of choosing actions *qua* their kind. That is, *human* agents are necessarily motivated to choose among *human* actions.[134] Thus they are always motivated to choose that which is good *qua* human, and this is a fact about motivation – a man's aims - *simpliciter*.

Both philosophers argue in their own way that motivation is necessarily and in itself responsive, to at least some degree, to facts about one's attributive goodness. Even if Geach and Anscombe are correct however, authority is not yet delivered. Geach and Anscombe both argue that human goodness stands to have *some* necessary influence over our motivations. That doesn't entail that it will have *decisive* influence over them. What is required is not only an explanation of why our ideal selves will be necessarily *somewhat* motivated towards attributive

[134] The truth of that might not be apparent in Geach's quote himself. He says, after all, that what a human must be motivated to choose is good actions – he doesn't specify they must be choosing good *human actions*. That detail though is apparent when we recall that he argues there is no such thing a good action, simpliciter. "[A]lthough we can speak of a good or bad *human* act, we cannot sensibly speak of a good or bad event, a good or bad thing to happen. 'Event', like 'thing', is too empty a word to convey … a standard of goodness' (emphasis mine) (Geach 1956, 41). Acts, like events, can only be evaluated in a context. In choosing how to act, a human agent must be choosing, to at least some extent, how to act *as a human*, i.e., choosing among *human actions*.

goodness, such that they might still be *more* moved by other considerations. We need an explanation of why our ideal selves would be *overridingly* motivated by attributive goodness.

Thomson provides the elements of such an explanation. In her defences of her own attributivist project, Thomson argues for a relationship of origination and dependence between desires and the perception of attributive goodness. She writes,

> people don't come to think a thing good because they discover that, lo and behold, they have come to take an interest in it; they come to take an interest in the thing because they come to think it good (Thomson 2011, 474).

Thomson's assertion here is that judgments about attributive goodness are those out from which motivations like desire arise. Presumably she means judgments about one's *own* attributive goodness. After all we may judge many things to be good *qua* their own kinds without coming to 'take an interest' in them. Rather, Thomson's view is that when we *do* take an interest in something, we do so typically because of how we judge it to be good in some way *for us*. She writes that when we take an interest in something, we do so "because we … believe it is good for us (thus in our objective interest) or because we … believe we would enjoy it" (Ibidem). Whether it's smoking, attending theatre, eating well, eating poorly, or fostering relationships with friends and family, we don't come to desire these things arbitrarily, but rather under and *because of* the understanding that they are good things *for us* to do.

The case that Thomson seems to be making, and which should be taken as the attributivist position overall, is the following. The nature and/or origin of desire is such that desires are *functionally oriented around*, or otherwise *dependent for their satisfaction upon*, promoting one's attributive goodness.

That is a crucial move in explaining how attributive goodness is supposed to be authoritative. In essence, what the attributivist does, and must do, is assert the existence of a

substantive aim of desire – to promote attributive goodness – in the same way that one might think corresponding to truth is the substantive aim of belief.[135] Arguably, the functional aim of belief is to generate *true* beliefs. Likewise, argues Attributivism, the functional aim of desire is to bring about attributively *good* things. The result is the following thesis which I take Attributivism to be committed to.

> **Substantive Aim Thesis.** Every desire for *x* is a desire for *x as* good *qua* one's kind. To desire *x* is to aim at realising *x* in a way that is attributively good by the standards of one's kind.

In **Substantive Aim Thesis**, we are given the crux of the attributivist argument for authority. Attributive Goodness is an authoritative normative property because to act in a way which fails to promote attributive goodness, or is detrimental to its instantiation, is to act in a confused and self-defeating manner. Acting in a way that fails to promote attributive goodness is acting in a way that fails to achieve, or actively hampers, the aim of the desire in virtue of which the action was taken. So it is easy to imagine why, as required by **Authority**, an **Ideal Agent** would be necessarily motivated to promote attributive goodness. As attributive goodness represents a sort of constitutive aim of desire, to act in a way that impedes the instantiation of attributive goodness is to act in a way that frustrates one's own goals. If we grant **Substantive Aim Thesis**, acting in a way that fails to cohere with attributive goodness is to act in a way that is overall practically incoherent. No **Ideal Agent** therefore could be motivated to take such an action.

[135] Attributivism thus seems to be in opposition to philosophers like Darwall, who has argued in 'Moore, Normativity and Intrinsic Value' (2003) and *The Second-Person Stainpoint* (2006) against there being a general, substantive aim to desire.

The attributivist now has an argument for why attributive goodness is authoritative. Is it a plausible one? On the face of it, yes. Attributivism's prospects of course depend almost entirely on whether or not desire actually has the *specific* aim ascribed to it above. However, attributivists have plausible reasons to expect that there should be at least *some* such aim. The attributivist view in general is that the dispositions and behaviours of individuals are shaped and given functional aims by their place within the reciprocally dependent systems of organic life.[136] Their view of the 'phenomenon of wanting', of desire and of psychological motivation in general, is that those processes don't arise out of nowhere, but emerge and are shaped by their role within those systems, just as with any other disposition. Psychological motivational states like desire are evolved processes like any other. They arise from and are determined by the same pressures which shape other mental and physical dispositions, and so it would be strange for them to be entirely unconnected to their place within a functioning biological system. If it's a sensible question to ask, as Philippa Foot does, what gives agents their goals, on the face of it the answer might as plausibly be 'their needs' as anything else.[137]

The plausibility of the specific aim attributivists ascribe to desire is less obvious. We might approach that question by asking whether the world appears as we would expect it to if their claims were true. In their favour, it at least sometimes seems to.

Intuitively most of us would agree that a starving, destitute, homeless and wounded man is one who is poorly off in a normatively meaningful way – he ought not to be this way, or accept his state, and *ceteris paribus* we ought to see his lot improved. We have a hard time judging these to be anything but bad; sure enough attributivism, recognising these as defective in our particular lifeform, corresponds to our findings. Similarly most of us will see a man who is friendless, isolated, notably cruel or exceedingly vain as exhibiting normatively poor

[136] See again earlier sections 3.1 and 3.2 for those arguments.
[137] See Foot 2001, 22.

behaviour, and once again attributivism will support our judgments in finding these defective in an essentially social organism. These are simple intuitions though, and fairly easily accommodated – what might be more convincing is if it should turn out typical that a vicious yet *successful* man rests nevertheless uneasy, or if the impoverished yet *virtuous* are typically more content.

Foot makes much of those latter sort of cases in *Natural Goodness*. She calls on Wittgenstein, who by all accounts led a troubled and often unhappy life, and yet who in the end found little room for regret, presumably – as Foot takes it – for having been dedicated to virtuous pursuits.[138] Likewise she references a prominent Nazi commander – Gustav Wagner – who for all accounts revelled in his sadistic treatment of holocaust victims and showed no overt remorse in later life, yet was plagued by suicidal tendencies and later took his own life. She speaks fondly of a group of prisoners about to be executed for opposing Nazi pogroms, whose letters prior to death she claims radiate "an extraordinary sense of happiness", despite the dire predicaments of the letter writers.[139]

The happiness of the suffering yet virtuous, and the misery of the successful yet vicious, Foot chalks up to success or failure in instantiating the attributive goodness of the human kind. Sure enough this is exactly the sort of result we would expect to find if the claims of Attributivism were true – if the *important* and *authoritative* motivations were those grounded in attributive goodness. Foot has, of course, focused on only a few striking cases – it's difficult to know if the sort of result she focuses on is typical, or if Foot is being overly optimistic. Yet if these cases can be believed they *do* seem to lend credence to attributivist claims, and in the absence of a wealth of evidence for either side, attributivist claims to authority at least don't

[138] "One recalls Wittgenstein's famous death-bed insistence that he had had a wonderful life" (Foot 2001, 85).
[139] See Foot 2001, 92-96, and *Dying We Live*, Gollwitzer H., Kuhn K., and Schneider R., 1956.

seem *glaringly* implausible. I am content then with taking the argument here outlined as being worth taking seriously.

§5.4 – ATTRIBUTIVE GOODNESS AND REGULATION

As in §5.2, I'll begin by reminding us of what regulation is. This section will provide a brief recap of the regulatory requirement as regards Attributivism. §5.5 will subsequently explain how Attributivism might accommodate it. To begin with,

> **Regulation.** For any property *x*, *x* is fundamentally normative *only if x* provides a standard of action that some, but not all, possible actions fail to satisfy.

Regulation is required for normativity, because **Regulation** is what introduces the idea of obligations we can fall short of, in virtue of introducing *wrong* or *ruled out* ways of acting. Without **Regulation**, all we have so far are facts about what we are and must be motivated towards. **Regulation** is what makes normativity *relevant* in our deliberations.

So what does **Regulation** require of Attributivism? The answer in the abstract is easy. **Regulation** – particularly as compared to **Authority** – seems to be a relatively simple requirement. All attributivists have to do is describe how the property of attributive allows for the generation of a standard according to which certain actions are ruled in or ruled out. That is to say, they need to give an account of how they understand good and bad action in terms of attributive goodness. The following section will explain how they do so.

§5.5 – REGULATORY ATTRIBUTIVISM

In this section I explain how Attributivism accommodates **Regulation**. I do so by first drawing attention to the parallels between Attributivism and general virtue-based ethical theories. I conclude that attributivists characteristically derive regulation from attributive goodness by situating agents on a spectrum. On the one end is the *fully virtuous agent*. On the other end is the *defective* agent. Attributivism explains virtue and defect in terms of attributive goodness: the fully virtuous agent is one who is successful by the standards of her kind, while the defective agent is one who is not. Action is regulated accordingly. Actions which are characteristically taken by the virtuous agent are ruled in, while those taken by the defective agent are ruled out. I'll provide more detailed explanations in the following.

Many attributivist theories model themselves as virtue theories – most prominently the accounts of Foot and Hursthouse. It is a reasonable comparison, and a good place to start. Much like typical virtue theories, Attributivism is *agent-centred* rather than *act-centred*; its primary focus is on standards of *being*, not *doing*, and where it dictates actions it justifies such dictates through the rationalization that *act x* will help realize a good *way of being* for the acting agent.[140] It takes the question 'how should I *be*' as more fundamental than 'what should I do', and treats *evaluative* concepts like goodness or excellence as basic, rather than *deontic* concepts like duty or obligation.[141]

[140] Attributivists like Geach deny that there can be such a thing as a good or bad 'event' or 'thing' – these categories are too broad, they must be supplemented with an account of what *kind* of subjects are concerned, i.e., a good event *for humans*, a good *human* thing to happen (Geach 1956, 19). Similar complaints it seems might be leveraged against 'acts' – before evaluating an act one must understand the context of the act, what *kind* of act it is or what it is for. Of course one might entertain, as has become more common, that action itself may have constitutive standards – that an act may be evaluated simply *qua action*. Even granting such a consideration though, I imagine the attributivist response should be to insist that any given act draws the content of its constitutive standards or aims from the constitutive aims or standards of the agent by or for which the action is taken. See Silverstein 2012 on the need for a reductive account of constitutive standards of action.

[141] Hursthouse provides these candidate 'defining features of virtue theories' in the first chapter of *On Virtue Ethics* – she goes on to argue that most such characterizations of virtue ethics are flawed or misleading, but also that there is at least some truth to them. See Hursthouse 1999, 25.

A common objection to virtue theories has been, historically, that they are incapable of informatively guiding action, not least because they primarily evaluate *agents* and not actions. Rosalind Hursthouse dismisses such objections decisively in the first chapter of *On Virtue Ethics* (Hursthouse 1999). The entirety of the chapter is worth reading, if only for how Hursthouse resolves the illusion that utilitarian or deontological approaches to act-regulation are somehow fundamentally different than those of virtue ethics. I'm content here however to provide only her approach to defining 'right action':

> P.I. An action is right iff it is what a virtuous agent would characteristically … do in the circumstances.

where

> P.Ia. A virtuous agent is one who has, and exercises, certain character traits, namely, the virtues.

and where

> P.2. A virtue is a character trait that … (Hursthouse 1999, 28-29).

It remains to the specific virtue theory in question to finish the above sentence. Within Attributivism, a virtue is a character trait that promotes the satisfaction of the requirements of the standards of an agent's kind.

To paraphrase, the right action in any given circumstance is that which a virtuous agent would do in that circumstance. What a virtuous agent does meanwhile is exhibit the virtues. So the right action in any given circumstance is to exhibit the virtues. It is only for the attributivist here to give a reductive account of what the virtues *are,* namely to claim that 'the virtues' vary in accordance with and relative to the kind of the acting agent, and are identified

for a given agent with the attributive 'goods' of its kind. So Hursthouse's formula comes out as the following: an action is right for a member of a kind iff it is what an attributively good member of that kind would characteristically do. What an attributively good member of a kind characteristically does is satisfy the standards of their kind.

Hursthouse provides a start at understanding Attributivism's method of regulation. She tells us how Attributivism allows for the classification and identification of *right* actions. Attributive goodness is instantiated by actions which satisfy the standards of one's kind, and those actions can be identified by knowing what satisfies those standards, or by knowing what is characteristically done by fully virtuous agents of one's kind. What of *wrong* action though? **Regulation** demands some actions be ruled *out* by the relevant property. Is attributive goodness the sort of property that can rule out action, and if so, what sort of actions?

In what ways might a property rule out an action? Consider how other types of ethical theory come to delineate their prohibited actions. A simple form of act utilitarianism makes the right action in any situation the one which instantiates the maximization of overall happiness. A wrong act, then, is any action which (i) *fails* to maximize happiness and so *impedes* the instantiation of the relevant property or (ii) *decreases* the level of overall happiness and so actively *detracts from* or *harms* said instantiation. Similarly, a deontological theory might understand the right action to be that which, say, instantiates in the agent the property of having done her duty. A wrong act, then, would be any action which (i) *fails* to uphold one's duty, *impeding* the instantiation of the property or (ii) conflicts with one's duty, say lying when one has a duty to tell the truth, and so being *incompatible* with the instantiation of the property.

Both of the theories just given plausibly allow for **Regulation**; both theories offer properties that seem capable of being instantiated, and it's relatively obvious that many genuinely possible actions will serve to impede or conflict with that instantiation. Harmful actions – like causing pain – conflict with maximizing happiness. Non-universalisable actions

– like murder – conflict with the deontological categorical imperative. Attributivism understands right action to be that which instantiates or promotes the instantiation of an agent's attributive goodness. What kind of action then should we expect to be ruled out for *conflicting* with that instantiation?

Hursthouse, above, might be said to have focused upon those actions Attributivism rules *in* – her focus was specifically on the actions of fully virtuous agents and what they would do. It's possible to think of what's ruled *out* in terms of what *isn't* ruled in – what one ought not to do is that which no fully virtuous agent would ever do. Judith Thomson however offers an account of regulation in terms of *defective* agents which may prove more informative.

As already given in previous sections, Thompson argues that "[x's] being marked as [defective] if he does not [φ] is what makes it the case *both* that [φ-ing] deserves our wanting it to occur, and that he ought to [φ]" (Thomson 2010, 759). In *Normativity*, she puts this more formally:

> (Improved Directive Thesis) For it to be the case that A ought to V is for it to be the case that there is a directive kind K such that:
>
> (α) A is a K, and
>
> (ß) if a K doesn't V, then it is a defective K, and
>
> (γ) there is no directive kind K+ such that K is a sub-kind of K+, and such that if a K+ does V, then it is a defective K+. (Thomson 2008, 214).

As an example – Stan ought to break promises only if (α) Stan is a Liar and (ß) if a Liar doesn't break promises it is a defective Liar and (γ) there is no other kind Human such that Liar is a sub-kind of Human, and such that if a Human breaks promises is it a defective Human. Since (γ) obtains, Stan ought not to break promises as he would be defective. Though formulaically focused on what it takes for something to be ruled in, Thomson's formula is illuminating in the

way it characterises the standards of kinds as standards in virtue of which one can be successful or *defective*, and categorises right actions in contrast with *defect*. Actions which lead to defectiveness in agents are the kind of actions ruled out in Attributivism. An action should be understood as *defective* iff (i) the acting agent can be said to have a *function* and (ii) Ving is an act of failing to perform that function and thus impeding satisfaction of the agent's kind standards, or (iii) Ving is an act of performing that function *badly*, thus conflicting with the satisfaction of the agent's kind standards.

Is attributive goodness then a property which plausibly allows for the classification of a set of actions as *wrong*? Certainly. Are there, meanwhile, any possible actions that *do* come out as defective? It seems likely. Seeing-eye dogs, as Thomson supplies, are defective if they fail to guide their owners across the street, or do direct them into on-coming traffic. Seeing-eye dogs, meanwhile, actually exist, and those actions are actions they might very well take. Is there, finally, a plausibly accessible method for determining which actions are actually ruled out? Probably yes: grasping the function of a kind *requires* an understanding of what would constitute a defective member of that kind, so it is as plausible that we can come to understand what actions are ruled out as it is that we can come to understand the function of our kind. Attributivism appears to allow for Regulation as easily as do many other theories of normativity.

It would have been surprising in the extreme had Attributivism *not* come out to be capable of regulating action. It is nevertheless informative to understand how it does so, and what kinds of actions it actually regulates. I am satisfied that Attributivism's regulatory prospects are at least as plausible as those of alternative theories.

§5.6 – The Spectre of Authoritative Regulation

In this chapter I have presented what I think Attributivism is committed to if attributive goodness is to be an authoritative and regulatory property, and thus a fundamentally normative property. I have argued that the characteristic and best way through which Attributivism accounts for **Authority** is by arguing for the **Substantive Aim Thesis**. According to that thesis, psychological motivational states like desire are essentially or functionally oriented around attributive goodness, such that the satisfaction of any given desire depends at least in part on satisfying it in a way that coheres with or furthers one's attributive goodness. I have argued also that Attributivism accommodates **Regulation** by being grounded in standards that can be satisfied or impeded by possible actions, and that those actions can be cached out and understood in terms of virtuous and defective agents and their characteristic actions.

I'm prepared to accept the plausibility of the above arguments on behalf of Attributivism. I'm even willing to believe that they might be correct, and that Attributivism can deliver **Authority** and **Regulation**. What I am not convinced of however is that it can do so *simultaneously*. A fundamentally normative property can't just be *sometimes* authoritative and *sometimes* regulatory, or just one *or* the other. It must be both at once; its specific regulations must be authoritative; it must be *authoritatively regulatory*. I think that is a particularly challenging requirement to meet and, over the course of the next two chapters, I will explain why Attributivism doesn't do so.

6 – ESCAPING ATTRIBUTIVISM

The explicit aim of this thesis is to explain the *failure* of Attributivism. Until now however, my treatment of that theory has been primarily *positive*. This chapter at last turns to criticism. So far I've undertaken two distinct projects. The first has been to develop the strongest version of Attributivism. The second has been to provide a minimal analysis of normativity. Both projects, however disparate, were undertaken with an aim to assessing Attributivism's normative capacity. I can now do so. Hereon I will argue that a critical defect exists in Attributivism. My argument is, roughly, as follows.

Attributivism argues that what is normative for agent x can be explained in terms of what is good for x's kind, such that if x belongs to kind K, and Φ is good *qua* K, then x ought to Φ. I've argued that that view is plausible only if what is good for kind K would overridingly motivate the ideal version of x. As I'll explain in this chapter, the way that Attributivism tries to accommodate that requirement both assumes and requires a sort of kind essentialism, i.e., that agents belong to, and can belong to, only one kind. I will argue that that is not the case. Agents can and do belong to multiple kinds, and can take steps to change their kind. Because of that, it turns out *not* to be the case that x's ideal self must be overridingly motivated by what is good for K. Thus the attributive goodness of x's kind – what is good for K – isn't authoritative, and so isn't fundamentally normative. So Attributivism, which claims that the attributive goodness of one's kind *is* fundamentally normative, fails.

§6.1, *The Guise of the (Attributive) Good*, revisits the **Substantive-Aim Thesis** of the previous chapter. I argue that the best way of explaining a substantive aim of desire is by

taking *all* desires to be conditional desires. I argue too that this makes a significant difference for the way we approach objections to Attributivism.

§6.2, *An Abundance of Kinds*, argues that the result of §6.1 is that Φing is only ever authoritatively ruled out when it isn't good for *any* kind the acting agent instantiates, or can instantiate. I argue that this is a problem, as kinds are possessed much more abundantly and much more fluidly than attributivists have thus far acknowledged – potentially such as to allow an agent to invalidate the authority of any particular kind.

§6.3, *Kind Manipulation*, develops a few scenarios to demonstrate the problem outlined in §6.2, exploring the difficulty posed by both multiple kind instantiation and the possibility of kind-manipulation or outright replacement.

§6.4, *Many Kinds of Objection*, entertains and rejects several attempts at rescuing the Authority of *specific* kinds. §6.5 concludes.

§6.1 – THE GUISE OF THE (ATTRIBUTIVE) GOOD

Before I explain the objection against Attributivism, something more needs to be said about how that theory tries to satisfy **Authority**. Once again, Attributivism argues that an agent ought to do only what is good *qua* their kind, such that if agent *x* belongs to kind *K*, *x* ought to do only what is good *qua K*. The question is, how can attributivists plausibly argue that, for example, ideal human agents would be overridingly motivated to take only actions that are good *qua* human? In this section, I'll explain the best answer to that question. I'll argue specifically that attributivists should claim that all desires are *conditional* desires, i.e., desires that are able to be satisfied only on the condition that doing so will be good for the agent's kind. Thus the only desires an ideal human agent would find to be satiable, and so worth acting upon, would be those that promoted human goodness.

