

Human Defect

A Thesis  
SUBMITTED TO THE FACULTY OF  
UNIVERSITY OF MINNESOTA  
BY

Mark Zimmerman Herr

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Valerie Tiberius, Adviser

November 2014

© Mark Zimmerman Herr 2014

## Acknowledgements

A thesis cannot be written without the help and support of others. It would be impossible to name everyone who was a part of writing this, so I will not try. But I would like to acknowledge the support of the philosophy department as a whole, and particularly the help of Anita Wallace in surmounting the nearly endless bureaucratic obstacles involved. I also profited a great deal from my discussions with William Bausman, and the helpful feedback I received from committee members Alan Love and Carl Elliot. But more than anything else, I would like to thank Valerie Tiberius. Without her positive encouragement, patience, and good advice this thesis would never have existed which, I believe, means that it is partially *her* fault, and not just mine.

## **Abstract**

I argue that under an Aristotelian understanding of human nature it is impossible for human beings to ever be fully good.

Some recent accounts of moral virtue, especially that of Philippa Foot, that start from a notion of a fully good life for human beings, and take human goodness to be what is required to live that life. But there is a problem with this, namely that the idea of an ideally good human life requires an understanding of human nature first, and that account of human nature is exactly what is expressed by a picture of what it is to be a good human being. Instead we should start from an idea of what it is to be a human being, expressed in functional terms. I argue that it is possible to do this by taking functions to be part of the membership criteria for certain kinds of things. Then I argue that we can identify which ways of evaluating human beings functionally are moral evaluations by reference to the characteristic use of those evaluations. Finally I argue that if virtue both involves a link to human nature and gets things right, as Aristotelian accounts do, many human virtues are unattainable, since humans characteristically get things wrong.

## Table of Contents

Introduction	1
Chapter 1: Ends	7
Chapter 2: Functions	37
Chapter 3: Virtue	69
Chapter 4: Defect	107
Conclusion	141
Bibliography	147

## **Introduction**

In 1958 Elizabeth Anscombe's essay "Modern Moral Philosophy" argued that recent moral philosophy had gone down the wrong path by focusing on rules for action, and suggested that philosophers would do well to consider the notion of virtue more thoroughly. Since this there has been an increase in interest in both the notion of virtue, and the role it plays in ethics. There have also been a number of accounts focusing on a notion of virtue. The earliest and most influential ones take their inspiration from the works of Aristotle, particularly Natural Goodness, by Philippa Foot<sup>1</sup>, On Virtue Ethics by Rosalind Hursthouse<sup>2</sup> and Dependent Rational Animals by Alasdair Macintyre<sup>3</sup>. Since then, however, there have also been an increasing number of philosophers arguing for virtue ethical accounts inspired by other historical figures such as Hume or Nietzsche, or which are independent of any particular historical tradition<sup>4</sup>

I will be exclusively focusing on accounts within the Aristotelian tradition. These accounts are usually considered *eudaimonistic* accounts. They start by taking there to be a good life for human beings, one peculiar to human beings involving certain characteristic elements such as, usually, loving relationships, fulfilling projects, health, and so on. What this good life consists of is not a purely moral matter. It is based in part on how humans actually go about doing things, what makes us up, and other basic empirical facts about human nature. The species specific nature of this account means that it is constrained by empirical facts to a greater degree than many other normative

---

<sup>1</sup> Foot (2001)

<sup>2</sup> Hursthouse (1999)

<sup>3</sup> Macintyre (1999) and other works, this is just the book most representative of the sort of account I will be discussing.

<sup>4</sup> For example: Annas (2009); Driver (2001); Hurka (2001); Slote (2001); Swanton (2003); and the excellent Tiberius (2010).

accounts are, since it is linked to human biology in a significant way. What makes individual human beings good examples of the human species, or good people, are the traits that fit human beings to live this life. Things that make human beings good human beings are things about them that are needed to live well in the human way. This does not necessarily mean that they are sufficient to live well – according to these accounts even a fundamentally virtuous person can live a poor life if circumstances conspire against them. And there are ways in which human beings can be good or bad which are not considered moral, such as evaluations related to health. Within the general Aristotelian project there are a variety of different questions, but I intend to focus specifically on the relationship between a conception of a good life for human beings and human virtue. Foot and Hursthouse provide what I take to be the most representative answers to this question, within the Aristotelian camp. I will only be discussing Foot specifically here, though given the similarities in their accounts some of what I argue could reasonably be applicable to Hursthouse's work as well.