Of course I've already given a partial answer to the above question in §5.3. There I argued that Attributivism is committed to the **Substantive Aim Thesis –** the idea that all desires are aimed at being satisfied in a way that is good *qua* the agent's kind. Still I think more needs to be said. It will be important to know the exact form of those desires, so as to know *exactly* what's involved in satisfying the substantive aim of desire. For example, do the desires of ideal human agents somehow aim *specifically* at *human* good, such that, *necessarily* they can only be satisfied if doing so is good *qua* human? Or do they aim generally at the good of the agent's kind, which just happens to be human, such that the satisfaction of the desire only *contingently* depends upon human goodness? Taking all desires to be conditional desires will, as it turns out, make the desires of ideal agents the latter sort, contingently dependent upon the good of the agent's kind, whatever that might be. That detail will be important in what comes after. For now though, I'll explain why attributivists should argue that all desires are conditional desires, what those are, and how they seem to deliver the results they need.

It is evidently false that agents have an *explicit* overriding desire to do only what is good for their kind. It isn't the case, for example, that among my motivational set is a desire with the specific, explicit aim of doing what is good *qua* human, whatever that might be, and which my ideal self would find to be overriding. It's entirely possible that I might go my entire life without ever thinking to myself, "I'd really like to do something that's good *qua* human right now." It's likewise possible that I might have no specific desire at all which my ideal self would find to be consistently overridingly motivating. Attributivism requires though that even in the absence of any such explicit, specific desires, my motivational set will be such that my ideal self would be motivated by human goodness. Evidently then, what Attributivism requires is that my more specific desires generally, whatever *those* may be, depend for their satisfaction on also promoting human goodness. Fortunately, there exists a theory of desire that might accommodate that requirement. That theory posits a sort of desire that is dependent for its satisfaction upon additional implicit conditions also being satisfied. They are called *conditional desires*. The best hope for Attributivism then is to argue that all desires are conditional desires.

Classically desires are separated into two types: unconditional and conditional desires.[142] An *unconditional* desire for *x* is a desire for *x* on its own merits alone. If my desire for *x* is unconditional, then just as long as I think *x* stands to be satisfied, I will be motivated to satisfy *x*. If, for example, I desire unconditionally that my descendants live happy lives, then I'll be motivated by that prospect, regardless of other conditions – whether or not, for example, I'll be alive to see it, or whether or not their lives will be good for other people, etc. If my desire for my descendants to be happy is unconditional, then, in the absence of a stronger desire, nothing more you can tell me about the conditions under which that will be the case will influence my desire, or cause me to cease being motivated by it.

[142] See, for example, the distinction made by Williams in 'The Makropulos Case' (Williams 1973).

A *conditional* desire, by contrast, is one which *is* receptive to further facts about the conditions in which the explicit aim will be realised. If I have a conditional desire for *x*, then I'll only be motivated by the prospect of *x* conditionally, given my understanding, implicit or otherwise, that some other condition will continue to be the case. Paradigmatic examples are what Stephen Schiffer calls 'bodily appetites' (Schiffer 1976, 202). When, for example, I desire a drink, I desire it only on the condition that, when I *get* that drink, I'll still be thirsty. If you told me that by the time my drink got to me I wouldn't be thirsty any longer, that would cause me to cease being motivated by that desire now. Likewise if I *did* receive that drink later, once I'd ceased to be thirsty, I wouldn't find that it satisfied any desire of mine.[143] Likewise, I might have a conditional desire to go to the pub only on the condition that the pub is *quiet*. If you told me the pub would *not* be quiet, I'd cease to be motivated by the desire to go to the pub. If I went to the pub anyway, I'd find that doing so didn't satisfy my desire. I might not even realise *why*, might never realise that my desire had been conditional in the above way, but nevertheless, its satisfaction relied upon further conditions also being the case.[144]

The claim attributivists must make then is that necessarily, whenever one has a desire for *x*, that desire is a conditional desire for *x,* conditional upon it being the case that *x* is in one's attributive good. If it were not the case, or ceased to be the case, that the latter condition obtains, a desire for *x* would be unable to be satisfied. It would therefore cease to motivate at least an *ideal* agent, aware of all the facts. In the case that satisfying the desire for *x* would impede the promotion of one's Attributive Goodness, doing so would be self-defeating. That

[143] It isn't unusual for philosophers to think of conditional desires *exclusively* in those terms, i.e., in terms of being conditional upon their own persistence. Parfit does in *Reasons and Persons* (Parfit 1984, 151). Likewise Schueler, paraphrasing Platts, argues that "what is desirable to the agent about the object of [a conditional] desire is completely exhausted by the fact that he wants it" (Schueler 1995, 40, Platts 1991). Yet persistence isn't the only plausible condition. Here I am, of course, assuming that a desire might be conditional upon it being the case that the desideratum is in one's attributive good.

[144] See also McDaniel and Bradley's 'Desires', where they write that "B's desire to get a beer might be conditional on the beer being less filling, whether the favourite beer-drinking location is open that night, or just about anything else. A conditional desire might relate a person to any two propositions at all; there need be no interesting connection between the two propositions, and neither proposition need be about that desire itself." (McDaniel & Bradley 2008, 278).

is, obtaining the desideratum *x* alone would *defeat*, not satisfy the agent's overall motivation, which was to obtain *x as* attributively good. Obtaining *x* at the cost of attributive goodness would be satisfying an impulse at the cost of the underlying desire by which it was motivated, and of which *x* was merely a partial desideratum.

Is it plausible to think that all desires are conditional in the above way? Plausible enough. It's very popular to conceive of desiring as inhabiting what's called an *evaluative perspective*. A broad range of philosophers including Scanlon, Railton, Brewer, Darwall and Tenenbaum, Quinn and even Anscombe herself can all be said to subscribe, implicitly or explicitly, to what is sometimes called the Evaluative Outlook conception of desire.[145] The general thrust of the view is that to desire *x* is to be in a psychological state that involves evaluating *x* as somehow reason giving, worthwhile, or good, and being motivated towards *x* on those grounds, where said motivation is subject to modification in light of the truth of the relevant evaluations. It isn't, *prima facie*, terrifically more implausible to think that the evaluative perspective inhabited might be one of evaluating *x* as attributively good.

It's likewise very popular to adopt the 'Guise of the Good' view of desire, intention and action (henceforth 'GG'). Tenenbaum broadly summarises GG as the view "that desire … always aim at the good" (2010, 3). More elaborately the view is that whenever we desire *x*, we do so with the implicit understanding that *x* is in some way a *good* thing to realise. Desiderata, under GG, aren't unconditional objects of desire – a desire for *x* isn't a desire for *x*, whatever *x* might turn out to be. Rather when *x* is desired it's desired under the guise of being good. The desideratum of a desire for *x*, fully cached out, is for *x as good*, and the persistence of our desire is taken to be responsive to the truth of that evaluation. In that sense then, GG is a view

[145] See Anscombe 2000, Brewer 2009, Darwall 2006, Quinn 1993, Railton 2010, Scanlon 1998 and 2002, and Tenenbaum 2007. See also Schapiro, Tamar 2014 for an informative overview.

that entails all desires are *conditional* desires, conditional upon the desideratum actually being that which we evaluate it as.

GG is, again, a very popular view. It is held in some form or another by Joseph Raz, Sergio Tenenbaum, Sebastian Rodl, Matthew Boyle and Douglas Lavin, and historically by a wide range of philosophers, from Plato to Kant, Davidson, and, tellingly, Aristotle, Anscombe and Geach.[146] If GG itself seems so plausible, it doesn't seem like a large leap to think that Attributive-GG, the view that all desires or endorsed desires are for the desideratum understood as being *attributively* good *for* the desiring agent, should also be plausible. Indeed, it may be *more* so, and even already recognised as such. Certainly that's the view that Anscombe and Geach held. Alongside them, Boyle and Lavin (2010) argue that the goodness recognised in GG must derive from teleological consideration, being 'good for' something or other, and Tenenbaum points out that Aristotle, a founding voice in GG, took the end of all intentional action to be "something that is good *for the agent*" (2010, 203). So Attributive-GG should in fact already be considered as a facet of an already established view.

But *is* Attributive-GG really as plausible as an unrestricted form of GG? It's one thing to think that everything one desires one also sees as good *in some way*. It's another thing to think that everyone one desires, one also sees as good *specifically* for their kind. If that were true, we should expect desiring to unpack into a sort of two-sided conceptual exercise, wherein I conceive (i) of myself as a certain *kind* of entity and (ii) of the desideratum as good *qua* that kind. I think though that isn't particularly far-fetched. Here's a way attributivism might explain desires along exactly those lines.

Take a basic bodily appetite. I thirst, and I have a desire to drink from this mountain stream here. I'm about to do so, when my hiking companion, having just tested the water

[146] See Raz (2010), Tenenbaum (2010, 2013), Rodl (2010), Boyle and Lavin (2010), Anscombe (2000), Geach (1956), and Davidson (1980). See also Moss (2010) and Barney (2010) for the Aristotelian and Platonic roots of GG, and Tenenbaum (2010) for an interesting look at the conflicting ways Kant and Aristotle approach the view.

upstream, points out that the water is contaminated with Giardia. I lose my desire to drink from the stream. What's happened?

One explanation is that my thirst, and my desire to quench it, caused me to see the stream as *good*, made it appear desirable, brought into perspective its clear, rushing water, sparkling as it tumbles over the rocks, sending up a cool, refreshing mist, etc. But I didn't come to see it as *just* good, says the attributivist – it's not that I took the stream to be desirable on its own merits, such that anyone would have a reason unconditionally to drink from it. I came to see it as good *for me*. It's important to note though that one can't sensibly come to conceive of something as good *for them* without being informed by at least some conception of what *they* happen to be. When I conceive of *x* as good for me, I must also implicitly conceive of *myself* as the *Kind* of thing for which *x* is good. That then, claims the attributivist, is what it is to desire *x*. To desire *x* is to be in a psychological, motivating state in which *x* is conceived of as giving reason to act in virtue both of *x's* being good for me and of me being the Kind of thing for which *x* is good. And that's what explains the evaporation of my desire when I am informed of the stream's contamination. When I desired to drink from the stream, I implicitly conceived of it as good for the Kind of thing that I am. A thirsting creature, certainly, but also a *Human Being.* Although information about the contamination of the stream doesn't affect its ability to quench my thirst one bit, it does make it the case that it's no longer true that it is good *for* me, because as a *Human* I'm not the Kind of thing for which Giardia is good.[147]

[147] Nobody doubts that bodily appetites are generally conditional. What about less obviously self-concerned desires? Imagine I desire to feed the poor. Is it plausible that such a desire could be conditional in the attributivist way? I think so. Consider the following. I am about to act on my desire to feed the poor, when my partner tells me a strange fact – the world we live in has suddenly become such that the only way for me to help feed the poor is by rounding them up into camps, forcing tubes down their throats, and force-feeding them in a painful and anguishing ordeal. I no longer want to do so. Why? The attributivist's explanation is that when I desired to feed the poor, I did so *as* a Human, with the implicit understanding that feeding the poor is a good thing for Humans to do. Something is good for a Human to do if, for example, it's socially cooperative. So I undertook to feed the poor, thinking that doing so was at least compatible with being socially cooperative. Thus when it turned out that there was no socially cooperative way by which I could feed the poor, I renounced the desire to do so. What's more, I was *right* to do so. Since my desire to feed the poor was conditional upon it being the case that I could do so in a way in keeping with the standards of my Kind, i.e., in a socially cooperative manner, acting upon it in the way available would have been counterproductive.

The *guise of the attributive-good* thesis does what attributivism requires. An analysis of desire that sees all desires, or at least all endorsed desires, as essentially conditional upon desiderata being in one's attributive goodness is both theoretically coherent, and is able to produce and explain intuitive results. The final claim then is that motivational sets are necessarily responsive to attributive goodness because satisfying desires generally is conditional upon doing so in a way compatible with promoting attributive goodness. Desires are so conditional because each act of desiring is an act of conceptualising some desideratum as good relative to an implicit conceptualisation of one's self as a certain Kind of thing. To satisfy some desideratum in a way that impedes attributive goodness is to satisfy some impulse at the cost of the desire that motivates it, or otherwise gives us reason to pursue it. Thus if some desideratum is such that it can't be realised while also being in the attributive good of the acting agent, then their ideal self would not act upon the desire.

I think attributivists should be happy with the above. I think in fact they already are. The thesis above, call it the Guise of Attributive-Good thesis, positions Attributive Goodness not as representing a distinct class of desires, but as something we *need* in the course of our more general activities. Tellingly, attributivist literature is replete with talk of needs and necessities. I won't repeat these arguments at length – they've been covered in detail in previous chapters – but a reminder won't hurt. In this section already we've been reminded that Geach's view was in line with the view above – he argued explicitly for the unavoidable, *necessary*, influence of human-interests on making human choices, decisions *as* a human. Anscombe, Foot and Thompson have made *Aristotelian Necessities* a staple of attributivist lexicon, arguing in doing so that (many) categorical features are those which *need* to be realised

by a given subject if it is to *succeed* in being what it is, and fulfilling its drives and desires therein.[148]

Talk of Attributive Goodness inevitably descends into talk of *needs*, *necessities* and *requirements.* It should be no surprise then that the best version of Attributive authority is one that sees Attributive Goodness as a requirement generated in and entailed by the act of desiring itself.

To summarise, Attributivism's best tactic to explain the authority of kind-interests over desires is to establish the furtherance of kind-interests as a constitutive part of endorsed desires, such that no such desire can be fulfilled while compromising kind-interests. No desire is undertaken unconditionally, says the attributivist, rather every desire is undertaken under the 'Guise of the Attributive Good', the good, that is, of the kind of thing *as* which the desiring agent acts. The authority of attributive goodness is predicated upon it being an unavoidable condition upon desire fulfilment, a *necessity* on the path to desire's attainment.

[148] See Anscombe, 'On Promising and Its Justice' (1981b, 15, 18-19), 'On the Source of the Authority of the State' (1978a), and 'Rules, Rights and Promises' (1978b), Foot, *Natural Goodness* (2001, 46), and Thompson, *Life and Action* (2008, 73).

§6.2 – An Abundance of Kinds

In this section I will argue, ultimately, that Attributivism as it's classically understood fails. I argue it does so because it depends upon a sort of kind essentialism – the idea that agents belong necessarily to only one kind – when in fact that's not the case. I argue that agents can and often do belong to multiple kinds. I argue that because of that, attributive goodness comes out to be non-authoritative, and thus not fundamentally normative. Before I develop that argument however, let's, one final time, review just exactly what Attributivism claims, and just what I'm arguing is false.

Canonically, Attributivism says the following. For any agent *x*, there is a kind *K* such that (i) *x* belongs to *K* and (ii) *x* ought to do only what is good *qua K*. It argues that attributive goodness, goodness *qua* one's kind, is *the* fundamentally normative property, the *source* of normativity, and that it produces normative obligations to act in the above way. In short, it argues that whether or not *Φ*ing is attributively good *qua* a kind *K* to which *x* belongs determines whether or not *x* ought to *Φ*.

For sake of simplicity, I'll put the above in terms of a single case, the *human* case. Canonically Attributivism argues the following: if *x* is *human*, then *x* ought to do only what is good *qua* human, and *because* it is good *qua* human. Exact formulations vary. Allyn Fives paraphrases that "[W]hat a person should do all things considered is that which he or she should do because human good hangs on it" (Fives 2008, 172). Anselm Müller says that human agents ought to act on reasons "that have to be acted on where a society of human beings is to get on well" (Müller 2004, 29). What all agree on though is something like the following: if murder is bad *qua* human, then a human agent ought not to murder. I take, furthermore, 'ought not to murder' as synonymous with 'ought not to act in ways that result in murder', and take the preceding claim to extend unreservedly to claims like the following: a human agent ought not to murder, or bring it about that they murder, or contribute to murdering. That is, a human

agent ought not to murder right now, they ought not to act in ways that will result in their murdering later – say by laying traps designed to murder – and they ought not to act in ways that will contribute to somebody else murdering. They, generally speaking, ought not to bring murder about.

As I have now argued several times, for the above to be true, human goodness must be *authoritative* for human agents – it must be the case that whenever an agent is *human*, that agent's ideal self would be overridingly motivated to act in the above ways. In short, for Attributivism to be true, the ideal human agent must be overridingly motivated not to murder, not to act in ways that will result in their murdering, not to help others murder, etc.

What I will argue across the next several sections is that attributive goodness *isn't* authoritative in the above way. I will argue that there is no reason to believe that whenever an *x* belongs to kind *K*, the ideal version of *x* will be overridingly motivated to act only in ways that are good *qua K*. I will make that argument for the following reason. As discussed in 6.1, Attributivism's best prospect for accommodating **Authority** is through the *guise of the attributive good thesis*. That thesis specifies that every desire for *y* held by agent *x* will be able to satisfied only if it's the case that obtaining *y* is good for the agent's kind. The problem for Attributivism is, however, that the thesis *doesn't* specify any *specific* kind. That is, it doesn't say that if *x* belongs to *K*, that *x*'s desire for *y* will only be satisfied if *y* is good *qua K*. It says only that obtaining *y* must come out as good *qua a* kind the agent belongs to. That, I'll argue, is a problem for Attributivism, for the following reason.

Agents can and do typically belong to *multiple* kinds. Because of that fact, for example, a human agent's desire for *y* might plausibly be satisfied even when *y* is *not* good *qua* human, just as long as *x* is good *qua* some *other* kind the agent belongs to. As such, that agent might have intelligible reason to disregard *human* goodness, in favour of goodness *qua* other kinds she belongs to. Thus, as I'll argue, it's false that if *x* belongs to kind *K*, then *x*'s ideal self will

have overriding motivations to act only in ways that are good *qua K*. Thus Attributivism, which claims that whenever *x* belongs to *K*, *x* ought to do only what is good *qua K*, is false.

To begin with then, consider the following. Grant the supposition of the previous section. Every desire is held under the guise of desiderata being good for a kind the agent instantiates. Likewise, acting in a way that is bad for the kind under the guise of which some desire is held will frustrate the satisfaction of that desire. Here then is a question: must every desire held by a human agent be held *exclusively* under the guise of *Human* goodness? Must it be held even *primarily* under the guise of Human goodness? No.

Recall the example discussed previously, in which a human being desires to drink from a stream that is, unbeknownst to her, contaminated. Grant that it's true that when she forms that desire, she does so under the guise of attributive goodness – its satisfaction is conditional upon drinking being good for her kind. That needn't entail though that when she forms her desire to drink she does so specifically under the guise of *human* goodness. That's obvious when we consider that she might form her desire without even knowing *what* kind she belongs to. All that must be true is that her desire to drink contains the implicit assumption that doing so is *in some way* attributively good for her, good for *a* kind she belongs to, *whatever* that kind turns out to be. So it seems her desire to drink is satiable, all things being equal, just as long as when she drinks she belongs to a kind for which drinking the water is good.

Consider also the case of a terrorist who desires, in the name of some ideology or another, to destroy the structures of society. We must grant, from the previous section, that when he desires violence that desire is two-sided – he desires both to commit violence and that doing so be good for his kind. He must, implicitly, conceive of himself as a kind of thing for which enacting violence is good. It strains credulity though to think that the kind he has in

mind, so speak, is *human*.[149] What seems rather more plausible is that when *x* sets out to bomb an embassy or exterminate a population, he acts with an implicit understanding of himself as a revolutionary, a political radical, a protector of a certain ideology – a villain if he's unusually honest – and that he acts with the understanding that doing so is good *qua that* kind.

Assume that I'm right about the above. Assume now also that in fact the subjects of those cases actually do belong to multiple kinds. The thirsty hiker is human, but is *also* a mutant, with a special secondary digestive system. Her human digestion will suffer if she drinks contaminated water, but her mutant digestive system will flourish. Likewise imagine that the bomber in the second case is human, and a revolutionary, and moreover that to be a revolutionary is to belong to a distinct *kind* of thing, a unique sustainable way of life, wherein destroying social structures is good. In these cases though, unless there is some reason for the subjects to prioritise their *human* way-of-being, it seems as though their desires stand to be satisfied *regardless* of whether or not it is bad *qua* human. The desiderata of their desires *will* be attributively good for *a* kind they belong to, and so those desires will be coherently actionable by their ideal selves.

I think the above represents a particular and insurmountable problem for Attributivism. That is because I think that agents *can* belong to multiple kinds with conflicting standards. Likewise I believe that there is no method whereby Attributivism can demand that one kind be prioritised over another. What that means, I will show, is that the attributive goodness of any *particular* kind won't, in itself, satisfy **Authority**. Despite an act being *bad qua* one kind, it may be good *qua* some other kind one belongs to, and so believably something one's ideal self might do.

[149] Or if he does so, his concept of Human Kind is so different from ours that we may as well be speaking of two different Kinds altogether.

Let's not get ahead of ourselves though; the above relies on it being true that there is no justification for normatively prioritising the Human kind, or indeed any particular kind at all, over any other. Why believe that? The details of a related debate can provide some insight. The attributivist account that we have come, by now, to recognize is similar in many ways to Constitutivist accounts developed over the last two decades. Such accounts are championed in one form or another by a wide range of philosophers: Katsafanas (2013, 2018), Korsgaard (1999, 2008, 2009), Silverstein (2008, 2012), Smith (2010, 2012, 2013), Street (2008, 2010, 2012), Rosati (2003) and Velleman (2000, 2009). All follow more or less the same process. They argue, roughly, for the existence of constitutive aims of agency – aims one must have to *be* an agent, and that one must have *insofar* as one *is* an agent. They then argue for the authority of such aims in virtue of the *inescapability* of *being* an agent. We can't, they claim, do otherwise than to be and embrace being an agent, at least insofar as we engage in deliberation, and so can't help but be committed to agency's constitutive aims. The parallels with Attributivism as we've come to understand it are stark.

An influential objection to the Constitutivist project comes in David Enoch's 'Agency, Shmagency' (2006). Enoch's objects that once constitutivists provide a reductive analysis of agency, it isn't clear why agency *is* something we can't help but engage with, or why we should care about its aims. Constitutivism is easy to accept when all that's asserted of agency is that it's the sort of thing we have in virtue of our capacity to deliberate. That's because at this point being an agent is compatible with whatever aims we might actually have. There's no way to fail at being one. Once Constitutivism provides a specific analysis of agency however, and being an agent becomes something we can do badly, it becomes much less clear why we should care, less clear that being an agent is something we can't help but be.

Enoch makes many arguments in 'Agency Shmagency', but he makes the above point explicitly in his objections to Rosati and Korsgaard. Sure, he says, grant that some of my aims

count as constitutive of a particular analysis of agency. Given that I am already sceptical of the normative authority of many, if not all, *particular* aims, then "why does it matter, as far as the question of normative arbitrariness is concerned, that some parts of [my] psychology have this necessary-agency status? Why shouldn't [I] treat the motives and capacities constitutive of agency as normatively arbitrary?" (Enoch 2006, 178). The gist of Enoch's point here is that once we've classified agency in a way such as to make it simply one of many projects we might or might not engage in, we lose justification for attributing normative priority to it.

> "Classify … me as you like. Perhaps I cannot be classified as an agent without aiming [at *x*]. But why should I be an agent? Perhaps I can't act without aiming at [*x*], but why should I act? … I am perfectly happy being a shmagent – a nonagent who is very similar to agents but who lacks the aim … of [*x*]." (*Idem* 179).

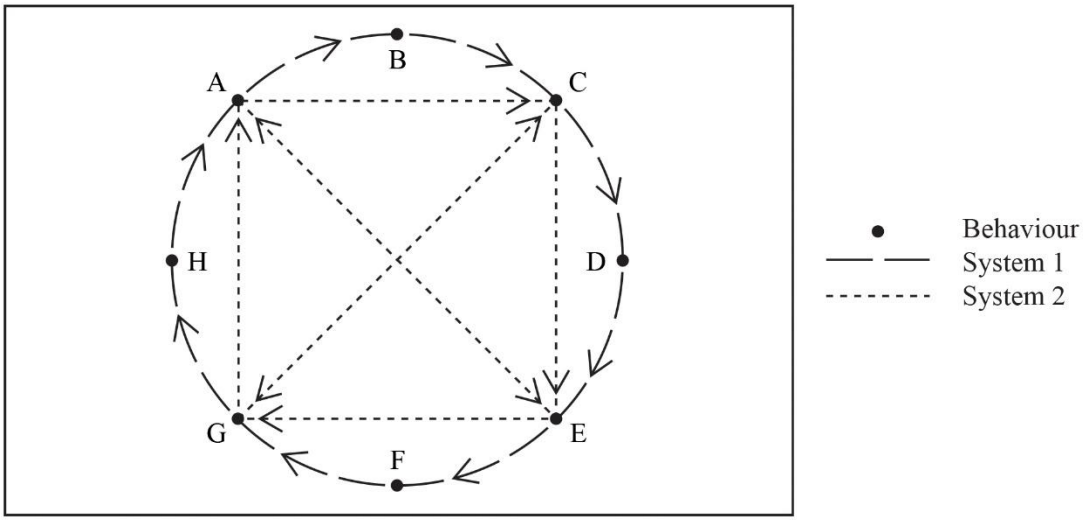
The gist of Enoch's argument is this: the fact that one can't *be* an agent without engaging in some specific activity means nothing if there are viable alternatives to being an agent, which there are once agency is given a specific analysis. Take for example a constitutivist view proposed by Velleman in *The Possibility of Practical Reason* (Velleman 2000). Velleman argued that intentional action has the constitutive aim of corresponding to the antecedent beliefs of agents about what they are going to do. Actions that don't correspond are therefore poor actions – a person who pours their whisky on the ground when they meant to drink it has committed a poor action, because it has failed at the constitutive aim of action, i.e., to be intelligible in respect to the beliefs of the agent about what they were going to do. But say I don't care whether or not my actions end up being what I expect them to be? What if I happen to enjoy the mystery of not knowing what I'm going to do, or live a charmed life where my instinctual movements always produce the best result, regardless of whether or not I meant to make them? Why then should I care about whether or not I am an agent, or committing good

actions? I seem entitled to be content with being a shmagent, and performing shmactions instead. In that case, I have no obvious reason to be motivated by what is constitutive of agency or action.

The above points apply naturally to Attributivism. So what if I belong to kind *x*, if there are viable alternative ways of being? I have no reason to be bothered by failing at *x*'s constitutive standards, if there is a viable alternative kind *y* to which I can also belong, and be successful as. Moreover, once *x* is given a restricted definition, there *are* viable alternatives. If belonging to a kind is a matter of instantiating a system of characteristic, reciprocally-dependent behaviours, then it becomes possible for one to belong to *multiple* kinds, by instantiating multiple, independently sustainable, systems of behaviours. At that point though, there's no reason any one kind should take priority over another.

A visual aid may be helpful in making the above clear. Take the following to visualise the instantiation of multiple kinds by a subject, in virtue of its instantiation multiple systems of independently sustainable reciprocally dependent behaviours.

(a.)



The above diagram demonstrates the ways the behaviours of a subject (indicated by points A-H) might interact with each other, such as to allow for the formation of multiple, distinct, inter-

dependent systems. The same subject might instantiate one kind in virtue of the subject's behaviours instantiating system *A-H*, while *also* instantiating a second kind in virtue of instantiating, for example, system *ACEG*. Although those systems overlap in some respects, neither *A-H* nor *ACEG* requires the persistence of the other for its own persistence, or maintenance. Thus it seems the subject has no particular need to care about its success as a member of the kind marked out by *A-H*, if it can maintain itself as a member of the kind marked out by *ACEG*.[150]

Perhaps, for example, being Human consists in instantiating a system wherein, among other things, rationality and social cooperation interrelate. Humans are creatures whose rational faculties allow them to establish socially cooperative social groups, and whose rational faculties stand to be, in turn, maintained and shaped by those social groups. Say I am Human. Say however that I'm *not* socially cooperative, and moreover that I instantiate a second set of interrelated behaviours, which allow me to maintain myself in a different way. It seems that I'm entitled to say: "Very well, I'm a bad Human. But I'm an excellent *shmuman*, and happy to be so. Even if I will not flourish in the particularly *human* way by rejecting social relationships, why should I prefer human flourishing to shmuman flourishing, given that both are viable ways of being?" I seem to be equally justified in acting *qua* one kind as I do the other. That's a point worth belabouring, and again we might visualise the problem as follows.

[150] If it helps to illustrate the above by way of example, say that instantiating system *A-H* marks one out as a wild dog, and maintaining system *A-H* is required to be a successful wild dog. Meanwhile, instantiating the system *ACEG* marks one out as a domesticated dog, and maintaining *ACEG* is required to be a successful domesticated dog. Say that points B and D and F represent the following dispositions or behaviours (B) hunting in packs (D) barking at threats and (F) fighting rival dogs for food. Neither system of interdependence requires the persistence of the other – a dog needn't be a domesticated dog to be a wild dog, nor vice-versa. A dog then, even one that *does* instantiate both systems, has no particular need to maintain any one system over the other – if it loses its ability to be a successful wild dog, its life won't be impacted in a meaningful way, given its ability to succeed as a domesticated dog.

(b.)

Subject	Kind	Required Behaviour *qua* Kind
x	K_1	$\rightarrow \Phi$
	K_2	$\rightarrow \neg \Phi$
	K_3	$\rightarrow \Psi$
	K_4	$\rightarrow \neg \Psi$

In the above, attempting to derive meaningful normative implications for X's action in virtue of *x*'s kind is futile: as *x* instantiates a wide range of kinds, each with their own, often conflicting, requirements, any action *x* might be 'required' to take in virtue of *x*'s membership in one kind might be justifiably disregarded under the auspices of *x*'s membership in yet another. What should it matter to *x*, on attributivist terms, that as a *K* it is right to *Φ*? As a K_2 it is wrong, and so *not Φ*ing is equally justifiable. Likewise if *x* feels disinclined to even worry about *Φ*ing, membership in kinds K_3 and K_4 justify *x* taking an interest in *Ψ*ing instead, positively or negatively.

The above sort of cases are plausible. To further elaborate on that plausibility, and on the problem it poses, consider the following cases.

A tiger is captured and featured in a zoo. Let's stipulate that among a tiger's categorical features are traits such as ferocity, a predatory instinct, the ability to hunt and kill prey, sharp teeth and claws, etc. As it happens though, *this* tiger is quite a poor one; its claws are blunt, its demeanour placid and amiable, it's largely uninterested in hunting its own prey. But, at the same time, those traits which make it defective as a tiger cause it to excel as a zoo animal. I'll stipulate too that this zoo is a *humane* one, an environment in which an animal can exist without experiencing any mental or physical strife that isn't, at least, compensated for. Is it of any normative importance whatsoever for the tiger in question that it is defective as a *tiger*, or that being a successful zoo-animal entails being a defective tiger? None at all. If anything, the fact

that it is a bad tiger seems to give us just as much reason to promote an *alternative* life-style for the animal, instead of expecting or pushing it to be a good tiger. If the kind 'zoo-animal' is a sustainable, viable alternative to the kind 'tiger', then attributivists seem to offer no reason for why the latter should be preferred over the former.[151]

If Attributivism has difficulty establishing authoritative norms for even relatively simple animals, we should be even more sceptical of its chances regarding more sophisticated models. Take the example of Tammy Baritone. Tammy is a human being. I'll even grant that in some sense she's necessarily so. I'll grant too the common stipulation that to be Human is to be a socially-cooperative and practically rational animal. A token Human is defective if unable to form cooperative social bonds or exercise rationality towards (human) ends. Tammy, as it happens, is a very *successful* Human: she's intelligent, even-tempered, comes to well-considered conclusions on practical issues. She's fit, healthy, with a robust social circle; her family adores her, she's well thought of and popular, she's fond of her friends, treats them well and enjoys their successes. She even runs a non-profit charitable organisation and holds a prominent political office, doing well by her constituents therein. To make this picture as generous as possible for Attributivism, it's even true that her interests are generally receptive to her status as a Human, such that being defective as a Human will bear *to some extent* upon her deliberations, and the satisfaction of her desires.

Tammy is also the head of a major crime syndicate. She rules a criminal empire through fear and violence: she murders, she rapes, she tortures, she trades in human flesh and abuses her political influence. She's tremendously successful in this criminal capacity, and being so allows her to maintain herself and to satisfy a wide range of interests and desires. Make no mistake though, by exemplifying her Criminal traits, Tammy is defective as a Human – she is

[151] The extent of attributivist responses to this sort of case seems to be summed up in the following from Foot: "Even in a zoo a fleeing animal like a deer that cannot run well is … defective and not as it should be, in spite of the fact that … this may be no disadvantage" (Foot 2001, 34). Foot writes as if there is, as such, no problem here, but the response as I see it can only be 'fine, this creature is defective as a deer; so what?'

monstrous in fact, by Human standards. Even should some Human-interests benefit from Tammy's criminal activities – say the money she extorts allows her to treat her friends and family well – this is incidental, and she could be just as successful *as* a Criminal were she without Human-interests at all, or if her criminal traits were in conflict with them. And they are – imagine now that Tammy has reached a point where both her Human-interests and Criminal-interests can't coincide, one must suffer for the benefit of the other. Is there any authoritative fact Attributivism can offer about which Kind Tammy ought to embrace, and thus any fact about any particular actions she authoritatively ought to take? No. Any reasons attributivists can appeal to – that is, reasons grounded in kind facts alone – to favour one kind can also be appealed to in favour of the other. Once Tammy has come to belong to a viable alternate non-human kind, the needs and interests attributed to the human kind become normatively arbitrary.[152] [153]

My view is that examples like the above are coherent, and that they pose an insurmountable problem for Attributivism. I'll summarise why.

(i) If Attributivism is true, then attributive goodness is fundamentally normative.

[152] One might anticipate this objection: being a criminal doesn't constitute belonging to a particular kind, in the way that being a human or a tiger does. My response to that objection will appear in §6.4, but in brief, given the non-historical, reciprocal-dependency method through which attributivists must understand kinds, I don't believe they have any grounds to reject a very permissive account of kinds, whereby being a criminal will indeed fit the bill.

[153] It may be worth taking a closer look at what it means for a kind to be a 'viable alternative'. Some may disagree that kinds like 'criminal' or 'zoo-animal' really are viable alternative ways-of-being. For example, Enoch, when he writes as if being a schmagent is a viable alternative to being an agent, seems to rely on the intuitive impression that schmagency is 'just as good' as agency. That intuition may not carry over into other cases. Where a shmagent stands to be just as happy, or lead just as good a life, as an agent, we might think an animal would be *happier* as a wild tiger than as a zoo-animal, or that a non-criminal leads an objectively *better* life than a criminal. So we might think that zoo-animals and criminals aren't really *viable* alternatives at all. The response to that is brief. The argument may well be true, but it's not the sort of argument *attributivists* can appeal to. To do so would be to ground normativity in facts like whether or not the subject would be *happy*, or whether or not it is leading the *good life*, and not, as attributivists require, whether or not kind standards are being met. Within Attributivism, whether or not being kind *x* is normatively viable way-to-be can only depend upon whether or not one can exist as an *x* while meeting its standards. If an animal can meet the standards of being a zoo-animal as well as, or better than, it can meet the standards of being a tiger, then normatively speaking being a zoo-animal is, at least, 'just as good'.

(ii) If attributive goodness is fundamentally normative, it is non-derivatively authoritative.

(iii) Attributive goodness is non-derivatively authoritative *iff* whenever *x* belongs to *K*, *x*'s ideal self will be overridingly motivated to act only in ways that are good *qua K*.

(iv) If *x* belongs to and can succeed as a member of distinct alternative kinds *K* and *K*+, then *x* has no greater reason to be good *qua K* than to be good *qua K*+.

(v) Plausibly, *x* does belong to and can succeed as a member of distinct alternative kinds *K* and *K*+.

Therefore,

(vi) Plausibly, *x* has no greater reason to be good *qua K* than to be good *qua K*+.

Therefore,

(vii) It is not the case that necessarily when *x* belongs to *K*, *x*'s ideal self will be overridingly motivated to take only acts that are good *qua K*.

Therefore,

(viii) Attributive goodness is not non-derivatively authoritative, and so not fundamentally normative.

I believe that the above is part of a decisive argument against Attributivism. Yet it may seem vulnerable to a particularly obvious objection. I'll end this section by explaining it.

Everything said thus far relies on subjects instantiating multiple kinds. That suggests a simple revision to Attributivism. Rather than taking Attributivism to say that when *x* belongs to kind *y*, the goodness of *y* is fundamentally normative for *x*, say instead that when *x* belongs to a *set* of kinds, goodness among members of that set is fundamentally normative for *x*, such that *x* ought to do only what is good *qua at least one* member of that set. It doesn't seem immediately obvious that *that* formulation fails to satisfy **Authority**. The claim now is that

attributive goodness among members of the *set* of one's kinds is authoritative, because whenever *x* belongs to a set of kinds *y, z ... n,* necessarily *x*'s ideal self will be overridingly motivated to take only actions that are good *qua at least one of y, z ... n*.

That seems plausible – if the *guise of the attributive good thesis* is correct, a desire for *y* won't be satisfied unless achieving *y* is good for a kind the agent belongs to. Although, as has been established, that doesn't require that achieving *y* must be good *qua* any one, *specific* kind the agent instantiates, it does still seem to entail that unless *y* is good *qua at least one* kind the agent instantiates, the desire will be unable to be satisfied. That seems to entail that an agent's ideal self will take only actions that will result in goodness *qua at least one* kind she instantiates, and not take actions that are bad *qua every* kind she instantiates. Moreover, there does seem to be a substantial range of actions that will fall into the latter category. Nobody, after all, will instantiate a set of kinds wide enough to simultaneously justify *every* action. Even if Tammy genuinely is both a human and a criminal, such that her ideal self might intelligibly care for *or* torture others, neither of those kinds make it good for her to, say, dedicate herself exclusively to seeking out and turning on radios, or to drink gasoline, or to renounce her life to live wild in the woods and eat nothing but berries and salmon. Those actions are defective *qua every* kind she instantiates, so at least she ought not to do *those* things.

If the above revision is plausible, then the objection laid out in this section fails. Fortunately I don't think it is plausible. The revised view says that attributive goodness is authoritative because, necessarily, facts about goodness among the *set* of one's kind entail overriding motivation in one's ideal self to take only actions that are good *qua* at least one member of that set. As I'll explain in the following section, I don't think that's true. That is because not only are agents able to instantiate multiple kinds, but the kinds that they do instantiate are liable to change. If agents can take control of the set of kinds they *do* instantiate – and I think they can – then it will not be true that when one belongs to a set of kinds, one's

ideal self would necessarily be overridingly motivated to act only in ways that are good for members within that set. So, as we will see, attributive goodness itself will remain non-authoritative.

§6.3 – KIND MANIPULATION

There are two arguments I will make in this section. First, if we can come to instantiate kinds beyond those in our current set, that's a problem for Attributivism. Second, we *can* plausibly do so.

I'll start with a reminder from the last section: the proposal now on the table is that attributive goodness among the set of kinds one belongs to is fundamentally normative. What that means is that any agent ought to take only actions that promote goodness *qua at least one* kind the agent actually belongs to. As we know by now though, for that to be true, it must be the case that the agent's ideal self would be overridingly motivated to take only actions that promote goodness *qua at least one* kind the agent actually belongs to. For example, if an agent is a human, a criminal, and a philosopher, we now propose that they ought to take only actions that are good *qua* at least one of human, criminal, or philosopher. For that to be true, it must be the case that their ideal self would be motivated only to take actions that are good *qua* at least one of human, criminal, or philosophers.

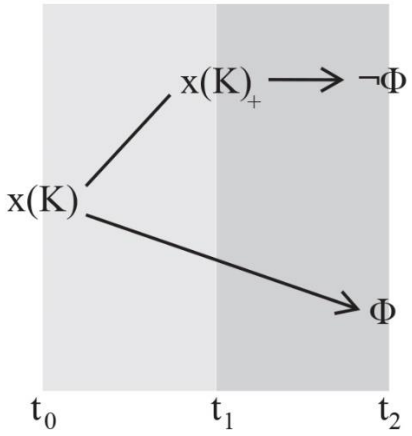
That seems *prima facie* plausible. The reason why comes from the *guise of the attributive good thesis.* If a desire for *y* can only be satisfied if achieving *y* will be good for a kind the agent belongs to, then it *seems* like the agent's desires will only be able to be satisfied if *y* will be good *qua* at least one kind they actually belong to. i.e., if the above agent's desires are to be satisfied, it seems that achieving the object of his desire must be good *qua* at least one of human, criminal, or philosopher.

Actually though I think that isn't true. As I'll argue now, it isn't true specifically because agents can act to instantiate kinds that are not within their current set. What that means is that a desire for *y* that is *bad qua* all the agent's actually instantiated kinds can still be satisfied, just as long as the agent *comes to* instantiate a kind *qua* which *y* is good, prior to obtaining *y*. All that's required to satisfy a desire under the *guise of the attributive good thesis*,

after all, is that when *y* is achieved, the agent belongs to a kind *qua* which *y* is good. Thus it isn't plausible to believe that, in virtue of an agent currently belonging exclusively to the kinds human, criminal, and philosopher, that the agent's ideal self will be motivated only to take actions that are good *qua* at least one of those kinds. They might very plausibly be motivated to do some act *y* that is *bad qua* each of those kinds, in virtue of knowing that they will come to instantiate a kind *qua* which *y* is good. There is, in short, no reason for *x* to care about whether or not *Φ*ing is bad *qua K*, if *x* will no longer *be* (exclusively) *K* by the time *x Φ*s. Thus, goodness among the set of one's kinds comes out, again, non-authoritative, and so non-fundamentally normative. The following will now elaborate on, and illustrate, the above.

Consider the following model.

(c.)



In the above x(k) represents an agent *x*, who belongs to kind *K*. At time t_0, *x* is poised to decide whether or not to *Φ*, which would occur at t_2. According the standards of *K*, *Φ*-ing is attributively good, and not *Φ*-ing is defective. Thus according to Attributivism, *x* ought to *Φ*. But *x* doesn't want to *Φ*, he wants to *not Φ*. What's more, the option is available for *x* to

instead act in such a way as to instantiate kind *K*+. For a member of *K*+, *not* *Φ*-ing is attributively good, and *Φ*-ing is defective, so a member of *K*+ ought not to *Φ*.