While I find this general approach appealing I intend to argue that there is a structural problem with this kind of account. If moral virtues are understood as qualities needed in order to live a truly good human life it would have to be possible to give an account of a fully good and fully *human* life prior to saying what kinds of things human beings are, because that is what an account of virtues amounts to. I will argue that this problem can be adequately addressed by focusing on talking about human goodness independently of a final end for human beings, or what a truly good human life amounts to. Talking about how human beings function in the world, and what makes us good or



bad at it, can be done separately, and in a way that preserves both the way in which we talk about human virtues and the way in which Aristotelian theories do so in practical cases. But this change will make one difference, which is that it will let us avoid having to posit that human nature is, in theory if not practice, perfectible. I will argue that this is a substantial advantage because it is not perfectible in the way that is needed if the good life for human beings plays an important structural role in an account of human goodness.

In chapter one I will be discussing the account of human virtue provided by Philippa Foot in Natural Goodness. I will start by summarizing her account of natural evaluations generally, and touch briefly on how she extends it to cover moral evaluations of human beings. The larger part of the chapter will be a criticism of the overall picture. I will argue that relying on an idea of a good *human* life to determine what does and does not support natural evaluations of human beings, including moral ones, is problematic. The idea of a good *human* life must presuppose an already existing understanding of what human nature amounts to and cannot straightforwardly be used to determine why some things or are not virtues since that understanding of human would have to include knowing which things are and are not virtues of human beings. I will conclude that a broadly Aristotelian account must be based in an account of how human beings function, and take the idea of a good human life only as a consequence of that not an explanation of it.

Then, in chapter two, I will suggest a way to understand how we could make sense of attributing functions to parts of human beings in a way that would both support the sort of natural evaluations that allow us to talk about virtues of human beings, and do

so without referring to an idea of an ideally good human life. Initially I will discuss the literature on attributing proper functions to things in biology, but conclude that none of the available accounts have the resources to do what I need them to do. I will then argue that we can understand functional evaluations as bound up in an understanding of the membership criteria of some kinds of things, and in this case, the parts of human beings that we evaluate in moral terms.

In chapter three I will proceed to talk about moral evaluations in particular. What I suggest in chapter two is neutral with regards to the different possible ways of evaluating human beings, including both moral evaluations and also, for example, evaluations related to health. But we treat moral evaluations of human beings differently from other kinds of evaluations of human beings, so it is important to know how to pick out which evaluations are moral ones, and to make sense of why it is that they seem distinctive. I will argue that the best way to pick out moral evaluations of human beings from other ones is by reference to the characteristic use we put them to and, specifically, their place in encouraging certain behaviors and discouraging others.

Finally in chapter four I will argue that there is a further advantage to taking the idea of a good human life to be purely a consequence of an account of human virtue, and not to play a direct role in determining what is and what is not a virtue. Taking an ideally good human life to determine what is and is not a virtue requires the possibility of there being such a life. But whether or not we are capable of living this kind of live, or whether human nature is theoretically perfectible is an empirical question. If human nature is not coherent enough to allow for this possibility then that would cause a

significant problem for an account which takes it to determine what is and is not a virtue, but not for an account which takes the idea of a good life for human beings to be purely a result of a separate account of human virtue. I will argue that human nature is not capable of being perfected in this way. I will appeal to three different morally relevant psychological features of human beings which prevent us from being fully perfectible with regards to those areas. From that I will conclude that human nature does not cohere in the right way to be capable of being perfected in every aspect: we are by our nature defective. This does not mean that we are all wretched or awful, but it does mean that referring to an idealized notion of human nature, or an understanding of it according to which we are theoretically perfectible will not work. We are directed at ends which we are fundamentally incapable of fully achieving. Finally I will argue that an understanding of human nature which is internally inconsistent, which lacks a coherent final end, is more valuable when it comes to dealing with the ways we actually live, and the situations in which we find ourselves.