The conclusion I draw from the above is that at t0 there is no fact of the matter as to whether or not *x* ought to *Φ*. That's *despite* it being the case at t_0 that *Φ*-ing would be defective *qua* his current kind(s). That's because I can see no reason that the analogue of Enoch's *schmagency* argument applied in the last section doesn't apply equally here. Just as there is no reason for *x* to overridingly care about the standards of one of his kinds if he is also a member of a viable alternative kind, there is no reason for *x* to overridingly care about the standards of his current kind(s) if he *can* be a member of a viable alternative. Even granting that at t_0 not *Φ*-ing is bad for *x*, and that the satisfaction of *x's* desire at t_0 to not *Φ is* conditional upon not *Φ*-ing being *good* for *x*, *x*'s desire will still be satisfied when the action occurs at t_2 just as long as *x* comes to instantiate *K*+ first, so that not *Φ*-ing *ends up* attributively good for *x*. So there's no reason to think that *x*'s ideal self wouldn't to be motivated to not *Φ*, despite not *Φ*-ing being defective qua the set of *x*'s kinds at t_0. Thus it's not the case that *x*'s ideal self will be overridingly motivated to act only in ways that are good *qua* at least one of the set of *x*'s kinds at t_0. Therefore attributive goodness within the set of *x*'s kinds at t_0 is not non-derivatively authoritative, and thus not fundamentally normative.[154]

The above has so far been abstract. Let's look at a more concrete example. In fact, in order to be as charitable as possible, let's look at a case that involves extreme defectiveness *qua* the set of one's current kinds. Take Tammy once again. The last section ended with the intuitive proposition that Tammy ought not to do things like drink gasoline, or dedicate herself primarily to turning on radios, or to living like a bear in the woods. Attributivism will suggest that that is because doing so is defective *qua* all of her currently instantiated kinds. So say that

[154] A possible objection is worth mentioning here. It might seem that what *x* is really motivated to do isn't to not Φ, but to *become K+ and not Φ*. That isn't exactly equivalent to being motivated to take the defective act of not Φ-ing. Yet nothing changes if we stipulate that becoming a member of K+ is itself defective qua *x*'s kinds at t0. It would still be intelligible for *x*'s ideal self to take that action despite its defectiveness, and so the point remains.

Tammy *did* develop a desire to drink gasoline – a persistent and significant desire to do so. Is the fact of that act's defectiveness *qua* all members of her current kind set enough to guarantee that Tammy's ideal self will nevertheless be overridingly motivated not to drink gasoline? I think not.

Imagine that Tammy lives in a fantastic, futuristic scenario, in which genetic and body modification is common place. Existing in this world are mechanical people who drink gasoline daily to fuel their internal-combustion hearts. What's more, Tammy's good friend Dr. Scifi has perfected a technique for installing robotic parts in organic bodies. That process too, as it happens, is defective *qua* the human kind – doing so is absolutely destructive to the human way of life, and requires sacrificing many human goods. Yet in virtue of her desire to drink gasoline, Tammy is willing to undergo the procedure. And I think there is no authoritative reason, based solely in facts about her current kinds, that she shouldn't do so. The thought that there is such a reason was based on the guise of the attributive good thesis, and the idea that if Tammy pursued a desire that was bad *qua* all of her current kinds it would be unsatisfiable under that thesis, and thus not something her ideal self would pursue. That's apparently not the case. Tammy's desires, whether or not their desiderata are bad *qua* her *current* kinds, *are* able to be satisfied. That's because by the time she is finished achieving her desires, they will *not* be bad for her kinds. As such there's no reason to believe her ideal self would be overridingly motivated against their undertaking. Thus it's not true that whenever one is a member of a set of kinds *y, z … n*, and that set exhausts one's kinds, that one's ideal self will be overridingly motivated to take only actions that are good *qua* at least one of those kinds. Tammy may sensibly take and pursue a wide number of actions that are defective *qua* her current kinds. Thus even the revised, set-based version of Attributivism is non-authoritative, and thus not fundamentally normative.

So I argue that, *if* agents can come to instantiate kinds beyond those they already belong to, that attributive goodness within the set of one's kinds is not fundamentally normative. Given the fantastic sort of example I've so far employed though, one might wonder if it's really plausible that we can do so. Is it plausible? Yes.

The only reason I can think of for assuming that the above is *not* plausible is if we adopt a particularly rigid view of kinds and membership therein. We might, for example, believe that kinds are purely biological classifications, and that membership within kinds is determined entirely by genetic lineage. A bear, for example, is an animal born to bear parents, its kinds are exhausted by that fact, and immutable in virtue of it. Whatever else happens, we might say, it is a bear for life in virtue of its parentage. That though isn't the sort of kind analysis available to Attributivism. For reasons given earlier, particularly in §3.1, Attributivism is untenable if we understand kinds in that way.

As we should now be familiar with, within Attributivism kinds are determined by the participation of a subject in a system of behaviours for and from which functional standards can be derived. As Thomson points out, that allows for a wide variety of kinds, many of which don't fit with the above: toasters, lawnmowers, seeing-eye dogs, tennis players, beefsteak tomatoes, tigers, humans, etc. (Thomson 2008, 19-22). If we understand kinds in that way though, membership within kinds rises and falls with the instantiation of such behavioural systems. Give that understanding, I take it as uncontroversial that we *can* come to belong to new kinds – and cease to belong to others. Nobody will deny that we can come to take part in new systems of behaviours. One isn't born a tennis player after all, one comes to *be* a tennis player by doing the things tennis players do.

So it is plausible that we can come to belong to new kinds, and thus that our ideal selves might be motivated in virtue of that possibility to act in ways that are defective *qua* our current kinds, and so implausible that goodness within the set of our current kinds is in itself

authoritative. It's not true that whenever *x* belongs to a set of kinds *y, z ... n*, and that set exhausts the kinds that *x* belongs to, that *x*'s ideal self will be overridingly motivated to act only in ways that are good *qua* at least one of *y, z ... n*. So goodness within that set is not non-derivatively authoritative, and not fundamentally normative.

§6.4 – Many Kinds of Objections

The range of potential responses to the material of the previous sections is extensive. In this section I'll consider four families of responses, and explain why I don't find them troubling. In §6.4.1 I'll consider the objection that I am too permissive with what I take to be kinds. §6.4.2 considers that the results of the cases I have employed are intuitively implausible. §6.4.3 finally considers that quite often circumstances *won't* be such as to allow agents to escape the influence of their kinds in the ways I have laid out, and that it's alright for attributive goodness to be just 'occasionally authoritative'.

§6.4.1 – The Wrong Kind of Kinds

I have been very permissive with what I take to be kinds. In my book, the sort of kinds one might belong to in order to derive meaningful normative standards includes just about anything at all. Organisms, animals, humans, etc., all these are kinds. But so too are criminals, sociopaths, and hedonists. Many attributivists might wish to deny that being one of the latter sort of things represents belonging to a *kind* at all. The point of this sub-section is to explore that objection, and to explain why attributivists can't be and aren't committed to that sort of view.

There might be many reasons why one would want to deny that, for example, being a criminal can constitute belonging to a distinct kind of thing. There are genuine arguments to be had about what the proper analysis of a kind is.[155] However, for our purposes here, nothing hangs on the final outcome of those arguments. What is important here is whether or not the method of kind individuation that is and must be employed under Attributivism validates my view, whereby being a criminal can constitute belonging to a kind. There are two reasons to

[155] Classically, for example, Aristotle himself may have wanted to deny that being a criminal is being a distinct *kind* of thing. Being a criminal, it might have been argued, is only an accidental property, defined relative to the interests of pre-existing entities, rather than marking out a distinct way of existing.

think it does. Both of those reasons can already be found in this thesis, but I will explicitly reiterate them here.

The first reason is precedent. As we were only just reminded in the previous section, some very prominent attributivists already very clearly adopt a similarly permissive view of kinds. Thomson takes the relevant kinds in Attributivism to be what she calls 'good making kinds', any sort of kind with a function or defining behaviour in virtue of which standards of success are derived. These include such disparate kinds as toasters, seeing-eye dogs, tennis-players, beefsteak tomatoes, tigers and humans (Thomson 2008, 19-20). So it's apparent that a degree of permissiveness about kinds is already an element of at least some very notable attributivist accounts. If being a tennis-player can constitute kind membership, then certainly so too can being a criminal.

The second reason is a conceptual one. As was reviewed most significantly across chapters 2 and 3, and particularly in §3.1 and §3.2, the many objections to Attributivism inevitably demand that attributivists present a clear, ahistorical, functional analysis of kind membership. Towards that end, the 'Foot-Thompson' account of lifeform analysis classifies individuals as belonging to certain kinds in accordance with whether or not their behaviours constitute a particular 'way of living'. That means that an individual is a member of a kind when that individual's behaviours can be interpreted as being part of a system of self-maintenance, involving characteristic ends and methods of attaining them (Thompson 2004, 52, 2008, 49-79). For example, an organism is a Bee when it instantiates a system oriented around and sustained by, amongst other things, the maintenance of the hive, social hierarchy, and pollination. An organism is a tree when it instantiates a system oriented around and sustained by the spreading of a photosynthetic canopy and a vascular root system. And so on. The result of that account, among others, is the reciprocal dependency test for kind

membership.[156] One is a member of a kind when one instantiates a system of reciprocally dependent behaviours – a sustainable system of causally connected and interdependent ends and behaviours, representing a distinct way of being.

Once we accept a view like the above however, there is no justification for, by default, excluding being a criminal from belonging to a kind. When one is able to instantiate a system of behaviours oriented around and sustained by the pursuit of criminal activity, then one has just as much right to be seen as a member of a distinct kind as a bee or a tree does in virtue of the systems *they* instantiate. That isn't to say that *whenever* one breaks the law, and so acts criminal-y, that one is a member of the criminal kind. When that sort of behaviour exists within a sustainable system of behaviours oriented around criminal activity though, as was the case in the arguments of the previous sections, *then* one is entitled to kind membership, and all that results thereby.

So I don't believe attributivists can coherently object to my permissive view of kinds. Many of them already accept it. Meanwhile, because of the sort of ahistorical lifeform analyses required *of* Attributivism, those who don't accept it, must.

§6.4.2 – Implausibly Ideal Actions

In this section I quickly respond to an objection that disagrees with certain claims I make. Namely, it disagrees with the plausibility of my examples, and that ideal agents could be motivated in the ways that I argued for. In particular this objection states that the acts my example agents engage in are just clearly absurd or unreasonable, and so not of the sort ideal agents would plausibly undertake. I will explain why that possibility does nothing for Attributivism.

[156] See again §3.2.1 for a more thorough explanation of these accounts.

I have argued that goodness *qua* one's kind does not guarantee any overriding motivations on the part of ideal agents. My examples included human agents who, as I've argued, might plausibly act in ways that are defective *qua human*. They might embrace the criminal lifestyle, torture others, disdain social groups, or even destroy their ability to enjoy traditionally human pleasures in the name of drinking gasoline. An objection to my arguments denies the plausibility of those scenarios for the following reason: taking such extremely defective actions is clearly absurd! Nobody's ideal self could countenance doing such things. Therefore goodness and defectiveness on the extreme ends at least *must* be authoritative.

I think that sentiment is a very intuitive one. But in response I have to ask why. On what grounds do we think taking such extremely defective actions is absurd? There might be many reasons. It might be because they're just obviously *wrong*. It might be because they are pathological behaviours, and so by definition *crazy*. It might be because, at the end of the day, we think it's just *better* to be human than to be whatever gasoline guzzling monstrosity we might otherwise be. So the saying goes, "it is better to be a human being dissatisfied than a pig satisfied" (Mill 1871, 14).

No arguments of that sort can count *in favour* of Attributivism though. Attributivism is built on the rejection of things being 'just better', of being just 'obviously wrong', or of meaningful *external* normative standards. That leaves the attributivists with a dilemma. They can make arguments like the above in order to reject the plausibility of taking extremely defective actions, but in that case they have abandoned the core tenet of Attributivism – that normativity derives solely from the standards of one's kind. That would be to refute, not redeem Attributivism. If, on the other hand, they maintain that commitment, they remain vulnerable to my objections, that goodness *qua* one's kind can plausibly fail to motivate one's ideal self, given the possibility of alternative kinds. So appealing to the 'clear absurdity' of extremely defective actions offers no respite.

§6.4.3 – Circumstantial Authority

In this section I turn to the final objection I'll consider. The gist is that my arguments in §6.2 and §6.3 didn't prove that attributive goodness is not authoritative. They proved that attributive goodness is *sometimes* not authoritative, under certain conditions – namely, when agents can take control of the kinds they belong to. The objection argues that the fact that there are certain conditions under which a property's influence isn't felt however doesn't, *prima facie*, mean that it doesn't ground normativity. It means only that sometimes those grounds are disabled, or otherwise circumvented. I argue here that even if that's true, the conditions under which attributive goodness fails to be authoritative *do* indicate that it doesn't ground normativity *fundamentally*, i.e., *alone*, *in itself*, and *that* is what we are interested in.

The short way of putting the objection here is this: *sometimes* attributive goodness is authoritative, and that's good enough. The slightly longer version is as follows. The circumstances I employ in the previous sections to demonstrate attributive goodness' lack of authority are generally unusual. Quite often one won't be able to instantiate new kinds, or instantiate the sort of multiple kinds that would justify disregard for the standards of those they belong to. If Tammy wants to drink gasoline, and she *doesn't* live in the outlandish futuristic scenario I stipulated, then she's out of luck. If her desire is conditional upon drinking gasoline being good for her kind, and she can't become such a kind, then the desire can't be satisfied. The fact that *in this case* drinking gasoline is unavoidably defective would motivate her ideal self not to drink. That means that it is only in some cases that attributive goodness fails to be authoritative.

The above seems to be a problem for my arguments, because we might not think that a fundamentally normative property needs to be authoritative in *every* circumstance. We might imagine a scenario wherein we take *pain* to be a fundamentally normative property. When an act has the property of inflicting pain upon us, we ought not to do it. Yet in that scenario there

will be circumstances whereby it will be permissible to take an action that *would* be painful, namely when we can make that action *stop* being painful. So the authority of pain is only circumstantial, present in situations where we don't have the ability to avoid it. Yet pain may still seem plausibly fundamentally normative, specifically because when we can't avoid it, we ought to be ruled by it. We might think attributive goodness works the same way. Even though there are *certain* cases wherein attributive goodness comes out non-authoritative, it's still fundamentally normative specifically because whenever we *can't* generate those circumstances, our ideal self would be guided by it.

There is an important difference between the case of pain and the case of attributive goodness that explains why the analogy doesn't work. Yes, it's not a requirement of a fundamentally normative property that its authority should *always* be felt. But I think it *is* a requirement of a fundamentally normative property that its authority should be felt *whenever it's present.* In the pain examples, we escape the authority of pain because we can *remove* pain from the equation, make it such that an act stops being painful. That's not the case with attributive goodness. In the cases laid out in §6.2 and §6.3, one has justification for ignoring the good of one's kind(s) even while still being a member of those kinds. *That* makes the problem.

Attributivism doesn't argue that one ought *sometimes* to act in the good of one's kind, or that one ought not to take defective actions *unless* one can get away with them. It argues that one ought to take only certain actions *if and because* one is a member of a kind, and the actions are good thereby. Attributive goodness is supposed be *non-derivatively* normative – normativity is supposed to be a consequence of its distribution *alone*. When we have attributive goodness, we are supposed to have normativity. That's why its failure in even extreme circumstances is a problem for Attributivism. If it's plausible that there is ever a time when

we can have the property *without* having normativity, then normativity isn't delivered by the property *alone*, and it's not fundamentally normative.

The above argument may not be entirely convincing, as it is popular to think of normative grounds as things that can be *enabled* or *disabled* by circumstances.[157] For example, imagine that I live in Canada, I have a modest income, and there is an airline strike. In this case the fact that there is an airline strike is a reason for me not to make plans to visit Japan in the near future. That fact is grounded, at least partially, in the fact that I live in Canada. That ground is *enabled* by the fact that I have only a modest income, and can't afford a private airplane – it's only because of my modest income that living in Canada becomes a reason for me not to plan to visit Japan. If normative grounds can be enabled and disabled like that though, then perhaps attributive goodness is the sort of ground that is only *enabled* by certain conditions, namely its being circumstantially impossible to instantiate other kinds.

I think the same argument I made in response to the pain example is relevant here. Even if, in the above example, living in Canada is a circumstantially normative property, nobody would think living in Canada is a *fundamentally* normative property. That's specifically because we can so often have the property 'living in Canada' *without* deriving anything normative from it. If *x*'s being normative is something that can be enabled or disabled in conjunction with certain circumstances, despite there being no change to the presence of *x* itself, then what we have is *not* that *x* itself is normative, but that *x* is normative *in conjunction* with something else. In the previous case, for example, having modest means makes living in Canada normatively relevant because *in those circumstances* it would be impossible or extremely unpleasant to travel to Japan, and, we might think, one ought not to plan on doing things that are impossible or extremely unpleasant. In that case, rather than being inherently normatively relevant, *x* – living in Canada – is *made* normatively relevant by some further

[157] See Bader 2016 for related arguments, particularly §3, pp. 31-39.

feature brought about by the conditions. It's reasonable then to think that living in Canada is not *itself* fundamentally normative, but that the grounds of normativity, at the very least, require the inclusion of those additional features.

In conclusion I think that the merely circumstantial nature of attributive goodness' authority is decisive evidence *against* its being fundamentally normative. For *x* to be fundamentally normative, authority must derive from the property *alone*. If we can have situations wherein we have the property but *not* authority, then the derivation does not follow, and the property is not normative in and of itself.[158]

[158] In essence I seem to be arguing that a property can't be fundamentally normative unless it is consistently normative. My claim is that, if *x* is supposed to be sufficient for normativity, then *if x then* normativity. If *x* only counts in favour of action in particular circumstances, then that is evidence that *x* isn't itself normative. I might rightly be interpreted as weighing in on the debate over particularism. Dancy, the particularist, argues that there is no necessary consistency when it comes to normativity – sometimes certain considerations just are normative, and sometimes they aren't. In some particular cases *x* counts in favour of acting, and that's all there is to it being normative (Dancy 2006). As such it might justly be said that Dancy disagrees with me here. On the other hand, Brad Hooker and Joseph Raz seem to support my view. Hooker argues that there is particular utility in theories of normativity that employ a consistent, underlying, unifying normative property (Hooker 2000, 19-22). He argues likewise that unified accounts make the best sense of 'overwhelming sensible' moral assertions (Hooker 2000b, 1-22). Raz, similarly, argues that normative reasons often undermine particular considerations, and that evaluative properties must extend beyond reasons to act in particular cases (Raz 2000). If I'm right about this analogy, then throw my argument on the pile against particularism.

§6.5 – WHAT NOW?

A lot has been done in this chapter. In §6.1 I elaborated on the **Substantive Aim Thesis** of the previous chapter, and argued that the best way for attributivists to accommodate **Authority** is through the *guise of the attributive good* thesis. That thesis argues that desires are essentially conditional upon their desiderata being attributively good for the desiring agent. I argued that that thesis is not *prima facie* implausible.

In §6.2 I argued that agents typically instantiate multiple kinds. I argued that that is a problem for Attributivism. Attributivism has historically made the mistake of treating agents as belonging to only one kind, or at least one kind with clear priority. I argued that they have no justification for that belief. I argued further that instantiating multiple kinds renders goodness *qua specific* kinds non-authoritative. It is the case that one can satisfy desires the desiderata of which are defective *qua* one kind, just as long as they are good *qua* another. Thus there are no overriding motivations on the part of one's ideal self that *must* derive from the goodness of a particular kind. Thus no specific attributive goodness is authoritative, and thus no specific attributive goodness is fundamentally normative.

In §6.3 I argued against a revised Attributivism, whereby goodness *qua* the *set* of one's kinds is taken to be authoritative. I argued that it is plausible for agents to come to instantiate *new* kinds. I argued too that it is reasonable for agents to act as regards the goods of those future kinds as opposed to their current kinds. Thus I argued that goodness *qua* the set of one's current kinds is also non-authoritative, and so not fundamentally normative.

§6.4 responded to a number of possible objections. I argued that attributivists can't deny my permissive view of kinds, that they can't rely on the implausibility of ideal agents actually taking extremely defective actions, and that they can't rely on circumstantial authority.

If I have been correct in all I have said, then I think it cannot be that attributive goodness *qua* one's kinds is fundamentally normative. Thus, Attributivism as we know it is false. In the

next chapter I will close by considering whether or not there is any way of *revising* Attributivism, or if it must be rejected all together. I think in fact that there *are* ways of revising Attributivism into something plausible, but that the result will be altogether unappealing.

7 – POST ATTRIBUTIVISM

I've argued that Attributivism as we know it fails as a theory of normativity. Yet I am not and should not be content to leave things there. If current accounts of Attributivism are wrong, what does that mean, going forward? What can we learn from Attributivism's failure? Perhaps more importantly what should *attributivists*, and those with attributivist sympathies, do now? If a theory has failed, it must either be modified, or relinquished in light of its failure. The focus of this chapter is on which of those courses now recommends itself.

If Attributivism can't work as it's currently understood, then perhaps it can be modified. That should be the first tactic for any attributivist, to seek to jettison problematic premises while retaining the core commitments of Attributivism. That strategy is the focus of §7.1. Therein I argue that Attributivism *can* be modified to produce some limited but genuine normative constraints. It can be so modified by widening the scope of normatively relevant kinds, from those that the agent actually instantiates to some wider subset of the set of all instantiable kinds. I will argue that the resultant *Global Attributivism* plausibly produces some range of authoritative regulation, at least as long as the conditional view of desire established in chapter 6 is accepted.

Even if Attributivism can be modified to produce some authoritative regulations, it remains an open question whether this can be done in a way that retains the benefits of the original theory. In §7.2 I argue that those prospects are dim. The best case scenario for Global Attributivism is implausibly permissive. Although it does manage to rule out some actions, the scope of that range is so limited as to appeal to no one, and so there is good reason to look elsewhere.