**1. Ends**

In Natural Goodness Philippa Foot argues that justice, kindness, and moral actions in general are rational not because of their results or effects but because the terms themselves pick out ways in which human practical reasoning can be excellent. Her argument rests on an account of what she calls natural goodness and defect. By natural goodness and defect Foot is referring to what she argues is a distinctive way of evaluating living things: that is, in terms of how they *ought* to be in as much as they are members of some kind of living thing, or share in a particular ‘life form’.<sup>5</sup> All living things are, on her account, subject to this general type of evaluation, though human beings are distinctive in that goodness for human beings is not simply biological goodness, as it is for animal and plants. Traits in humans that are commonly taken to be virtues are, on her account, ways in which the practical will of human beings can display natural excellence, and vices are a sort of natural defect. Here I will be focusing primarily on the general account of natural evaluation that Foot provides in chapter two of Natural Goodness (“Natural Norms”), and chapter three (“Transition to Human Beings”).

I will start in section 1 by briefly discussing Michael Thompson’s article “The Representation of Life” before turning to Foot’s account in section 2, which builds on it. After discussing Thompson and Foot on natural evaluations of living things generally I will discuss in section 3 how Foot takes this account to extend to human beings in particular. In section 4 I will analyze what Foot means when she appeals to a natural end for living things. And, in section 5, I will argue that the account she gives of it results in a dangerous circularity: that end must simultaneously explain and be explained by

---

<sup>5</sup> “By contrast, ‘natural’ goodness, as I define it, which is attributable only to living things themselves and to their parts, characteristics, and operations, is intrinsic or ‘autonomous’ goodness in that it depends directly on the relation of an individual to the ‘life form’ of its species.” (Foot, 27)

reference to what makes some living thing a good example of its species. I will conclude in section by arguing that this problem means that the idea of a distinctively human good cannot play a role in explaining what makes people good or bad human beings without abandoning a recognizably Aristotelian account. Instead it can only be a consequence of human goodness.

## 1.

In “The Representation of Life” Michael Thompson argues that we have a distinctive way of speaking about living things, one that cannot be eliminated while still understanding them as living things. Thompson claims that our way of talking about living things includes a particular sort of general claim about living things that he calls “natural historical” judgments<sup>6</sup>. These judgments can take many forms, but Thompson suggests that the most common is what he calls ‘Aristotelian Categorical’ sentences. Examples of Aristotelian Categoricals include, most distinctively, sentences in which the subject is definite and singular, such as “The mayfly breeds shortly before dying”, but also can take indefinite plural subjects, such as “Oak trees have deep, spreading roots.” Sentences of this sort attribute properties to natural kinds of living things, usually, Thompson suggests, species or more general kinds of living things, and “express one’s *interpretation* or *understanding* of the life-form shared by the members of the class.”<sup>7</sup> Aristotelian categoricals do not exhaust the ways in which natural historical judgments can be expressed, according to Thompson, but are especially indicative of what is expressed by them.

---

<sup>6</sup> Thompson, 281.

<sup>7</sup> Thompson, 288.

Natural historical judgments, and Aristotelian categorical sentences, are strange in certain important ways. While they appear to express general claims they do not do so in a way reducible to other ways of speaking about groups of objects<sup>8</sup>, specifically, they cannot be understood as quantifying over some group of members. The most natural way to attempt to read Aristotelian categorical sentences as quantified would be to take them to express some form of universal quantification. The best reason to take them to be universally quantified statements is that, Thompson suggests, “The S is F” and “The S is G” together entail that “The S is F and G”. This rule for combining Aristotelian categorical distinguishes these sentences from ones that might express an existentially quantified sentence, or even one expressing a statistical generalization. However, natural historical judgments accommodate exceptions in a way that universally quantified statements cannot. If taken as expressing universal quantification, the sentence “The mayfly breeds shortly before dying” is clearly false: in fact, most do not breed at all. And, Thompson argues, this cannot be covered with the addition of a *ceteris paribus* clause.