§7.3 concludes the chapter by discussing what to do if we are not content with modifying Attributivism. To some extent or another, that option involves relinquishing attributivist commitments. The question is, to what degree? §7.3 lays out two options. The first involves giving up on attributive goodness and any hopes for an attributive normative property, just as attributivists have given up on predicative normative properties. For the dedicated attributivist, that option gives up on normative realism entirely. The second option involves trying to learn from Attributivism's near successes. Although Attributivism fails at the extremes, in many non-fantastical cases its demands seem appropriate. If it is so *nearly* plausible, we might conclude that it is, if not correct, *on the right track*, and research further where that track leads. §7.4 concludes.

§7.1 – MODIFYING ATTRIBUTIVISM

Throughout this section I make the following case. Current attributivist accounts don't generate the normative constraints they promise, yet some *other* constraints *do* derive from the arguments developed in their favour. In light of them, a modification to Attributivism seems plausible. In §7.1.1 I explain how the cases I have developed in previous chapters suggest modifying Attributivism by widening its scope. In §7.1.2 I argue for just how wide that scope should be. In particular, I argue that rather than grounding normativity in goodness *qua* one's actual instantiated kinds, normativity may plausibly be grounded in goodness *qua* the members of the set of all kinds one could instantiate. §7.1.3 then provides examples of what sort of action regulation actually falls out of the modified theory.

§7.1.1 – New Attributivism and Attributivism Classic.

Current accounts of Attributivism are unsupportable, or so I've argued. Until now those accounts have been synonymous with Attributivism as a whole. Their defining feature, canonically, has been to propose grounding normativity in goodness *qua* the kind(s) one (currently/actually) belongs to. In light of the pending modification I'll call the view espoused by those accounts *Local Attributivism*, or **LA**. It says the following.

> **LA**. (i) Goodness *qua* one's actual, instantiated kind(s) is fundamentally normative. (ii) One ought to do only that which is good *qua* one's actual instantiated kind(s).

As I've argued, **LA** is false. Attributivism though has always maintained three claims. First, the general claim that goodness is always attributive. Second, the claim that attributive goodness is fundamentally normative, and third, that one's particular normative obligations are determined by the goodness of one's actual, instantiated kinds. Arguably however, the third

claim neither follows from the others, nor is it required for a view to be attributivist. As long as a view retains the first claim, and the claim that normativity is *in some way* grounded in attributive goodness, then I think we have an attributivist view. And so I think there is room within Attributivism for a view of the following sort. Call it *Global Attributivism*, or **GA**.

> **GA.** (i) For any agent *x*, goodness *qua* the members of the set of all kinds *x* does and can instantiate is fundamentally normative. (ii) One ought to *either* not Φ *or* belong to a kind at the time of Φing *qua* which Φing is good.

I think that **GA** is plausible and, moreover, is suggested as a modification by the very examples I employed against **LA**. In the subsequent section I'll explain how and why the circumstances of **LA**'s failure plausibly motivate the adoption of **GA**.

§7.1.2 – What Tammy Ought To Do, and Why.

In this section I'll elaborate on the motivation behind, and plausibility of, adopting **GA.** I'll also explain the sort of normative obligations **GA** actually entails – that is, what we ought to do if **GA** is right. In order to do that, I think it best to return to the examples of the previous chapter.

Take Tammy once again, who desires to drink gasoline. Chapter 6 demonstrated that even if drinking gasoline is bad *qua* Human, and even if Tammy is Human, and even if the *guise of the attributive-good* thesis is true, that Tammy's ideal self might nevertheless be intelligibly motivated to drink gasoline. That's because even if drinking gasoline can't be done in a way that's good *qua* Human, it may be done in a way that's good *qua* another kind Tammy belongs to, or even one she can arrange to belong to. So Tammy's ideal self might intelligibly be motivated to pursue goals that are not good *qua* any of her actually instantiated kinds. So

LA is not right, as goodness *qua* Tammy's actually instantiated kinds is not authoritative, and thus not fundamentally normative.

Even, however, if goodness *qua* Tammy's instantiated kinds doesn't determine what she ought to do in the above scenario, that's not the same as there being *nothing* she ought to do therein. In fact, the details of the case imply that there's something very specific she ought to do, or, more precisely, ought *not* to do. What makes it permissible for Tammy to act in ways that are contrary to the goodness of her instantiated kinds is that she can and will come to instantiate *other* kinds, *qua* which her actions will be good. What that implies is that *if* there are no other kinds she can instantiate, *then* contradicting her instantiated kind standards would be impermissible. As such it seems true that Tammy *ought not* to drink gasoline *unless* she can (and does) come to instantiate a kind *qua* which doing so is good.

The appearance of such an ought motivates a move towards **GA**, since what seems to be determining the permissibility or impermissibility of Tammy's actions is whether or not a kind is available to instantiate that will cause her actions to come out good, if she instantiates it. Thus we might think that it's that range of kinds – possibly instantiable kinds – and goodness among those kinds that grounds Tammy's normative obligations. Moreover, that seems plausible. If we continue to accept the *guise of the attributive-good thesis* – i.e., that agents implicitly desire things only on the condition that they will be attributively good for them – then that gives us reason to think that the motivations of ideal agents are bound by facts about goodness *qua* the range of instantiable kinds. For an action can only come out attributively good if there exists a kind *qua* which it *is* good. Thus whether or not an ideal agent will be motivated to take that action depends on the existence of such a kind. Thus facts about goodness *qua* the range of instantiable kinds entail overriding motivations on the part of ideal agents, and are thus authoritative, and so seem like the sort of thing which might serve as grounds of normativity.

So the case of Tammy and her gasoline drinking gives attributivists reason to widen their scope, from goodness *qua* one's currently instantiated kinds, to goodness amongst a wider range. What's more, Tammy's case gives us an idea of just what sort of regulation that entails. Since Tammy's ideal self will only drink gasoline *if* she also comes to instantiate a kind such that drinking gasoline comes out *good*, we can derive two sorts of authoritative regulations. (i) If there *were* no kind *qua* which drinking gasoline were good, Tammy's ideal self would not drink gasoline. Thus any agent ought *not* Φ insofar as there is no kind the agent can instantiate, *qua* which Φing is good. (ii) The only case in which Tammy's ideal self chooses to drink gasoline is one in which she also comes to instantiate a kind *qua* which doing so is good. She would not do so otherwise. Thus, one ought *either* to do what is required to instantiate a kind *qua* which Φing is good, *or* not Φ. What specifically that amounts to is determined by which kinds are available to instantiate, *qua* which Φing is good.

The above, then, explains both (i) why it seems plausible for attributivists to revise their theory from **LA** to **GA**, and (ii) the sort of obligations that derive from **GA**.

§7.1.3 – Global Attributivism, Authority and Regulation

I've argued that attributivists have reason, inspired by the cases of the previous chapter, to revise their theory. I've argued that there's a specific and plausible theory they might adopt. It is, as a reminder, the following:

> **GA.** (i) For any agent *x*, goodness *qua* the members of the set of all kinds *x* does and can instantiate is fundamentally normative. (ii) One ought *either* not Φ *or* belong to a kind at the time of Φing *qua* which Φing is good.

My view is that **GA** is a plausible attributivist theory. I think that it satisfies **Authority** and **Regulation**, and thus meets the requirements for fundamental normativity set out in chapter 4. In the rest of this section, I'll explain first how it satisfies **Authority**, and then take a closer look at what regulation it actually provides.

GA satisfies **Authority** provided that we continue to accept the *guise of the attributive-good* thesis. As a reminder, what **Authority** stipulates is that for any *x* to be fundamentally normative, ideal agents must derive overriding motivations from facts about *x*. What that means for **GA** is that ideal agents must derive overriding motivations from facts about goodness *qua* members of the set of all kinds they can instantiate. That's plausible, as follows.

1. According to the *guise of the attributive-good* thesis, any agent *x* desires to Φ only on the condition that Φing will be good *qua* a kind *x* belongs to at the time of Φing.
2. From 1, no desire to Φ can be satisfied if Φing will not be good *qua* a kind *x* belongs to at the time of Φing.
3. From 1 and 2, the ideal *x* will not be motivated to Φ, if Φing won't or can't be good *qua* a kind *x* will belong to at the time of Φing.
4. If Φing is bad *qua every* kind instantiable by *x*, then Φing can't be good *qua* a kind *x* belongs to at the time of Φing.
5. From 3 and 4, the ideal *x* will not be motivated to Φ if Φing is bad for *every* kind *x* can instantiate.

Two sorts of overriding motivations are thus determined by facts about goodness *qua* the set of kinds instantiable by agents. Since ideal agents desire – and are thus motivated to take – actions only insofar as taking those actions will be good for a kind they belong to, the following will be true. Any ideal agent will be (i) overridingly motivated *only* to take actions when the agent will *also* belong to a kind *qua* which that action is good, thus (ii) overridingly motivated *not* to take actions that are bad *qua every* instantiable kind. Thus, any ideal agent

will be overridingly motivated *either* to *not* Φ *or* to belong to a kind at the time of Φing *qua* which Φing is good. And so the ought stipulated in **GA** comes out authoritative.

So I accept that **GA** satisfies **Authority**. If it also satisfies **Regulation**, then if there were no other normative grounds other than those stipulated in **GA**, there would be genuine normativity. So now let's look at **Regulation**. What does **GA** actually rule in and out, and when? The answer seems to be: not a lot.

GA has the potential to regulate action in two ways. At the minimum, it rules out *combinations* of actions. It may also, however, rule out specific actions with varying degrees of strength, depending on certain conditions. I'll elaborate now on these two varieties of regulation.

Prima facie **GA** doesn't regulate over any specific actions. That is to say, it doesn't require that one does or does not do any single specific action. What **GA** specifies is that one ought to *either* not Φ *or* belong to a kind at the time of Φing *qua* which Φing is good. Notably, that says neither that one actually ought not to Φ, nor that one actually ought to belong to a kind of the specified sort. Only a disjunctive normative claim is true: one ought to do *either a or b*, but both *a* and *b* remain permissible options in isolation. **GA** does rule out something though, which is the *conjunction* of the two, i.e., the *combined* action of *both* Φing *and* not belonging to any kind *qua* which Φing is good. If it's the case that one ought to do one thing or another, after all, then doing neither would be failing to do what one ought to do.[159]

[159] The truth of that can be demonstrated more concretely via rules of negation and disjunction, or what's known as a DeMorgan equivalence. As a rule, (*a*) is equivalent to ~~(*a*), or not not-(*a*). Likewise the disjunction (*a* v *b*) is equivalent to ~~(*a* v *b*). The negation of a disjunction however is equivalent to the conjunction of the negation of the disjuncts; ~(*a* v *b*) is equivalent to (~a ^ ~b). Thus (*a* v *b*) equates to ~~(*a* v *b*) which equates to ~(~*a* ^ ~*b*). From the very same chain of logic, we can see that **GA**, which states that one ought to (~Φ v (belong to a kind at the time of Φing *qua* which Φing is good)), equates to the imperative that one ought to ~(~~Φ ^ ~(belong to a kind at the time of Φing *qua* which Φing is good)). In other words, one ought to not *both* Φ *and* not belong to a kind at the time of Φing *qua* which Φing is good.

So minimally, **GA** rules out the combination of Φing while also belonging exclusively to a kind or set of kinds *qua* which Φing is not good. That though is a very permissive form of regulation. Since no single act is ever *itself* ruled out, but only ruled out in conjunction with not instantiating an appropriate kind, *any* single act remains permissible, just as long as one has the capacity to instantiate said kind. Much like under the purported normativity of instrumental rationality, everything remains 'on the table', so to speak, for a sufficiently resourced agent.

GA has the potential to rule out particular actions however if certain conditions obtain. Namely, it rules out specific actions if it is the case that there is no kind available to be instantiated by the given agent according to which that action is good. Since **GA** stipulates that one ought either to ~Φ or belong to a kind *qua* which Φing is good, if the latter case is unable to be fulfilled, then ~Φing is the only remaining option, and so Φing is ruled out. This sort of regulation too breaks down into two categories however: actions may be ruled out in this way either circumstantially, or categorically, depending on *why* the agent is unable to instantiate an appropriate kind.

In one case, the agent merely lacks the resources to instantiate a kind *qua* which Φing is good. In which case, Φing is ruled out *circumstantially*, as a matter of the agent's circumstances. It's very plausible that this will often happen. Many kinds that might be theoretically instantiable may nevertheless not be instantiable *by* me, if I lack the resources to instantiate them. If, unlike Tammy, I don't know a mad-scientist and never will, then a wide range of kinds are unavailable to me, namely those which would require his help to instantiate. Thus in any case where I ought either to ~Φ *or* belong to such a kind, I would be left with only one option, to ~Φ, in virtue of my circumstantial inability to bring about the latter of the disjuncts. Thus certain actions may be circumstantially ruled out by **GA**.

GA rules out actions categorically in the case that there are actions for which there are no kinds *whatsoever qua* which the actions are good. That is to say, it's not just the case that the given agent is circumstantially unable to instantiate an otherwise instantiable kind *qua* which Φing is good, but that there is no such thing as a kind *qua* which Φing is good. In that case, for any agent in any circumstance, the only viable option under **GA** is to ~Φ, given the impossibility of belonging to a kind *qua* which Φing is good.

The above seem theoretically possible. In practice it is difficult to think of what sort of behaviour could *never* be considered good *qua any* kind. I can think of only suicidal or otherwise self-destructive behaviours as potential candidates. Attributivist kinds are after all, as a reminder, determined in response to the existence of reciprocally dependent structures of operations. These are systems within which two or more behaviours or dispositions interact, reinforce, and require one another, producing together a result that could not be effected or maintained otherwise. Self-destructive behaviours seem, *prima facie*, inimical to this sort of structure, aimed at tearing down rather than reinforcing other behaviours. At least, *non-instrumentally* self-destructive behaviours. Although there's room to quibble on the definitions of self-destruction vs. self-*sacrifice*, on the face of it at least suicide and self-harm can be functional when undertaken in the service of some other virtue – parents sacrificing themselves for their children, for example. Yet self-destruction for its own sake seems like a different beast. *That* seems incompatible with the maintenance of some larger system. So we have reason to think at least that non-instrumental self-destructive behaviours may be ruled out by **GA** in every case, regardless of the motives, resources or circumstances of any agent.

I've now, across this section, explained both what regulation **GA** provides, and why its standard is authoritative. I conclude that I've proposed a viable alternative formulation of Attributivism. **GA** promises to be able to ground some authoritative regulations in a way Local Attributivism never could. Adopting **GA** allows the attributivist to retain the plausibility of

their claim that normativity is cached out entirely in terms attributive goodness. Whatever Tammy, or any agent, does, there is a range of appropriate and inappropriate ways to do it, based on facts about what is attributively good across a range of kinds. One really ought not to Φ without *also* instantiating a member of the set of kinds *qua* which Φing is good. Moreover, whatever capabilities Tammy Baritone possesses, whatever mad-scientists she knows, she is at least normatively obliged not to self-harm non-instrumentally. A less capable, less connected, alternate world counterpart of Tammy, Tony Alto, has significantly fewer options. Having neither the capabilities to flourish in a life of crime, nor the connections to transcend his Human limitations, he is obliged, on the basis of his Ideal Self's dispositions, to pursue neither murder nor drinking gasoline nor living wild in the woods. If this is a dark world, and the loss of his dreams turns Tony to thoughts of self-harm or suicide, he's obliged to refrain even from that. These are genuine normative constraints.

§7.2 – Global Attributivism, Merits and Benefits

I think **GA** is coherent, and that attributive goodness within the set of kinds instantiable by an agent has the potential to serve as a fundamentally normative property set. I have no knock down arguments against it. Yet when I proposed reformulation as a viable option for the attributivist, I presupposed two conditions. The question facing attributivists, I said, was whether Attributivism could be modified into something that (i) 'works' to ground normativity, and (ii) retains the benefits of the original theory. **GA** satisfies (i). I question whether it satisfies (ii). The answer to that depends on what we take the benefits of Attributivism to be. I think that largely the answer is no. If so, we have reason to hesitate about accepting **GA**.

We must ask then what are traditionally seen as the benefits of Attributivism. Why do attributivists promote it? There seem to be at least four reasons. (i) Its ontological commitments are fewer and more reasonable than alternatives. (ii) Its regulations are not dependent on desire. (iii) It's significantly action guiding. (iv) It accommodates popular and strongly held intuitions. **LA** was taken to provide all of those benefits. I think that **GA** provides (i) and (ii), but to a lesser and unsatisfying degree. I think it fails to provide (iii) and (iv). I'll look now at each in turn.

*§7.2.1 – Is **GA** Parsimonious?*

The primary motivation behind Attributivism is, of course, that it promises to make sense of normativity without committing to any theoretically suspicious ideas about predicative goods and their grounds. Rather than commit to any ontologically strange properties, Attributivism offers to ground normativity only in facts about what things *are*. That's a tantalising promise for anyone attracted to naturalism or reductivism. In that, and only that, can **GA** be said to succeed to any significant degree.

GA has widened the range of kinds that Attributivism appeals to. There's nothing particularly objectionable about that. In widening the range it hasn't made any new *ontological* commitments. It remains committed only to positing the existence of kinds and standards of goodness according to them. Nor is there anything ontologically objectionable about positing the existence of a set of instantiable kinds. So **GA** is just as ontologically parsimonious as **LA** was. **LA** argued that normativity requires only reference to the goodness of one's kind and how it stands to influence one's desires. **GA** does the same, only across a wider range.

So **GA** remains pleasingly parsimonious. Even still though, I think it is *less so* than was **LA**. The tantalising prospect of grounding normativity in a straightforward analysis of what any given subject *is* is, I think, a *significant* part of Attributivism's allure. It certainly tempts me. Yet the degree to which **GA** now does that is suspect. Once Attributivism proposed that all normativity was of a common kind – to say that one ought to do good things was taken to mean the same thing as when we say that a tree ought to put out leaves, or that a toaster ought to toast. In all cases, we were promised, it would be possible to move from a straightforward understanding of what a subject *is* to what it *ought* to do. Now however the normativity of *agents*, of beings that can exercise rationality and freedom of choice, seems of a rather *different* kind. What an *agent* ought to do is determined not just by what she *is*, but all the things she *could* be. An understanding of what kind of thing the agent *is* no longer binds her actions; what we *ought* to do *isn't* constrained in the same way that a bee *ought* to buzz or a cow *ought* to graze. Even if **GA** remains ontologically parsimonious, it seems to have suffered in *conceptual* parsimony. So while **GA** does indeed accomplish the central attributivist aim, reducing the normative to a set of non-predicative, naturalistic properties, the appeal of that reduction has been reduced.

*§7.2.2 – Is **GA** Desire Independent?*

To repeat the conclusion of numerous chapters and sections, the relationship between desire and Attributivism is an intimate one. Geach recognised that a normative property must be a motivating one, and suggested that an intimate relationship with the act of desiring was how Attributive Goodness meets that demand. Anscombe did likewise, making it a fact about "the phenomenon of *wanting*" (Anscombe 2005c, 178) that one cannot help but desire, and be moved by, some facts about the goodness of their kind. Yet she did so in the course of *disagreeing* with the view that desires, or 'passions', are reason-giving, or that desires are inherently motivating in a way that beliefs aren't.

No attributivist disagrees with Geach or Anscombe; the normative status of Attributive Goodness *is* tied up with facts about one's motivational set.[160] It's a necessary connection with desire that makes the Attributive Goodness of kinds more than just another system of hypothetical norms like etiquette or the rules of chess, which matter only insofar as we accept them. At the same time, every attributivist agrees that reasons do not depend alone on any "individual peculiarities of desire" (Geach 1956, 40). Foot recognized Hume's 'practicality requirement', that normative judgments must serve to guide action, and that thus the cognitive and the conative must interact. She was firm in the belief that conative attitudes were part of practical rationality. She was firm too however that they were not the *grounds* of practical rationality, but rather that the question "why should I do *x*" is sufficiently answered by facts about the needs and goods of one's kind. For Foot, conative attitudes arise from cognitive beliefs – judgments about normativity create desires, they don't depend on them.[161]

Thomson also seems to accept that reasons – or beliefs about reasons – must be capable of explaining action, and that they can only do so in the presence of some *want* – or the belief

[160] The truth of that premise is even analytical – if a subject exhibited *none* of a specific set of dispositions or motivations, they would not *be* the specific Kind that is described at least partially in terms of those dispositions. The particulars of one's Attributive Goodness depends at least partially on elements of one's motivational set.
[161] See Foot 2001, particularly pp.17-23.

that doing *x* would satisfy some want of the agent. Yet she believes too "that reasons-for are facts [and that] no desire is a reason to do anything" (Thomson 2008, 163). Desires for Thomson are *background conditions*, not reasons themselves.[162] Like Foot, Thomson is a representative arch-attributivist, maintaining both that desires play an important role in the establishment of reasons, but that they are nonetheless *not* normative grounds themselves.

In short, attributivists have classically maintained that ought doesn't depend upon desire; rather the force of an ought is, at most, delivered *through* desire. What one ought to do is *not* classically believed to fluctuate along with one's desires. Rather, one's desires are expected to conform to what one ought to do.

GA upholds that commitment somewhat. *Some* oughts remain desire independent under **GA**. Specifically, that one ought *either* to Φ *or* belong to a kind *qua* which Φing is good, and that one ought not to Φ when one *can't* instantiate a kind *qua* which Φing is good. *Those* oughts plausibly constrain desire, and don't vary with any peculiarities of the same. Whether or not one actually desires to Φ without instantiating a kind *qua* which Φing is good, one ought not to.