It is worth noting that by repeated application of our apparently unexciting rule of inference – ‘Ss are F’, ‘Ss are G’, *ergo* ‘Ss are both F and G’ – we will presumably always be able to produce a true statement of our form involving a complex conjunctive predicate that is not true of *any* member of the kind denoted by its subject, living or dead. I mean: nobody’s

---

<sup>8</sup> This is a common feature of several sorts of general claims, often referred to ‘generic sentences’, though not all generic sentences are Aristotelian categorical sentences, or express natural historical judgments. With the exception of the final important feature of Aristotelian categorical sentences that I will discuss here, that is, their capacity to be combined in a distinctively teleological way, however, the interesting features of Aristotelian categorical sentences are common for most generic sentences.

perfect. (Will anyone say, by the way, that anything is, *ceteris paribus*, what it never is?)<sup>9</sup>

Since Aristotelian categoricals do not describe members of some kind as such but instead the life form that they share as members of that kind it is possible to infer from, for example, “The monarch butterfly flies south in the fall” and “The monarch butterfly flies north in the spring” that “The monarch butterfly flies north in the spring and south in the fall,” according to the inference rule that Aristotelian categoricals have in common with universal generalizations.. No particular monarch butterfly does so, however, or ever could since their average lifespan is approximately two months. It is unclear how any *ceteris paribus* clause could cover this sort of situation.

Aristotelian categoricals are also, Thompson suggests, unique in that they can be combined to express weakly teleological judgments, which he calls ‘natural teleological’ judgments, though not by means of some direct rule of inference. These judgments are similar to the natural historical judgments that make them up in that they are in some sense about the life form shared by members of a kind: “Their linguistic expressions, that is, are fit to enter into certain sorts of ‘final causes’. ‘They have blossoms of such-and-such type in order that such-and-such insects should be attracted and spread their pollen about.’”<sup>10</sup> Instead of describing some characteristic features of members of that kind, however, these describe the role or function that some characteristic feature of members of that kind plays in the life form of members of that kind.<sup>11</sup> Finally, and importantly,

---

<sup>9</sup> Thompson, 287-288

<sup>10</sup> Thompson, 292-293

<sup>11</sup> Unlike stronger teleological claims these do not express or entail any sort of intention as shown by, Thompson suggests, the fact that in these cases “P in order that Q” entails the truth of *both* P and Q, instead of merely the truth of P as in the case of sentences expressing some sort of intention.



Thompson suggests that it is possible to use natural historical judgments to support evaluations of living things to a certain limited extent.

If, though, we want to apply ‘normative’ categories to subrational nature, and apart from any relation to ‘our interests’, then the questions inevitably arise, and not so unreasonably: Where does the standard come from? What supplies the measure? The system of natural historical propositions with a given kind as a subject supplies such a standard for members of that kind. We may implicitly define a certain very abstract category of ‘natural defect’ with the following simple-minded principle of inference: *From* ‘The S is F’ and ‘This S is not F’ *to infer* ‘This S is defective in that it is not F’.<sup>12</sup>

Since natural historical judgments provide a description of the characteristic features of some kind, in effect, they can provide grounds for evaluating some particular member of that kind. If oak trees characteristically have deep roots, and one oak tree in particular has shallow roots, then it can be inferred that that particular oak tree has defective or poor roots for an oak tree, or that the tree itself is a defective or poor oak tree. This sort of evaluation is limited at best: it allows for evaluations only to the extent that it allows one to say that some living thing is characteristic or not characteristic of its type. And while, as in the case of the oak tree with shallow roots, it can sometimes capture an intuitive sense of what it is to have a defect there are less congenial examples available as well. Natural history judgments can just as easily pick out unimportant but characteristic features of living things, or even characteristic defects. It would be peculiar to infer from “Maple trees are susceptible to rot as they age” and “This particular maple tree is not susceptible to rot” that “This particular maple tree is defective in that it is not susceptible

---

<sup>12</sup> Thompson, 295

to rot.” For the sort of evaluations in question to be interesting or useful beyond what Thompson suggests there would have to be some way of specifying which characteristics are important or significant to the life form in question, and which are unimportant or worse. This is not a particular problem for his account, which only suggests the possibility of this sort of inference, but it is a problem for any account, like Foot’s, which attempts to use it to arrive at a broader or more interesting account..

## 2.

In Natural Goodness Philippa Foot makes a similar point concerning Thompson’s suggestion, and suggests a way in which the account can be modified in order to arrive at an account of natural evaluations that lacks this problem. The problem according to Foot’s analysis is that Thompson failed to separate natural historical judgments such as “The male peacock has a brightly colored tail” from other superficially similar propositions such as “The blue tit has a round blue patch on its head” that do not entail evaluative conclusions.<sup>13</sup> While both appear to be Aristotelian categorical sentences, it is only the first that should entail evaluations of the natural goodness or defect of individual members of the species in question. Foot concludes as a result that only the first could be a proper Aristotelian categorical, and that the difference between the two must be on the level of the meaning of the sentence itself, not the features that are attributed to the species.

“His talk of ‘natural history propositions’ was perhaps misleading in that it did not explicitly separate out what I would like to call the teleological

---

<sup>13</sup> Foot, 30

from the non-teleological attachment of predicates to a subject term that is the name of the species.”<sup>14</sup>

Foot suggests that the distinction here is largely between Aristotelian categoricals, on the one hand, and statistical generalizations on the other hand. Sentences like “The blue tit has a round blue patch on its head” may appear to be Aristotelian categoricals but simply express statistical generalizations, and do not relate to the teleology of types of living things.

If Aristotelian categoricals are distinct from apparently similar sentences that attribute features to kinds of living things, it is important to be clear on precisely how they can be distinguished. The answer, for Foot, is that Aristotelian categoricals describe features that are related to the natural ends of living things, namely, to self-maintenance, including development, and reproduction.

Aristotelian categoricals are propositions having to do with the way that certain features appear or that certain things are done in organisms of a given species either by the whole organism or by their characteristics or parts. But, speaking now for myself rather than for Thompson, I should say that in plants and non-human animals these things all have to do, directly or indirectly, with self-maintenance, as by defense and the obtaining of nourishment, or the reproduction of the individual, as by the building of nests. This is ‘the life’ characteristic of the kind of animals with which the categoricals here have to do. What ‘plays a part’ in this life is that which is causally and teleologically related to it, as putting out roots is related to obtaining nourishment, and attracting insects is related to reproduction in plants.<sup>15</sup>

---

<sup>14</sup> Ibid.

<sup>15</sup> Foot, 31.

The features in question are all important because they ‘play a part’ in ‘the life’ characteristic of living things of that kind. What it is for something, an activity, feature, etc., to play a part in the life characteristic of some sort of thing is for that thing to be needed for something of that sort to live a life that is characteristic of things like it, or to achieve the good for things of that sort.<sup>16</sup> The good life that is characteristic of plants and non-human animals is given in purely biological terms: it amounts to a life of self-maintenance and reproduction. Foot also, in places, includes ‘development’ as an additional part of the life of living things, though this could equally fall under self-maintenance to a certain extent. It is this life characteristic of living things of a certain species or kind, or good for members of that kind, that determines what features are important or unimportant and, as a result, determines how living things of that kind *should be*.

This same structure is reflected in the propositions that describe these features, though here there is a need for a very careful reading of what Foot argues. The preliminary point to be made is that while Thompson distinguishes between what he calls ‘natural *historical*’ judgments, expressed by Aristotelian categorical sentences, and ‘natural *teleological*’ judgments, Foot does not make this distinction, and considers *both* to be ‘Aristotelian categoricals’. For example, “There is an Aristotelian categorical about the species *peacock* to the effect that the male peacock displays his brilliant tail *in order* to attract a female during the mating season.”<sup>17</sup> Similarly when Foot claims that “What is

---

<sup>16</sup> In this quotation Foot does not explicitly state that ‘the life’ characteristic of some sort of living thing can be identified with the *good* for things of that kind. But she does at other places make this claim explicitly, e.g: “For such an evaluation is based on the general relation of this kind of feature to the pattern of life that is the *good of* creatures of this species.” (41).

<sup>17</sup> Foot, 31

crucial to all teleological propositions is the expectation of an answer to the question ‘What part does it play in the life cycle of things of the species S?’ In other words, ‘What is its function?’ or ‘What good does it do?’<sup>18</sup> she is including under ‘teleological proposition’ both “The male peacock has a brightly colored tail” and “The male peacock displays his brilliant tail *in order to* attract a female during the mating season.” There is a slight awkwardness here as the second sentence appears to be an *answer* to the question “what is its function?”, but functions can themselves have further functions within the characteristic lives of kinds of living things. The reason this is important is that Foot’s account of what makes the Aristotelian categorical sentences true (those that can support inferences to natural evaluations) will cover all of what she calls Aristotelian categoricals.