Yet **GA** is still less satisfying in this regard than **LA** was. One thing that often motivates the pursuit of desire-independent normative theories is the belief that doing what one ought to do isn't equivalent to satisfying one's desires. Likewise motivating is the belief that there are certain things we ought and ought not *desire*. Philosophers who are motivated by desire independence are motivated to reject Hume's claim that "'[t]is not contrary to reason to reason to prefer the destruction of the whole world to the scratching of my finger" (Hume 1989, II:iii:3 416). **LA** satisfied those motivations – a very specific set of actions are prescribed under **LA**, those that are good *qua* a *specific* way of being, and many things ought not to be done no matter *how* well they satisfied our desires. **GA** though is, by and large, extensionally equivalent to a

[162] See also Thomson 2008, 44-53 and 160-162.

theory of normativity that is grounded in desire. Save for the compound desire to Φ *and* not belong to a kind for which Φing is good, *any* desire is permissible to act upon, and any action which serves to satisfy a desire is one that is, as Hume put it, 'not contrary to reason'.

Under **GA,** then, a suitably resourced agent is left to do whatever she desires, without any further guidance. So **GA** constrains the desires of agents in a less satisfying way than **LA**, and in a way that is apt to seem overly permissive.

*§7.2.3 – Is **GA** Significantly Action-Guiding?*

One appeal of **LA** was that it gave attributivists the ability to normatively evaluate an exceedingly wide range of actions. A vast number of possible actions could be absolutely ruled out for a Human agent, for example, in virtue of their being defective *qua* Human. According to **LA**, if one is Human, a social and rational animal, one should not murder, rape, pillage, lie, steal, ingest toxins, embrace disease, abandon society, fail to keep promises, cheat on one's spouse, overeat, under-eat, act without thinking, believe obvious falsehoods, commit arson, ignore one's education, or spend too much time writing lists. Likewise, *good* acts, of the sort we *ought* to do, had the potential to be ranked against one another, in deference to *how* good they were, *how* essential they were in the functioning of the Human Kind. Perhaps one ought to protect one's family over saving a stranger's life, and save a stranger's life over keeping promises, and keep promises over donating to charity, and donate to charity over reflecting philosophically, and reflect philosophically over visiting the new Syrian café. *Most* actions stand to be good or not *qua* one's actual, instantiated kind, to at least some degree, and so **LA** promised a normative ground that would make sense of the thought that there are *many* things we often ought and ought not to do, both grand and mundane.

The regulation **GA** provides, on the other hand, can be vanishingly small. As a reminder, **GA** rules out *specific* actions in two sorts of cases. (i) Actions are circumstantially

ruled out, when *due to circumstance* agents are unable to instantiate a kind *qua* which the action is good. (ii) Actions are categorically ruled out, when there is no instantiable kind at all *qua* which the action is good. As regards (i), the actions that are ruled out depend heavily upon the resources of the individual. A poorly positioned agent might be confronted with a wide range of regulation. If they have no money, no skills, no connections, in short no *ability* to come to instantiate other kinds, there may be much they ought not to do. But these restrictions vanish as they gain in those resources, and as the resources of society allow agents to take greater and greater control over what kind of being they are. One who is rich and connected enough to successfully maintain an existence as a crime lord may come to instantiate such a kind, and become thus entitled to murder, torture and destabilise society. Theoretically, **GA** is compatible with a sufficiently resourced agent being subject to no circumstantial regulations at all. That leaves **GA**'s regulation entirely dependent upon (ii), and that which can be distilled from the idea of universally defective actions.

Unfortunately for **GA**'s prospects, I'm sceptical that there *are* any actions that are universally defective. I argued in §7.1.4 on behalf of the attributivist that non-instrumentally self-destructive kinds seem like a theoretical impossibility. I suspect that's not really the case. Consider the classic spy-thriller missive, which proclaims "this message will self-destruct". As a member of a certain class of espionage tools, it's absolutely imperative that those oracular messages should self-destruct as part of the function of their kind – they really ought to, and if they didn't they would be defective. Further easily recognised examples are bombs. These are kinds of things for which self-destruction is not just an action that *can* be undertaken virtuously, but an act that is built fundamentally into their function. It's an essential part of the system of interrelated operations we take to identify a distinct kind of object. The bomb's fuse burns, or its chemicals mix, or its timer counts down, it self-destructs, and the fact that it will self-destruct is what prompts and allows for all the previous actions in the first place.

I suspect that part of what motivates the thought that self-destructive kinds are somehow theoretically incoherent is that self-destruction is antithetical to a kind's *persistence*, or at least that of its members. Self-destructive behaviours threaten to bring down the system of which they are a part, and so run contrary to that system's indefinite maintenance. But I don't see why we should ever think that to represent a coherent kind a system must aim at *indefinite* persistence. Many kinds recognise their own eventual end in their operations. It's common knowledge that many insects kill themselves as part of their mating processes. Of course those deaths are instrumental towards the end of genetic reproduction, but I can think of no particularly good argument why terminal ends shouldn't be a coherent type of ends in themselves.[163]

Of course a reasonable defence on behalf of the incoherence of self-destructive kinds is to point out that even in the cases of bombs and self-destructive messages, such self-destruction is again instrumental. These devices are artefacts, and self-destruction is only part of their behaviours because Humans design and maintain them to do so instrumentally, towards other ends of external destruction, entertainment or espionage. Moreover the various operations in bombs that lead to their destruction can only be said to be *aimed* at such destruction because of the responses of *Humans* to a failure to explode, not because of anything in the system itself.

I think a simple thought experiment defuses the above. Actual technological advances are making it easier and easier to conceive of something like a bomb that is not beholden to Human artifice for its maintenance and aims. It's fairly easy to imagine a 'Smart Bomb', a bomb into the system of which an independent, monitoring intelligence is built. Such a system

[163] Spinoza thought it self-evident that no thing could contain essentially elements of its own destruction, and that everything thus strives essentially for its own perseverance. See principally Spinoza's *Ethics*, IIIP4, IIIP5 and IIIP6, (Spinoza 1996). I take it though that Spinoza meant this analytically. I don't believe he would argue that there can be no systems within which self-destruction can play a part, only that any such system must be made of at least two distinct things. If he *did* mean to deny the coherence of such systems, I suppose I disagree with him.

which cared not for when or why it explodes, and made changes to its own operations in response to a failure to do so, should very plausibly be seen as holding self-destruction as an end in itself. I think such a creature is entirely possible. I think thus that self-destructive kinds are entirely theoretically coherent, and that in fact there are no actions that are defective *qua every* possible instantiable kinds.

The authoritative regulation offered by **GA** is far from robust. It is, in fact, vanishingly minimalistic. Its most plausible range of regulations are tied intimately to, and vary dramatically with, the resources of a given agent. Rather than providing an overarching and holistic standard by which to evaluate a wide range of actions, **GA** suggests essentially that one ought not to do what one doesn't have the resources to do in a particular way. A suitably resourced agent in this picture is entitled to pursue any aim whatsoever. The very best case for **GA** assumes that I am wrong in my arguments above, and that functionally self-destructive kinds are theoretically incoherent, and thus that self-destruction is unable to ever be good. Even in that case, **GA** is left with one firm rule, 'do not pursue non-instrumental self-destruction'. While not, perhaps, an unreasonable rule, this one proscription doesn't promise significant guidance in our everyday lives. Another advantage of Classic Attributivism has been surrendered.

*§7.2.4 – Is **GA** Intuitive?*

It's not uncommon to think that the merits of a proposed theory of the normative depend at least to some extent on how well it satisfies certain common and deeply felt intuitions. Any theory of the normative that has as a consequence the permissibility of cannibalising live babies or torturing the elderly is a theory about which we should be sceptical. Does **GA** satisfy our common normative intuitions then? No. Of course not. Everything I've now said in the preceding subsections can stand in evidence there. Essentially nothing intuitively profane is

definitively ruled out. *Many* people might be in a position that allows them to justifiably murder or torture, if not now then in the future as the potential resources of mankind develop. Likewise many things we think are intuitively good – keeping promises, aiding the needy, being healthy in body and mind – **GA** sees as potentially irrelevant. This is a problem for anyone who thinks those sort of intuitions ought to be accommodated. And, as it turns out, attributivists typically do.

I needn't say much to prove that assertion. Attributivists readily do so for me. Anscombe advocated for the elements of Attributivism under the assumption that Attributivism would give the proper, normative role to Human flourishing. Such flourishing was understood to be in line with things like keeping promises, justice, and avoiding pain, hunger, destitution and social ostracism (Anscombe 2005c, 186-193). In fact, Anscombe decried alternatives precisely because they failed to produce norms of the sort we would call 'just' (Ibidem).

Likewise Hursthouse and Foot, as virtue theorists and Neo-Aristotelians both, predictably endorse Attributivism specifically because they take it to entail and explain the normative force of canonical, Aristotelian, social and rational 'virtues'. Hursthouse takes Attributivism to justify 'the virtues' because she takes all organisms to share the ends of survival, reproduction, freedom from pain and participation in pleasure, and maintenance of social order (Hursthouse 1999, 197-202). In Humans these ends are supposed to endorse behaviours like honesty, justice, loyalty, mental and physical health, etc. Foot largely agrees. Although she insists that she wants not to 'smuggle in' any preconceptions about what will count as 'good' in the Human lifeform, she also admits that she began her project taking the common 'vices' to be forms of natural defect (Foot 2001, 36-37). She is clear too that the plausibility of transitioning from the normativity of Attributive Goodness in plants and animals to the same in Humans depends on whether or not it can capture the "special subject of goodness of the will" which has to do with "the choice of lives, the education of children, or

with decisions of social policy." Her argument is that "human strengths and weaknesses, and even virtues and vices, are to be identified by reference to … 'biological' cycles" (Foot 2001, 38-41).[164]

It was a conceptual starting point for the philosophers above that the normative status of 'the virtues' could be explained by Attributivism. I think it is very unlikely that any of them would have been compelled to develop their theories had they thought that Attributivism would see the Human lifeform and its 'virtues' as apt to be rendered normatively inert. That though is precisely what **GA** accomplishes. A human agent under **GA** can look to the future and envision, and act to bring about, a future in which the human virtues play no part, and would do nothing wrong in doing so. **GA** accommodates almost no popular intuitions, the likes of which Foot, Hursthouse, Anscombe, even Geach make so much noise about.[165] So **GA** fails, once again, to accommodate some of the most basic reasons attributivists were drawn to Attributivism in the first place.

§7.2.5 – Unappealing Attributivism

The question to ask, I have said, is whether or not Attributivism can be modified into (i) a theory that works to ground normativity and (ii) a theory which retains the benefits of the original theory. Those benefits are the reasons Attributivism has been promoted and accepted in the first place. I think now I've shown that **GA**, the only promising modification of Attributivism, does *not* do (ii). It may do (i), and certainly doing so goes some way towards (ii); (i) is *part* of Attributivism's canonical benefits. Yet **GA** fails much more than it succeeds. **GA** is generally extensionally equivalent to desire-based theories of normativity, and its range of regulation is implausibly limited. It offers little to no serious guidance in everyday life, and

[164] See also all of her chapter 3.
[165] Geach commented that "Men need virtues as bees need stings" (Geach 1977, 15). A telling comment as to his view of the importance of the virtues in Human life, although I don't make much of here. His quote is compatible with a picture wherein the virtues are nonetheless non-authoritative.

satisfies almost no intuitions most ethicists, most agents and most *attributivists* hold. It has, in the end, very little to recommend it.

I don't deny that **GA** might represent a plausible, minimal normative ground. Someone could insist upon **GA** and only **GA**, and insist thereby that genuine, non-hypothetical normativity exists. But why? **GA** makes nobody happy. There are vanishingly few reasons to promote it, and an increasingly large number of reasons to look elsewhere.

§7.3 – ABANDONING ATTRIBUTIVISM

I think the conclusion to be drawn from all the preceding is that Attributivism in all its forms is a failure. Local Attributivism fails outright. Global Attributivism however is almost entirely bereft of the merits that motivated the original theory. It's possible to disagree; the failure of **GA** rests at least partially on intuitions that can be contested. But I think there are few reasons to champion it, and I know of no philosopher whose work would indicate a willingness to do so. Accepting its inadequacy then, this section asks as a parting query what the next step for attributivists is. §7.3.1 paints a pessimistic picture, that trying to retain some place for Attributivism in a wider ethical landscape is pointless. §7.3.2 touches briefly on whether or not the attributivist must thereby abandon normative realism entirely, and responds optimistically. The near successes of Attributivism suggest a third, alternate direction for research.

§7.3.1 – No Place for Attributivism

What should we do about Attributivism if **LA** fails and **GA** is so profoundly unsatisfying? If modification has failed to make Attributivism any more plausible, then rejection seems the appropriate course. Yet the attributivist might want to avoid abandoning their commitments outright. At the least, the plausibility of the *guise of the attributive-good* thesis makes it seem like there ought to be *some* place for Attributivism in our considerations. One thought suggests that perhaps Attributivism can be saved if it is taken as one *part* of a wider community of fundamentally normative grounds. Specifically, as most of the flaws of **GA** stem from its permissiveness in one way or another, what might seem a promising move is to *pair* **GA** with some additional normative ground, which could further specify what agents ought *to* do within the range **GA** makes permissible.

There are at least two problems with that proposal. The first is straight-forward. Perhaps the most significant motivation behind the development of Attributivism was the supposed unintelligibility of non-attributivist grounds for normativity. If we are now willing to *accept* additional, non-attributive grounds to supplement **GA**, then we've already given up the game. This tactic defeats the very motivation behind the adoption of **GA** in the first place.

Even if we think there remains reason to adopt both **GA** *and* some further, non-attributivist ground however, that strategy falls prey to a second problem. It either renders **GA** redundant, or trivialises it.

Imagine that we've posited a second fundamentally normative property *P*, which is supposed to supplement **GA**. From *P*, we derive a set of additional standards one supposedly ought to follow. Now we will always have two cases. Either (i) *P* and **GA** say we ought to do the same thing, they are *co-extensive*, or (ii) *P* and **GA** each say we ought to do a *different* thing, they *conflict*.

Take case (i). In this case, we imagine that *P* and **GA** never conflict, they never entail anything different about what one ought to do. If so, then **GA** is made redundant. If *P* and **GA** are co-extensive, then we have no reason to hold on to **GA** at all; *P* already does what **GA** provides. In that case we haven't preserved a place for **GA**, we've merely replaced it with something else.

So let's image that case (ii) applies instead. At least sometimes, **GA** and *P* conflict, and say we ought to do different things. The problem for **GA** here is that the very reason we proposed *P* to begin with was due to the fact that **GA** was unsatisfying. If, however, we were already motivated, when confronted with **GA** alone, to develop an account based on *P*, there's no reason to think we would ever privilege **GA** over *P* in cases of conflict. Since *P* in this case is what we apply to in order to generate intuitive normative conclusions, if the choice is

between what **GA** stipulates – which we've already found disagreeable – and what *P* stipulates – which, by stipulation, we prefer – **GA** might not as well be considered at all.

Thus **GA**, when paired with another normative property, is either redundant or irrelevant. I can't see, therefore, an easy way for Attributivism to retain a meaningful normative presence in conjunction with another prescriptive normative ground. That's not to say of course that there isn't one, only that insofar as I can tell, the prospects are slim. I won't consider them further.

§7.3.2 – Renouncing or Re-evaluating Normative Realism.

With finding a place for **GA** within a wider normative landscape unlikely, the attributivist is, I think, *finally* out of options. So now I close by asking what this means for those who have hitched themselves to the attributivist wagon. One obvious suggestion is that in abandoning Attributivism such philosophers should renounce normative realism, or at least cognitivism, entirely. That seems obvious, because so much of Attributivism's initial motivation stemmed from a firm denial of the plausibility of alternative, particularly predicative, cognitivist theories of the normative. The push to develop attributive goodness from a mere, commonly acknowledged, descriptive sense of goodness to a full blown normative property came about specifically because philosophers like Geach and Anscombe took it as evident that goodness was essentially attributive, and so that no intelligible claims about goodness could be entailed by non-attributive properties. Non-attributive properties, the attributivist has long maintained, cannot ground the central concern of ethics and so any theory which relies on them is doomed to fail.[166] Now that Attributivism has also been shown to fail, it seems like attributivists have

[166] Geach of course was adamant that goodness was essentially attributive, and also that goodness was the core concept in ethics, "deliberately ignoring the supposed distinction between the Right and the Good" and, like Aquinas, seeing the former cached out in terms of the latter (Geach 1956, 41). See also Ancombe's tirade against contemporaries in *Modern Moral Philosophy* for a succinct review of her views about non-cognitive, non-attributive theories (Anscombe 2005c, 170-72).

left themselves nowhere to go, save to renounce either all their anti-predicative judgments, or the entire exercise.

I don't have much to say on renouncing normativity entirely, other than that I don't expect it to be tremendously appealing to a group of people whose intuitions have thus far had them endorse cognitivist realism. Perhaps their familiarity with having rejected predicative theories will leave them ready to do the same again here. Regardless, if outright rejection is the path that recommends itself, then I imagine attributivists should look to the wealth of quasi-realist or Error Theory literature that exists. As this is not the place to launch an extended defence of or objection to error theories however, I can't comment on those prospects. I don't however expect outright rejection of normative realism to be a popular choice.[167]

Granted, neither do I expect attributivists to easily shed all their previous intuitions as to the prospects of alternative realist theories. That too would be an unpopular move. Of course, whether or not these moves are unpopular only matters if there is some alternative. I think perhaps there is. An alternative tactic is to employ what has been learned from the near successes of Attributivism in the development of a new, quasi-attributive theory. A theory, that is, which caches out the normativity of attributive goodness contingently in terms of how its instantiation coincides with the instantiation of some further property which (i) fundamentally grounds normativity, (ii) is instantiated *through* the instantiation of attributive goodness, yet (iii) is neither exclusively nor exhaustively instantiated through the instantiation of attributive goodness. I'll end this chapter with a very brief sketch of what I think such a theory might look like.

[167] Nor do I think there are any currently well regarded theories to which the attributivist could easily switch their allegiance. Constitutivism might seem like a reasonable choice – I have already enumerated a number of similarities and intersections between Consitutivist and attributivist thought. I suspect though that most attributivist-palatable Constitutivist theories will fall prey to either 'Agency Schmagency' type objections or, due to treating 'Agent' as just another Kind, the very same objections I have levelled against attributivism. Those that don't will have to rely on some objective, normative metric of rational agency of the sort that attributivists have categorically rejected.

§7.3.3 – Quasi-Attributivism

Attributivism has a lot to recommend it, if we disregard the inconvenient fact that it's false. In many ways it just seems *right*. At least in some sense, things really *should* do what it is their function to do, or else *something* has gone wrong. Likewise, lifeforms really *do* seem deprived in some important way if they can't succeed in their characteristic manner of living. What's more, these judgments don't seem grounded in anything more than an understanding that there *are* such characteristic, or functional, ways of being. Aristotelian necessities *do* seem important, and specifically *because* of the role they play in one's life. Moreover it seems eminently sensible to think that what something *is* should be closely related with how it ought to develop and act. The many philosophers I've discussed have demonstrated too that many of our common normative intuitions as to well-being, social responsibility, happiness, family, etc., can be explained by the role of those things in the functioning of the Human kind. Parsimony too speaks in Attributivism's favour. We readily expect the normativity of plants, animals and artefacts – if there is any – to be cached out in attributive terms; it is pleasingly parsimonious to think that the normativity of rational agents should be no different. Finally, the *guise of the attributive-good* thesis seems plausible, and so gives independent reason to believe that facts about kinds and their goodness should factor in our normative decision making, provided that we believe such is in any way determined by the dispositions of ideally informed agents.

With so much explanatory and intuitive plausibility, it seems odd to think that attributive goodness overlaps in *no* way with the grounds of normativity. Doing what is attributively good very often seems to be just what we ought to do. A good explanation for that would be if whatever *does* ground normativity is the sort of thing that *can* be instantiated via the instantiation of attributive goodness, but not exclusively nor exhaustively. I mean that in the way that travelling to Tibet can be done by travelling to Lhasa, but also and alternatively

by travelling to Shigatse. Say it were fundamentally normative that one ought to travel to Tibet. Since traveling to Lhasa so often satisfies that obligation, we might be forgiven for mistakenly believing traveling to Lhasa was our basic normative obligation. But of course it is not, it's to travel to Tibet, and we might come to see that by investigating what *else* travelling to Lhasa accomplishes. Attributivists might try the same thing.

In essence I'm proposing we take a sort of error-theory approach to Attributivism. Attributivism is false, but granting that, we can ask what explains why we found it so intuitive to begin with, i.e., what explains our making the 'error' of adopting Attributivism in the first place. The answer may be that Attributivism, though false, is indicative of something that *is* normative, or that promoting attributive goodness often happens, coincidentally, to promote something else that we really ought to promote.

Fully fleshing out such a theory goes far beyond what I can do here and now. In the interest of demonstrating that such a project isn't entirely hopeless however, I'll advance one suggestion. Geach wrote in 'Good and Evil' that he disregarded any supposed distinction between the Right and the Good, normatively speaking. He thought there was no meaningful distinction, and that Right, in any normatively relevant sense, is cached out in terms of Good. I think Geach might have been correct on the first point, but that he would have been better served had he understood the relation to hold in the opposite direction. One thing the promotion of attributive goodness might *also* be promoting is the *right* way to be. That suggests *rightness* as an alternative normative property.

Rightness can of course be understood in a variety of ways. One way to understand something – an act or an entity – as 'being Right' is to understand it as being *correct*. Being *correct* though seems very much like what attributivists think of as being attributively good. One way to speak of a tree or a toaster which meets all the standards of its kind is as a good tree or toaster, one that is doing what is good for it. An equally sensible way of speaking of

such a tree or toaster though is as a tree or toaster that is acting *rightly*, i.e., *correctly*, by the standards of its kind. Any *x* therefore which can be described as acting *correctly*, as promoting *rightness*, will often coincide with an *x* that is realising its own attributive good. Thus it's plausible to think that what explains the intuitive appeal of attributive goodness is that promoting it often happens to coincide with what we ought to do, which is to act in ways that are *right*.

So the intuitiveness of attributive goodness might plausibly suggest its own alternative. That alternative might succeed where attributive goodness fails. Attributive goodness is unable to adequately explain why it is sometimes permissible to do things that are bad *qua* one's kind. *Rightness* though, although *involving* facts about one's kind, can take a wider view. An act can be *correct* in many ways. When we ask if it was *right* for *x* to Φ, we might ask if it was right *qua* the standards of *x*'s kind, but we might also mean to ask if it was right within many other contexts. Crucially, it can sometimes be *right*, in the sense of *correct*, for attributively bad things to happen. Although burning down is a bad thing to happen to a tree *qua* tree, for example, there's nothing *wrong* with that event if it happens due to a lightning strike or other natural occurrence. So rightness as a property offers to explain the frequent intuitive merit of attributive goodness, while explaining or mitigating things attributive goodness has difficulty with.

I'm not of course suggesting that the sketch I've now presented is at all correct. For one it mentions nothing as to why we should expect rightness to be authoritative, or how to address conflicts between rightness in different contexts. I offer the view merely as an example that the sort of positive project I'm suggesting the attributivist engage in is not, *prima facie*, hopeless. I expect that should the attributivist wish neither to renounce all their intuitions, nor renounce normative realism entirely, that such a project is their only recourse.

§7.4 – AN ENDING.

In this chapter I've assessed what to do in light of Attributivism's failure. Attributivism, I've argued, should be modified or rejected. In §7.1 I argued that Attributivism can plausibly be modified to work by expanding its scope. Rather than grounding normativity in the goodness of one's actually instantiated kinds, I've argued that attributivists can ground normativity in goodness within the range of kinds agents can instantiate. The resulting *Global Attributivism*, or **GA**, generates the following conclusions. (i) One ought *either* not Φ *or* belong to a kind at the time of Φing *qua* which Φing is good. (ii) One ought *not* to Φ if there is no instantiable kind *qua* which Φing is good.

In §7.2 I argued that **GA** provides almost none of the benefits that have historically motivated the adoption of Attributivism. That, I argued, gives us reason to pursue alternatives to **GA**.

In §7.3 I explored what should be done if neither classical Attributivism, nor **GA**, are plausible. I argued that attributivists must either (i) reject their previous attributivist intuitions, (ii) reject moral realism entirely or (iii) apply the lessons of Attributivism in the pursuit of a new theory of the normative. I've argued that their most promising tactic is to investigate properties the promotion of which coincides with the promotion of attributive goodness. Grounding normativity in such a property could explain why Attributivism's dictates so often overlap intuitively with what we ought to do. That, I've argued, is a promising new avenue of research, and a positive outcome of Attributivism's failure.

8 – CONCLUSION

A lot has been said in this thesis. As a closing act, let's remember what.

I began by introducing Attributivism. An attributivist view is in essence any view that holds (i) that judgments about goodness are attributive – dependent upon standards relative to the kinds of subjects being evaluated – and (ii) that from those judgments we can draw normative conclusions. Attributivism is, in short, the doctrine that says what *x* ought to do is determined by what is attributively good for *x*. Attributivism takes the fundamental and founding concern of normativity to be goodness, and identifies goodness with the property of attributive goodness.

In chapter 2 I explored the attributivist canon. Across §2.2 I detailed the origins and motivations behind Attributivism as a theory. I explained the attributive/predicative distinction, and the arguments behind understanding 'goodness' as an essentially attributive adjective. In §2.3 I introduced the key objections to Attributivism. Specifically I discussed the problems of Motivation and Identification – how we should understand the good of agents as being connected to their motivations; whether or not we can meaningfully identify the good of rational agents; and how. In §2.4 I distilled the canonical answers to those questions. Attributivists assert a necessary connection between the phenomenon of desiring and facts about attributive goodness. They believe that one's kind is determined by one's end-oriented dispositions. They believe that these dispositions are neither optional nor subjective, but that they ground those that are. In essence attributivists answer both problems in one. They draw one's good from what one *is*. They identify what one is through identifying what one does. They ground one's motivations in the same fact.

In chapter 3 I assessed and responded to the wealth of objections that have arisen in response to recent attributivist accounts, like those of Philippa Foot and Judith Thomson. In §3.2 and §3.3 I responded to objections that Attributivism can't 'get off the ground' – that it relies on evolutionary function analyses that are either obviously divorced from the good of the individual, or unable to provide non-arbitrary functional standards. I argued that these objections arise from an insufficiently fine-grained understanding of the interrelation between a subject's 'categorical' behaviours – their essential functional operations – or from a misunderstanding of the sort of ahistorical functional analyses attributivists employ to derive the teleological standards of a subject. Attributivism *is* able to adequately reflect the needs of the individual, and ahistorical function analyses *are* capable of drawing a firm line between functional and incidental operations. In §3.4 I responded to the objection that the norms Attributivism must endorse entail obviously 'evil' or otherwise intuitively objectionable or contradictory outcomes. I resolved the apparent contradictions, and argued that attributivists can and should 'bite the bullet' on some intuitive disagreements, as their theory is, at heart, a revisionist one. In §3.5 I responded to the objection that counting rationality as part of Human essential operations makes the attributivist's appeal to kind standards just a stand-in for an appeal to independent standards of rationality. I argued that attributivists should reject that claim, as Attributivism makes sense of rationality and reason-responsiveness in terms of responsiveness to the various goods of one's kind, without which rationality is an empty concept.

Chapter 4 broke from Attributivism to discuss what is minimally required of a property if it is to be attributed the status of being fundamentally normative. Through §4.1-§4.3 I disagreed with analyses of normativity that understood it in terms of (i) reactive-attitudes, (ii) naïve-instrumentalism or (iii) non-reductive reasons. I argued against (i) that reactive-attitudes require an appeal to some master-set of norms to avoid an intuitively implausible range of

norms, and so endorse a further reduction. I argued against (ii) that naïve-instrumentalism, instrumentalism within any norm-implying framework, is (a) insufficiently connected to motivation, and (b) renders every action equivalently justified, thus trivialising or eliminating normativity altogether. I argued against (iii) that analysing normativity in terms of reasons insufficiently distinguishes between normative and non-normative reasons. I argued that in order to make this distinction, an analysis of normativity that is not in terms of reasons is required. Ultimately I argued that however characteristic and tied up (i) (ii) and (iii) are in our intuitive understanding of normativity, they are insufficient for grounding fundamental normativity. In §4.4-4.5 I argued that what *does* ground fundamental normativity is (i) Authority and (ii) Regulation – a form of inescapable connection to the motivations of agents as understood through the dispositions of their idealised selves, and a capacity for weighing, ordering or otherwise ruling in and out genuinely possible options within the range of potential actions.

In chapter 5 I explained how Attributivism aims to meet the requirements of chapter 4. Much of chapter 5 was devoted to reviewing the material to that point and combining the conclusions of chapter 4 into the dialogue of Attributivism. §5.3 explained how Attributivism must be understood as aiming at Authority. Attributivists assert a 'substantive-aim thesis', or the idea that there is a substantive or constitutive aim to desire, such that all desires are necessarily oriented around the promotion of one's attributive goodness. §5.5 explained Attributivism's method of Regulation. Attributivism orders genuinely possible actions in a predominantly scalar fashion, analysing 'ruled in' actions in terms of virtue, and 'ruled out' actions in terms of defect. Although there is a relatively clear distinction between what is good, or virtuous, and what is defective, and so what is permissible and forbidden, there remains a wide and nuanced range of better and worse actions even within those categories.

In chapter 6 I developed my argument against Attributivism. In §6.1 I further fleshed out how attributivists must understand the connection between desires for the attributive-goodness of one's kind and desires generally. I argued that attributivists must adopt the *guise of the attributive-good* thesis, a view of all desires as conditional desires, conditional upon being fulfilled in a way that's compatible with the promotion of one's attributive-goodness. Across §6.2-§6.3, I argued that the *guise of the attributive-good* thesis doesn't deliver Authority in the way attributivists require, because such conditional desires only grant Authority to the good of one's kind *if* one's kind is inescapable, and that isn't the case. I argued that 'the guise of the attributive-good' thesis can be satisfied while pursuing any action whatsoever, provided that one instantiates a kind for which that action is good. I developed several examples through which I argued for the unrestricted mutability of kinds. I argued that that mutability demonstrates that normativity is not grounded in the good of one's kind; one can find ways of justifiably ignoring its regulations, because one can control the means through which its influence is extended. §6.4 responded to some foreseen objections to my view on the unrestricted mutability of kinds, and the plausibility of the cases I rely upon.

Finally in chapter 7 I evaluated the future of Attributivism and its influence. In the interest of giving Attributivism it's full due, in §7.1 I developed a modified version of Attributivism, I called Global Attributivism, or **GA**, a theory that grounds normativity in the attributive goods of the range of instantiable kinds made relevant by the agent's desires. I argued that this version of Attributivism is plausibly normative. In §7.2 however I argued that the normativity provided by **GA** is exceedingly minimalistic, and arguably either entails or depends upon additional sources of normativity. The result is a normative theory that accommodates very few of the motivations prompting its development, and which is arguably redundant. I argued that there is no reason to accept even **GA**. Finally to conclude, in §7.3 I briefly discussed what options the attributivist might have going forward in the light of

Attributivism's failure, aside from renouncing normativity entirely. I briefly sketched a positive proposal for applying the lessons of Attributivism to a new, quasi-attributivist theory predicated on rightness rather than goodness. I remain hopeful that the near-successes of Attributivism can inform the future of naturalist, cognitivist ethical philosophy.

I'd like to close with a final thought on the place of my arguments in the literature on Attributivism. In particular, I wonder why they haven't been developed sooner. In some ways it is strange that the conclusions I've reached haven't been acknowledged by attributivists. Their own work makes the issue of kind malleability exceedingly relevant; that such is impossible can only be read as an unquestioned, underlying assumption in their philosophy. Yet they have, at times, had to make concessions on other points that should have brought my concerns in to view much earlier. Take, for example, the attributivist response to bad luck and adaptation.

Even the most diehard of attributivists – I'll take Foot here as that representative – must admit to a degree of luck being involved in one's living a good life, even while one embodies only the good characteristics of their kind. A bee that behaves as it should, stinging in defence of the hive, can instead bring on its doom, by drawing the ire of Humans for example. A deer that reacts as it should to danger, by leaping quickly away, might leap into a trap, or into traffic. Even when ones acts as one – attributively – ought to, one runs the risk of *worsening* one's condition. Doing what one ought to in terms of attributive goodness doesn't guarantee that things will go well. Yet shouldn't that imply that one ought to avoid doing what is in their attributive good, in those cases when doing so would go badly?

This of course is the sort of conundrum that inspired philosophers like FitzPatrick or McDowell to object: if an elephant-seal could mate *without* brutal competition, or a wolf learn to hunt *without* relying on their pack, then their categorical goods seem no longer to ground

their obligations. And as I've argued, Attributivism *can* respond to that sort of observation. If it is *better* for some subject to renounce certain categorical behaviours, the attributivist should respond that it's only so in virtue of the alternate action better supporting other, more fundamental categorical behaviours. The *needs* of the subject, in virtue of which it is better for them to renounce certain behaviours, remain rooted in facts about their kind. But what happens if a subject gains the ability to change the whole system? To take control of their needs as well as their behaviours?

Attributivists might even be read as having acknowledged the potential for kinds to change. Foot allows that attributivists must recognise sub-species adaptations – the development of *new* species from old as the old change their behaviours to adapt to challenges.[168] She lets it go at that though, as though not realising the damning consequences of that admission. In the face of which challenges are subjects entitled to adapt? Foot intends environmental challenges, but it seems that admission opens the door to challenges to success in *any* arena. Adaptation is a possibility in the face of challenges to even the most frivolous desire. Once the capability of metamorphosis from one kind into another is acknowledged, the Authority of kinds over desires loses all meaning.

Attributivists have tried to ground the Authority of attributive goodness in an inescapable fact about desire. We never just desire to realise *x*, they've maintained, we desire to realise *x* under the guise of our good, in a way that is good for us. That is, when we desire *x* we desire not just *x*, but also to *be* something for which *x* is good! They have believed that such a fact makes it the case that our motives are realisable *only* by adapting our behaviours to the good of our kind. As should now be clear though, our motives are equally realisable by adapting our kind to our desires; by *becoming* something for which *x* is good, whatever *x* might be.

[168] See Foot 2001, 29.

In short, the attributivist assumption has been that actions must be tailored to fit kinds, in order to satisfy the 'guise of the good' condition on desire satisfaction. They have ignored the viable alternative, that kinds may be tailored to fit actions. In the latter case, agents are given control over the facts upon which Normativity supervenes, and so rob it of its power. Attributivist sanctions are based entirely off the erroneous assumption that it's not possible to change what we are. It is.

BIBLIOGRAPHY

Adams, Robert Merrihew (2006). *A Theory of Virtue*. Oxford University Press.

Amundson, R. and Lauder, G.V. (1994). Function Without Purpose: The Uses of Causal Role Function in Evolutionary Biology, *Biology and Philosophy* 9, pp. 443–469.

Andreou, Chrisoula (2006). Getting On in a Varied World, *Social Theory and Practice* 32 (1), pp. 61-73.

Anscombe, G.E.M. (1978a). Rules, Rights, and Promises, *Midwest Studies in Philosophy* 3, pp. 318- 323.

Anscombe, G.E.M. (1978b). On the Source of the Authority of the State, *Ratio* 20 (1), pp. 1-28.

Anscombe, G.E.M. (1981a). The Question of Linguistic Idealism, *From Parmenides to Wittgenstein*, Basil Blackwell Publisher, pp. 112-133.

Anscombe, G.E.M. (1981b). On Promising and its Justice, and Whether it Need be Respected *in Foro Interno*. *The Collected Philosophical Papers of G.E.M. Anscombe, iii, Ethics, Religion and Politics.* Basil Blackwell Publisher, pp. 10-22.

Anscombe, G.E.M. (1981c). You Can have Sex without Children: Christianity and the New Offer, *The Collected Philosophical Papers of G.E.M. Anscombe, iii, Ethics, Religion and Politics.* Basil Blackwell Publisher, pp. 82-96.

Anscombe, G.E.M. (2000). *Intention*, Cambridge, MA: Harvard University Press.

Anscombe, G.E.M. (2005a). Good and Bad Human Action, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe,* eds. Mary Geach and Luke Gormally. Imprint Academics, pp.195-206.

Anscombe, G.E.M. (2005b). Human Essence, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe,* eds. Mary Geach and Luke Gormally. Imprint Academics, pp. 27-38.

Anscombe, G.E.M. (2005c). Modern Moral Philosophy, *Human Life, Action and Ethics: Essays by G.E.M. Anscombe,* eds. Mary Geach and Luke Gormally. Imprint Academics, pp. 169-194.

Aristotle (1984). *The Complete Works of Aristotle*. The Revised Oxford Translations, 2 vols., ed. Jonathan Barnes. Princeton: Princeton University Press.

Austin, J.L. (1962). *Sense and Sensibilia*, Oxford: Clarendon Press, ed. G.J. Warnock.

Bader, Ralf (2016). Conditions, Modifiers and Holism, *Weighing Reasons*, eds. Errol Lord and Barry Maguire, Oxford University Press, pp. 27-55.

Baker, Derek (2017). The Varieties of Normativity, *The Routledge Handbook of Metaethics*, eds. T. MchPherson and D. Plunkett. Routledge.

Barney, Rachel (2010). Plato on the Desire for the Good. *Desire, Practical Reason, and the Good*, ed. Sergio Tenenbaum. Oxford University Press, pp. 34-64.

Björklund, Frederick, Gunnar Björnsson, John Eriksson, Ragnar Francén Olinder, and Caj Strandberg (2012). Recent Work on Motivational Internalism, *Analysis* 72 (1), pp. 124-137.

Blackburn, Simon (1993). *Essays in Quasi-Realism.* Oxford University Press.

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

Blackburn, Simon (1998b). *Spreading the Word.* Oxford University Press.

Blackburn, Simon (2006). Must We Weep for Sentimentalism? *Contemporary Debates in Moral Theory*, ed. James Dreier. Blackwell Publishing, pp.144-160.

Boyd, R. (1999a). Homeostasis, Species, and Higher Taxa, *Species: New Interdisciplinary Essays*, ed. R. Wilson. Cambridge: MIT Press, pp. 141–185.

Boyd, R. (1999b). Kinds, Complexity and Multiple Realization: Comments on Millikan's 'Historical Kinds and the Special Sciences', *Philosophical Studies*, 95: 67– 98.

Brandt, R. (1979). *A Theory of the Good and the Right*, Oxford: Oxford University Press.

Brewer, Talbot (2009). *The Retrieval of Ethics*. Oxford University Press.

Brock, Dan W. (1977). The Justification of Morality, *American Philosophical Quarterly* 14 (1), pp. 71-78.

Buller, D.J. (1997). Individualism and Evolutionary Psychology (or: In Defense of "Narrow" Functions), *Philosophy of Science* 64, pp. 74–95.

Buller, David J. (1998). Etiological Theories of Function: A Geographical Survey, *Biology and Philosophy* (13), pp. 505-527.

Copp, D. (1997). The Ring of Gyges: Overridingness and the Unity of Reason. *Social Philosophy and Policy* 14 (1), pp. 86–106.

Copp, D., & Sobel, D. (2004). Morality and Virtue: an Assessment of Some Recent Work in Virtue Ethics, *Ethics* 114, pp. 514–554.

Cummins, R. (1975). Functional Analysis, *Journal of Philosophy* 72, pp. 741-765.

Cummins, R. (1977). Programs in the Explanation of Behaviour, *Philosophy of Science* 44, pp. 269-287.

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge, MA: MIT Press.

Danielsson, Sven (2007). On Geach on Good, *Homage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. Eds. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson & D. Egonssson.

Dancy, Jonathan (2006). *Ethics Without Principles*, Oxford University Press.

Darwall, Stephen (1989). 'Reviewed Work: *Ought, Reasons, and Morality* by W.D. Falk', *The Journal of Philosophy* 86 (4), pp.208-214.

Darwall, Stephen (1990). Autonomist Internalism and the Justification of Morals, *Noûs* 24 (2), pp. 257-67.

Darwall, Stephen (2003). 'Moore, Normativity and Intrinsic Value', *Ethics* 113, pp.468-489.

Darwall, Stephen (2006). *The Second-Person Standpoint: Respect, Morality and Accountability*. Cambridge, MA: Harvard University Press.

Dasgupta, Shamik (2017). Non-naturalism and Normative Authority, *Proceedings of the Aristotelian Society* 117 (3), pp.297-319.

Davidson, Donald (1980). "How Is Weakness of the Will Possible?" *Essays on Actions and Events,* ed. Donald Davidson. New York: Oxford University Press, pp21–43

Davidson, Donald (1987). Knowing One's Own Mind, *Proceedings and addresses of the American Philosophical Association*, volume 60, pp. 441-458.

Davies, P. (2000a). Malfunctions, *Biology and Philosophy*, *15*, pp. 19–38.

Davies, P. (2000b). The Nature of Natural Norms: Why Selected Functions are Systemic Capacity Functions, *Noûs* 34 (1), pp. 85- 107.

Davies, P. (2001). *Norms of Nature: Naturalism and the Nature of Functions*. Cambridge, MA: MIT Press.

Enoch, David (2005). Why Idealize? *Ethics* 115 (4), pp.759-787.

Enoch, David (2006). Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action, *The Philosophical Review* 115 (2), pp.169-198.

Enoch, David (2010). How Objectivity Matters, *Oxford Studies in Metaethics* 5, pp. 111-152.

Enoch, David (2011). *Taking Morality Seriously: A Defence of Robust Realism*, Oxford University Press.

Falk, W.D. (1986). *Ought, Reasons, and Morality*, Cornell University Press, Ithaca, NY.

Fine, Kit (1994). Essence and Modality, *Philosophical Perspectives* 8, pp. 1-16.

Finlay, Stephen (2007). Four Faces of Moral Realism, *Philosophy Compass* 2 (6), pp. 820–849.

Finlay, Stephen (2009). The Obscurity of Internal Reasons, *Philosopher's Imprint* 9, pp.1-22.

Finlay, Stephen (2014). *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press.

Finlay, Stephen and Mark Schroeder (2017). Reasons for Action, Internal vs. External, *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>.

FitzPatrick, William (2000). *Teleology and the Norms of Nature*, New York: Garland Publishing.

Firth, Roderick (1952). Ethical Absolutism and the Ideal Observer. *Philosophy and Phenomenological Research* 12, pp.317-45.

Fives, Allyn (2008). Human Flourishing: the Grounds of Moral Judgment. *The Journal of Value Inquiry,* 42 (2), pp.167-185.

Foot, Philippa (1972). Morality as a System of Hypothetical Imperatives, *The Philosophical Review* 81 (3), pp.305-316.