Since, according to Foot’s account, Aristotelian categoricals express teleological claims, including when they appear to express only general claims about living things, and since the teleology, and teleological status of parts of kinds of living things depends on the natural end of those kinds of living things, then Aristotelian categoricals must depend for their truth upon the natural end of the kind of living thing in question. This link between the natural end of living things and the Aristotelian categoricals that describe their way of life is what distinguishes them from other, apparently similar generalizations, and also the reason that Aristotelian categoricals can support natural evaluations. The relationship is not simply that the features they describe are

---

<sup>18</sup> Foot, 31-32

teleologically linked to the natural ends of the living things, but that this link is reflected in the meanings of the Aristotelian categoricals themselves<sup>19</sup>:

What, then, determines the truth of the teleological propositions of the non-quantifiable form that meet Thompson's conditions? We start from the fact that it is the particular life form of a species of plant or animal that determines how an individual plant or animal should be: the Aristotelian categoricals give the 'how' of what happens in the life cycle of that species. And all the truths about what this or that characteristic does, what its purpose or point is, and in suitable cases its function, must be related to this life cycle. The way an individual *should be* is determined by what is needed for development, self-maintenance, and reproduction: in most species involving defence and in some the rearing of the young.<sup>20</sup>

Since on Foot's account it is development, self-maintenance, and reproduction that is the good for non-human living things, the way any individual of some species should be, or ideally is, is set by what is needed for that good. The understanding of the good for things of that species determines what it is to be a good member of that species, and actual members of that species can be evaluated according to the account of what it is to be a good member of that species. The specific way in which the life forms of different species do achieve this good is what is described by Aristotelian categoricals. They attribute features, behaviors, etc., to kinds of living things, and they are true of them to the extent that they are necessary for living well for members of that kind. As a result

---

<sup>19</sup> This quotation directly follows Foot's quotation of a passage from Thompson describing natural teleological propositions. It is *possible* that she intends this to apply only to what he calls natural teleological propositions and not natural historical ones. But given the general context and Foot's suggestion that the two are both of the same category generally (Aristotelian categoricals) as I've argued immediately above I think this covers not just what Thompson calls natural teleological propositions but the more general category.

<sup>20</sup> Foot, 32-33

sentences like “The male peacock has a brightly colored tail” are different in meaning from a apparently similar sentences like “The blue tit has a round blue patch on its head”. In the former case the sentence attributes a certain property, having a brightly colored tail, in a specific way which entails that that property is necessary for obtaining the good for peacocks in the way characteristic of peacocks. In the latter case the sentence lacks this distinctive feature, amounting to no more than a general description of blue tits. Foot argues that this is what distinguishes Aristotelian categorical sentences from other sorts of general statements about living things of that kind, and, in combination with descriptions of particular living things, is what makes them entail evaluations of those living things.

This structure also allows us to account for virtues of living things, or something very much like virtues, as falling under a more general category of things that are necessary for creatures of some kind to live well. This does not mean that a peacock with a brightly colored tail is guaranteed to live a good peacock life, even if he is excellent in other important respects as well. He would only be well fitted to live that sort of life, not guaranteed it. Changes in circumstance, especially involving situations that are not the natural habitat of the peacock, can affect the life he lives, and perhaps other factors as well. It is this that allows us to say, on Foot’s account, that some things are good for peacocks, and others bad: they are good and bad for peacocks to the extent that they help in bringing out a good life for peacocks, in the same way that having bright tails or safe drinking water are good for peacocks. In all of these cases it is the notion of a natural end

common to either all members of some kind of living thing or to living things in general<sup>21</sup> that is the central theoretical notion standing behind the evaluations of natural goodness or defect. “What conceptually determines goodness in a feature or operation is the relation, for the species, of that feature of operation to survival and reproduction, because it is in that that good lies in the botanical and zoological worlds.”<sup>22</sup>

### 3.

When it comes to evaluations of human beings, however, Foot suggests that the account would have to be somewhat different. This is not because the *sort* of evaluation in question would be different, or because the basic structure involved would be different. It is that in the case of human beings a good life is not identified with things like reproduction, and self-maintenance: the good for human beings is different from that of other living things. The good for human beings is substantially more complex and contains more than simply reproduction and survival (though it is very likely that it does also contain those as elements): the

“human good must indeed be recognized as different from good in the world of plants and animals, where good consisted in success in the cycle of development, self-maintenance, and reproduction. Human good is *sui generis*.”<sup>23</sup>

---

<sup>21</sup> Or, as Foot suggests in another point, to all non-human living things in general. The ambiguity on her account between the natural ends being the same for all living things or different for different kinds is due to the fact that when speaking of ends she speaks of things quite generally, and refers to Aristotelian categoricals as describing *how* those ends are characteristically achieved. But similarly she argues that the truth conditions of the Aristotelian categoricals depend on and are distinct from the “life cycle”, meaning the natural ends, of that sort of living thing which suggests equally that the natural ends are proper to kinds of living things not living things generally. (The difference here could be as small as saying that one end of, say, horses is reproduction, and saying that it is *sexual* reproduction. But it is a substantial difference on a theoretic level.)