Foot, Philippa (2001). *Natural Goodness*, Clarendon Press - Oxford.

Frank, R. (1985). *Choosing the Right Pond*. New York: Oxford University Press.

Frankena, William K., (1976). On Saying the Ethical Thing, *Perspectives on Morality: Essays of William K Frankena*, ed. Kenneth E. Goodpaster. Notre Dame: University of Notre Dame Press.

Frankfurt, Harry (1988). Freedom of the Will and the Concept of a Person, *The Importance of What We Care About*, Cambridge: Cambridge University Press, pp.11-25.

Frey, Jennifer (2018). How To Be An Ethical Naturalist. In J. Hacker-Wright (Ed.), *Philippa Foot on Goodness and Virtue,* (pp. 47-84). Cham: Palgrave-Macmillan.

Geach, P. T. (1956). Good and Evil. *Analysis* 17 (2). Analysis Committee, Oxford University Press. pp. 33–42

Geach, P. T. (1964). Assertion, *Philosophical Review* 74, pp. 449-465.

Geach, P T. (1977). *The Virtues*. Cambridge: Cambridge University Press.

Gibbard, Allan (1990). *Wise Choices, Apt Feelings.* Cambridge, MA: Harvard University Press.

Gibbard, Allan (2003). *Thinking How to Live.* Cambridge, MA: Harvard University Press.

Hacker-Wright, John (2009). What is Natural about Foot's Ethical Naturalism? *Ratio* XXII, pp. 308-321

Hampton, Jean (1998). *The Authority of Reason*. Cambridge: Cambridge University Press.

Hare, R.M. (1965). *Freedom and Reason.* Oxford University Press, Oxford.

Hare, R.M. (1973). "Rawls' Theory of Justice", *Philosophical Quarterly*, 23 (91), pp.144-155.

Harman, Gilbert (2011). Judith Jarvis Thomson's *Normativity*, *Philosophical Studies,* 154 (3), pp.435-441.

Harman, Gilbert (2012). Moral Relativism Explained. Available through: http://www.princeton.edu/~harman/Papers/Moral_Relativism.pdf (last accessed: 07/02/16)

Heathwood, Chris (2011). 'Desire-Based Theories of Reasons, Pleasure, and Welfare', *Oxford Studies in Metaethics* 6, pp.79-106.

Heim, Irene & Angelika Kratzer (1998). *Semantics in Generative Grammar*. Blackwell Publishing, Oxford.

Hlobil, U. & Nieswandt, K. (2019). Foot with Achilles' Heel. *Philosophia*. https://doi.org/10.1007/s11406-019-00062-y (Last accessed, 16/09/19)

Hooker, Brad (2000). *Ideal Code, Real World: A Rule-consequentialist Theory of Morality*. Oxford University Press.

Hooker, Brad, (2000b). Moral Particularism: Wrong and Bad, *Moral Particularism,* eds. Brad Hooker and Margaret Olivia Little, Oxford University Press, pp.1-22.

Hrdy, Sarah Blaffer (1999). *Mother Nature*. New York: Ballantine Books.

Huemer, Michael (2000). Naturalism and the Problem of Moral Knowledge, *Southern Journal of Philosophy* 38 (4): pp. 575-597.

Huemer, Michael (2006). *Ethical Intuitionism*, Palgrave Macmillan.

Hull, David L (1998). On Human Nature, *The Philosophy of Biology*, eds. Hull, David L. and Michael Ruse. Oxford: Clarendon Press.

Hume, David (1739/1986). *A Treatise of Human Nature*, 2d ed., ed. L.A. Selby-Bigge, rev. P. Nidditch. Oxford: Oxford University Press.

Hursthouse, Rosalind (1999). *On Virtue Ethics*, Oxford: Oxford University Press.

Jackson, Frank and Michael Smith (2016). The Implementation Problem for Deontology, *Weighing Reasons*, eds. Errol Lord and Barry Maguire, Oxford University Press, pp.279-291.

Johnston, Mark (1989). Dispositional Theories of Value, *Proceedings of the Aristotelian Society,* Suppl. Vol. 63, pp.139-74.

Jones, Karen (2005). "Moral Epistemology", *The Oxford Handbook of Contemporary Philosophy*, eds. Frank Jackson and Michael Smith, pp.63–85. Oxford: Oxford University Press.

Joyce, Richard (2008). *The Myth of Morality*. Cambridge: Cambridge University Press.

Katsafanas, Paul (2013). *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.

Katsafanas, Paul (2018). Constitutivism about Practical Reasons, *Oxford Handbook of Reasons and Normativity*.

Kennedy, Christopher (2007). Vagueness and Grammar: the semantics of relative and absolute gradable adjectives, *Linguistics and Philosophy* 30 (1), pp 1–45.

Kitcher, Philip (1993). Function and Design, *Midwest Studies in Philosophy* 18, pp 379–397.

Kitcher, Philip (1999). Essence and Perfection, *Ethics*, 100, pp 59–83.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. (1999). "Self-Constitution in the Ethics of Plato and Kant", *Journal of Ethics* 3, pp.1-29.

Korsgaard, C. (2008). *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Korsgaard, C. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford: Oxford University Press.

Kraut, Richard (1994). Desire and the Human Good, *Proceedings and Addresses of the American Philosophical Association* 68 (2), pp 39-54.

Kripke, Saul (1972). Naming and Necessity, *Semantics of Natural Language*, eds. D. Davidson and G. Harman. Dordrecht, Holland: D. Reidel Publishing Co.

Lack, David (1968). *Ecological Adaptations for Breeding Birds.* London: Chapman and Hall.

Lavin, Douglas, and Matthew Boyle (2010). "Goodness and Desire," *Desire, Practical Reason, and the Good,* ed. Sergio Tenenbaum. New York: Oxford University Press, pp158–99

Law, Iain (2009). Motivation, Depression, and Character. *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, eds. Bortolotti, L. and Broome, M. Oxford: OUP.

Lenman, James (2005). The Saucer of Mud, the Kudzu Vine and the Uxorious Cheetah: Against Neo-Aristotelian Naturalism in Metaethics. *European Journal of Analytic Philosophy* 1 (2), pp 37-50.

Lenman, James (2006). Moral Naturalism, *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/ (last accessed: 21/06/16).

Levins, R. (1968). *Evolution in a changing environment.* Princeton University Press.

Levins, R. and R. Lewontin (1985). *The dialectical biologist.* Harvard University Press.

Lewis, David (2000). Dispositional Theories of Value. *Papers in Ethics and Social Philosophy*. Cambridge, Cambridge University Press, pp.68-94.

Lewis, G. (2006). *Sunbathing in the Rain: A Cheerful Book About Depression*. London: Jessica Kingsley.

Lewtonin, R. (1991). *Biology as ideology.* Harper-Collins.

Lord, Errol and Barry Maguire (2016). An Opinionated Guide to the Weight of Reasons, *Weighing Reasons*, eds. Errol Lord and Barry Maguire, Oxford University Press, pp.3-26.

Lott, Micah (2012). Have Elephant Seals Refuted Aristotle? Nature, Function, and Moral Goodness. *Journal of Moral Philosophy* 9, pp 1-23.

Lott, Micah (2012b). Moral Virtue as Knowledge of Human Form. *Social Theory and Practice* 38 (3), pp 407-431.

Lott, Micah (2014). Why be a Good Human Being? Natural Goodness, Reason, and the Authority of Human Nature. *Philosophia* 42 (3), pp 761-777.

Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.

McDaniel, Kris and Ben Bradley (2008). 'Desires'. *Mind* 177 (466), pp.267-302.

McDowell, J. (1998). Two sorts of naturalism, *Mind, Value, and Reality*. Cambridge: Harvard University Press.

Merricks, Trenton (1998). Against the Doctrine of Microphysical Supervenience. *Mind*, 107(425), 59–71.

Mill, J.S. (1871). *Utilitarianism*, Longmans, Green, Reader and Dyer.

Millgram, Elijah (2009). Critical Notice of *Life and Action*, *Analysis* 96 (3), pp.557-564.

Millikan, Ruth (1984). *Language, Thought and Other Biological Categories*. Cambridge MA: MIT Press.

Millikan, R. (1989a). Biosemantics. *The Journal of Philosophy*, *86*(6), 281–297.

Millikan, R. (1989b). In Defense of Proper Functions. *Philosophy of Science*, *56*, 288–302.

Millikan, R. (1993). Propensities, Exaptations, and the Brain. *White Queen Psychology*, ed. Ruth Millikan. Cambridge: MIT Press, pp. 31–50.

Millikan, R. (2002). Biofunctions: Two Paradigms. In A. Ariew, R. Cummins, & M. Perlman (Eds.), *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford: Oxford University Press.

Millum, Joseph (2006). Natural Goodness and Natural Evil. *Ratio* XIX, pp.199-213.

Moss, Jessica (2010). Aristotle's Non-Trivial, Non-Insane View That Everyone Always Desires Things Under the Guise of the Good. *Desire, Practical Reason, and the Good.* Oxford University Press, pp. 65-81.

Moore, G. E. (1903). *Principia Ethica*. Dover Publications.

Morzycki, Marcin (2015). *Modification.* Cambridge University Press.

Müller, A. W. (2004). Acting Well, *Modern Moral Philosophy*, ed. A. O'Hear, Cambridge: Cambridge University Press, pp.15-46.

Nair, Shyam (2016). How Do Reasons Accrue, *Weighing Reasons*, eds. Errol Lord and Barry Maguire, Oxford University Press, pp.56-73.

Neander, K. (1991a). Functions as Selected Effects: The Conceptual Analyst's Defense, *Philosophy of Science*, *58*(2), pp. 168–184.

Neander, K. (1991b). The Teleological Notion of "Function", *Australasian Journal of Philosophy*, *69*(4), pp. 454–468.

Norcross, Alastair (2006). The Scalar Approach to Utilitarianism, *The Blackwell Guide to Mill's Utilitarianism*, Wiley-Blackwell, pp.217-232.

Nussbaum, Martha C. (1995). "Aristotle on Human Nature and the Foundations of Ethics," *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, eds. Altham, J.E.J. and Ross Harrison. Cambridge: Cambridge University Press, pp.86–131.

Odenbaugh, Jay (2017). "Nothing in Ethics Makes Sense Except in the Light of Evolution? Natural Goodness, Normativity and Naturalism", *Synthese* 194 (4). pp.1031-1055.

Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Parfit, Derek (2006). Normativity, *Oxford Studies in Metaethics* I, pp. 325-80.

Parfit, Derek (2011). *On What Matters*, vols. 1 & 2. Oxford University Press.

Pigden, Charles (1990). On Geach on 'Good', *Philosophical Quarterly,* 40 (159), pp. 129-154.

Platts, Mark (1991). *Moral Realities*. London: Routledge.

Prior, E.W. (1985). What Is Wrong with Etiological Accounts of Biological Function? *Pacific Philosophical Quarterly* 66, pp. 310–328.

Putnam, Hilary (1977). Meaning and Reference, *Naming, Necessity and Natural Kinds,* ed. Stephen P. Schwartz. Ithaca: Cornell University Press.

Quine, W.V. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Quinn, Warren (1993). Putting Rationality in Its Place, *Morality and Action,* Cambridge University Press, pp 228-252

Radzik, Linda (2000). Incorrigible Norms: Foundationalist Theories of Normative Authority, *Southern Journal of Philosophy* 38 (4), pp 633-649.

Railton, Peter (1986). Facts and Values. *Philosophical Topics* 14, pp.5-31.

Railton, Peter (2010). *Toward a Unified Theory of Rationality in Belief, Desire, and Action*, Draft. Accessed at: https://deepblue.lib.umich.edu/handle/2027.42/78460 Last Accessed 27/06/18.

Ratcliffe, M. (2014). *Experiences of Depression: A Study in Phenomenology*. Oxford: OUP.

Raz, Joseph (2000). The Truth in Particularism, *Moral Particularism*, eds. Brad Hooker and Margaret Olivia Little, Oxford University Press, pp.48-78.

Raz, Joseph (2010). The Guise of the Good, *Desire, Practical Reason, and the Good,* ed. Sergio Tenenbaum. New York: Oxford University Press, pp. 111–37.

Read, Stephen (1994). *Thinking about Logic: An Introduction to the Philosophy of Logic*, Oxford University Press.

Rind, Miles and Tillinghast, Lauren (2008). What is an Attributive Adjective? *Philosophy*, 83 (323), pp. 77-88. Cambridge University Press on behalf of Royal Institute of Philosophy.

Roberts, R. (2001). Mental Illness, Motivation and Moral Commitment. *The Philosophical Quarterly* 51 (202), pp.41-59.

Roca-Royes, Sonia (2011). Essential Properties and Individual Essences, *Philosophy Compass* 6 (1), pp. 65-77.

Rodl, Sebastian (2010). The Form of the Will, *Desire, Good, and Practical Reason: Clasical and Contemporary Perspectives*, ed. Sergio Tenenbaum, OUP 2010.

Rosati, Connie. (2003). "Agency and the Open Question Argument." *Ethics* 113: 490-527.

Ross, W.D. (1930). *The Right and the Good*. Oxford University Press, New York.

Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge: Harvard University Press.

Scanlon, T.M. (2002). "Reasons and Passions." *Contours of Agency: Essays on themes from Harry Frankfurt*, eds. Buss, Sarah and Lee Overton, pp.165-183.

Scanlon, T.M. (2011). The Unity of the Normative, *Philosophical Studies*, 154 (3), pp 443-450

Scanlon, T.M. (2014). *Being Realistic about Reasons,* Oxford University Press.

Schapiro, Tamar (2014). “What Are Theories of Desire Theories of?” *Analytic Philosophy* 55 (2), pp.131-150.

Schiffer, Stephen (1976). “A Paradox of Desire.” *American Philosophical Quarterly* 13: 195-203.

Schroeder, Mark, (2005). Realism and Reduction: The Quest for Robustness, *Philosophical Imprint* 5 (1): 1-18

Schroeder, Mark (2010). *Noncognitivism in Ethics*. London: Routledge.

Schueler, G.F (1995). *Desire: Its Role in Practical Reason and the Explanation of Action*, Cambridge, MA: MIT Press.

Sepielli, Andrew (2019). Consequentialism and Evaluation of Action qua Action. *Methodology and Moral Philosophy*, eds. Jussi Suikkanen and Antti Kauppinen. Routledge.

Shafer-Landau, Russ (1998). Moral Judgment and Moral Motivation, *Philosophical Quarterly* 48 (192), pp 353–358

Shafer-Landau, Russ (2003). *Moral Realism: a Defence.* Oxford University Press.

Shoemaker, David W. (2003). Caring, Identification, and Agency, *Ethics* 114 (1), pp.88-118.

Sider, Ted (2013). *Writing the Book of the World.* Oxford University Press.

Silverstein, Matthew (2008). *Normative Authority and the Foundation of Ethics* (Doctoral dissertation). Retrieved from: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/58480/msilverz_1.pdf?sequence=1

Silverstein, Matthew (2012). Inescapability and Normativity, *Journal of Ethics and Social Philosophy* 6 (3): 1-26.

Silverstein, Matthew (2017). Ethics and Practical Reasoning, *Ethics* 127 (2), pp.353-382.

Skorupski, John (2012). The Frege-Geach Objection to Expressivism: Still Unanswered, *Analysis* 72, p. 9-18.

Smith, Michael (1994). *The Moral Problem*. Oxford: Basil Blackwell.

Smith, Michael (2009). Desires, Values, Reasons, and the Dualism of Practical Reason, *Essays on Derek Parfit’s On What Matters*, eds. Jussi Suikkanen and John Cottingham, Blackwell Publishing Ltd. pp.116-143.

Smith, Michael (2010). On Normativity, *Analysis,* 70 (4), pp 715-731.

Smith, Michael (2012). "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*, CXII, pp.309-330.

Smith, Michael (2013). “A Constitutivist Theory of Reasons: Its Promise and Parts,” *Law, Ethics and Philosophy* 1: 1–30

Smith, Michael (forthcoming). 'Parfit's Mistaken Metaethics'. Available through: http://www.princeton.edu/~msmith/mypapers/ParfitsMistakenMetaethics.pdf (last accessed 16/09/19).

Spinoza (1996). *Ethics*, ed. Edwin Curley. Penguin Classics.

Strawson, P.F. (1962). "Freedom and Resentment", *Proceedings of the British Academy*, 48, pp.1–25, reprinted in Watson 2003.

Street, Sharon. (2008). "Constructivism about Reasons," *Oxford Studies in Metaethics* Volume 3, Russ Shafer-Landau (ed.). Oxford: Oxford University Press.

Street, Sharon (2009). In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters. *Philosophical Issues* 19, Metaethics, pp.273-298.

Street, Sharon. (2010). What is Constructivism in Ethics and Metaethics? *Philosophy Compass* 5(5), pp. 363- 84.

Street, Sharon (2012). Coming to Terms with Contingency: Humean Constructivism about Practical Reason, *Constructivism in Practical Philosophy*, eds. James Lenman and Yanatan Shemmer. Oxford: Oxford University Press.

Sturgeon, Scott (2007). Normative Judgment, *Philosophical Perspectives* 21 (1), pp 569-587.

Suikkanen, Jussi (2018). Judgment Internalism: An Argument from Self-Knowledge, *Ethical Theory and Moral Practice* 21 (3), pp.489-503.

Sullivan-Bissett, Ema (2017). "Malfunction Defended", *Synthese* 194 (7), pp. 2501-2522.

Tenenbaum, Sergio (2007). *Appearances of the Good: An essay on the nature of practical reason.* Cambridge: Cambridge University Press.

Tenenbaum, Sergio (2010). *Desire, Practical Reason, and the Good*. Oxford University Press.

Tenenbaum, Sergio (2010). Good and Good For, *Desire, Practical Reason, and the Good,* ed. Sergio Tenenbaum, Oxford University Press.

Tenenbaum, Sergio (2013). Guise of the Good. *The International Encyclopedia of Ethics,* First Edition, ed. Hugh LaFollette. Blackwell Publishing Ltd.

Thomson, Judith Jarvis (2001). *Goodness and Advice.* Princeton, NJ: Princeton University Press.

Thomson, Judith Jarvis (2007). The Right and the Good, *The Journal of Philosophy,* 6, p.277.

Thomson, Judith Jarvis (2008). *Normativity.* Open Court.

Thomson, Judith Jarvis (2010). Reply to Critics, *Analysis*, 70 (4), pp 753-764.

Thomson, Judith Jarvis (2011). Reply to Critics, *Philosophical Studies*, 154 (3), pp 465-477.

Thompson, Michael (1995). The Representation of Life, *Virtues and Reasons,* eds. Hursthouse, R, G. Lawrence and W. Quinn. Oxford: Clarendon Press.

Thompson, Michael (2003). Three Degrees of Natural Goodness. Available through: http://www.pitt.edu/%7emthompso/three.pdf (last accessed 20/06/16)

Thompson, Michael (2004). Apprehending Human Form, *Modern Moral Philosophy*, ed. Anthony O'Hear. Cambridge: Cambridge University Press, pp 47-74.

Thompson, Michael (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*, Harvard University Press.

Thornhill, R. and C. Palmer (2000). *A Natural History of Rape*. Cambridge MA: MIT Press.

van Roojen, Mark (2013). Moral Cognitivism vs. Non-Cognitivism, *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/moral-cognitivism/>.

Velleman, J. David (2000). Introduction, *The Possibility of Practical Reason*, pp.1-31. Oxford University Press.

Velleman, J. David (2000). The Possibility of Practical Reason, *The Possibility of Practical Reason*, pp.170-199. Oxford University Press.

Velleman, J. David (2009). *How We Get Along*. New York: Cambridge University Press.

Wallace, R. J., (1994). *Responsibility and the Moral Sentiments*, Cambridge, MA: Harvard University Press.

Walsh, D.M. and Ariew, A. (1996). A Taxonomy of Functions, *Canadian Journal of Philosophy* (26), pp. 493–514.

Watson, Gary (1987). Free Action, Free Will, *Mind* 96 (382), pp.145-72.

Wedgwood, Ralph (2007). *The Nature of Normativity*, Oxford University Press.

Wiggins, David (1995). Categorical Requirements: Kant and Hume on the Idea of Duty, *Virtues and Reasons,* eds. R. Hursthouse, G. Lawrence, and W. Quinn. Oxford: Clarendon Press, pp. 279–330

Williams, Bernard (1973). The Makropulos Case: Reflections on the Tedium of Immortality, *Problems of the Self*, Cambridge: Cambridge University Press, pp.82-100.

Williams, Bernard (1979). Internal and External Reasons. Reprinted in *Moral Luck*, Cambridge: Cambridge University Press, 1981, 101–13

Williams, Bernard (1989). Internal Reasons and the Obscurity of Blame. Reprinted in *Making Sense of Humanity, and other philosophical papers*, Cambridge: Cambridge University Press, 1995.

Wittgenstein, Ludwig (1953). *Philosophical Investigations* , eds. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.

Wittgenstein, Ludwig (1994). *Remarks on the Foundations of Mathematics,* The MIT Press, Third Printing.

Woodcock, Scott (2006). Philippa Foot's Virtue Ethics has an Achilles' heel, *Dialogue* 45 (3), pp. 445-468.

Woodcock, Scott (2015). Neo-Aristotelian Naturalism and the Indeterminacy Objection, *International Journal of Philosophical Studies* 23 (1), pp.20-41.

Wright, Larry (1973). Functions, *The Philosophical Review,* 82 (2): 139-168.

Ziff, Paul (1963). *Semantic Analysis.* Ithaca, NY; Cornell University Press.