<sup>22</sup> Foot, 42

<sup>23</sup> Foot, 51



There are, Foot suggests, more ways in which human beings can be deprived of good than is true for other animals: there are many things that humans need to live well that are not needed by other living things. These things – a capacity for language, to appreciate art, tell and laugh at jokes, etc. – are not necessary for reproduction or survival. And without them a human being might well succeed as far as these criteria go, but they would still be noticeably lacking in one or more things necessary for living a properly good human life. For other living things there are, Foot argues, no equivalent set of further ends or goods.

But there is still a strong similarity between evaluations of non-human living things and evaluations of human beings on Foot's account. The difference between evaluations of human beings and that of animals is that the good for human beings is more intricate, composed of more diverse goods, and requiring a more complex system of norms than that of animals. But the conceptual structure underlying the evaluations in question is identical: the difference is merely that the foundation for those evaluations has a greater degree of complexity, and that this is mirrored in the evaluations themselves.

In spite of the diversity of human goods – the elements that make up good human lives – it is therefore possible that the concept of a good human life plays the same part in determining goodness of human characteristics and operations that the concept of flourishing plays in the determination of goodness of plants and animals.<sup>24</sup>

Despite the difference at the level of individual evaluations, then, the basic structure underlying evaluations of non-human living things also underlies evaluations of human

---

<sup>24</sup> Foot, 44. Here by “flourishing” Foot clearly means a more limited notion of a good life - one characterized in largely biological terms – than applies to the good life for human beings instead of a more general notion of flourishing which might also encompass human goodness as well. This is largely an unimportant terminological difference.

beings. In both cases there are natural ends for members of the kind in question, and a set of Aristotelian categoricals describing how they are characteristically achieved. And these Aristotelian categoricals, depending for their truth on the natural end or good life for living things of that kind, are capable of supporting inferences to a sort of natural evaluation of particular members of the kind in question, whether that kind is a non-human living thing like the wolf or a kind of living thing like a human being with a more complicated good.<sup>25</sup>

It is on this basis, then, that Foot goes on to argue that the traditionally considered virtues, such as justice, courage, etc. are rational. They are, the argument goes, excellences of the practical reason and excellent in that they are needed in the way that a brightly colored tail is needed for a peacock: they are necessary conditions for living well in the way characteristic of or proper to human beings. The fact that they are linked in this way to the good for human beings is not, however, the justification for acting according to them: it is the justification for *taking* them to be excellences of practical reason. The justification for acting justly, and the like, just is that doing so is acting according to reasons, namely reasons of justice. The question of whether or not these things give us proper reasons to act is, as a result, a poor question: they are this by definition and not as a result of some consequence of the sort of action in question.<sup>26</sup> In

---

<sup>25</sup> “I maintain that a common conceptual structure remains. For there is a ‘natural-history story’ about how humans achieve this good as there is about how plants and animals achieve theirs. There are truths such as ‘Humans make clothes and build houses’ that are to be compared with ‘Birds grow feathers and build nests’; but also propositions such as ‘Humans establish rules of conduct and recognize rights’. To determine what is goodness and what defect of character, disposition, and choice, we must consider what human good is and how human beings live: in other words, what kind of a living thing a human being is.” Foot, 51

<sup>26</sup> This argument is somewhat tangential to the topic under consideration here, so the treatment of it here is quick. The fullest version of this argument in Natural Goodness occurs on pages 62 to 64.

this way Foot's answer depends on the account of natural evaluations described here: it depends on an account of what is good for human beings (or a good life for human beings), and how human beings characteristically achieve it. And while evaluations of human beings differs in some ways from that of other living things the basic conceptual structure here is equivalent to that of other living things.

#### 4.

However there is a significant problem for Foot's account. Before showing what this problem is, it is necessary to spend sometime on what Foot means by the good for things of a certain kind, which is something that I have left relatively ambiguous until now. It is important because Foot's account of how to go about evaluating living things depends on the natural ends of those living things. Those ends are what give characteristic features the sort of importance in the life of members of that kind that makes them a subject for that sort of evaluation.

It is not at a first glance clear what the natural ends of living things are on Foot's account, though there are two relevant possibilities. The first possibility is the simplest, namely that the good for living things can be identified with reproduction, development, and self-maintenance. As a group of ends these can be referred to as the life cycle for individual members of some kind or other, according to Foot's account. On this way of talking there is a distinct sense in which all living things share precisely the same natural ends<sup>27</sup>, though they achieve them in different ways. There at least two places in Foot's account of natural evaluations where she seems to be explicitly arguing for something

---

<sup>27</sup> With at least one exception, according to Foot, that exception being human beings. Foot does not give any justification for treating human beings differently from other animals.

akin to this view. The first comes when she summarizes her general picture on pages 33 and 34:

- (a) There was the life cycle, which in these cases consisted roughly of self-maintenance and reproduction.
- (b) There was the set of propositions saying *how* for a certain species this was achieved: how nourishment was obtained, how development took place, what defenses were available, and how reproduction was secured.
- (c) From all this, *norms* were derived, requiring, for instance, a certain degree of swiftness in the deer, night vision in the owl, and cooperative hunting in the wolf.
- (d) By the application of these norms to an individual member of the relevant species it (this individual) was judged to be as it should be or, by contrast, to a lesser or greater degree defective in a certain respect.<sup>28</sup>

The most important or telling feature here is the distinction between the natural ends for living things of some or other kind, called the life cycle, and the distinctive way in which each species obtains those ends, as expressed by certain propositions. If there is a sharp distinction between the natural ends for living things and the ways in which members of a species achieve these ends then it is necessary to view the end in question as being relatively abstract, common to many if not all living things, and justified separately from concrete questions about the ways in which particular sorts of living things go about living. Another place in which Foot appears to say something conducive to this way of reading her account is on page 42, where she claims that

What conceptually determines goodness in a feature or operation is the relation, for the species, of that feature or operation to survival and

---

<sup>28</sup> Foot, 33-34.

reproduction, because it is in that that good lies in the botanical and zoological worlds.<sup>29</sup>

This example is less suggestive than the preceding one, but if what is being asserted here is that it is the relation to reproduction considered *generally*, or in the abstract, that makes some feature or operation good or defective, say, then it is conducive to the first way of reading Foot's account of natural ends for living things.

I do not think that this is the best way of reading Foot. The first reason to avoid reading Foot as arguing for this account of natural ends is that there is a substantial problem for this way of talking about ends, and the second reason for avoiding reading Foot as arguing for this account is that there is another, more plausible account which could be read into the above passages. In other passages Foot seems to suggest a slightly different account, one in which it is not achieving some particular ends set out as good, such as reproduction, that is the good for living things but instead the good for living things is to instantiate the life form characteristic of the sort of living thing in question. This position could be read into the above quotations if it is assumed that, as a general matter, the life forms of living things are directed at the ends that Foot mentions, though not necessarily, and not all in the same way. This does require some care in reading, but the additional claims needed here do not seem obviously false so it is not off limits.

Foot can be read claiming this second position most explicitly on page 91, where she states that "To flourish is here to instantiate the life form of that species, and to know whether an individual is or is not as it should be, one must know the life form of the species." And, immediately following on page 92, she states that "goodness in respect of

---

<sup>29</sup> Foot, 42.

bodily health, of faculties such as intelligence and memory, and so on is precisely that which fits a living thing for the instantiation of the life form of its species, and that this counts as the good of a living thing...” Finally, earlier on page 41, she claims that “such an evaluation is based on the general relation of this kind of feature to the pattern of life that is the *good of creatures of this species.*”<sup>30</sup> According to this position the natural end for a kind of thing is the instantiation of an idealized form of life for members of its kind. Since Foot also claims that the living things are generally directed at ends like reproduction and self maintenance, the good for living things of some particular kind would amount to instantiating a form of life concerned with achieving those ends in a specific, characteristic way. Along these lines there would be a unique, though for the most part structurally similar set of ends for each sort of living thing, not simply a different way of achieving the natural ends held in common by all living things.

The central difference between these two different accounts of the good for living things, which may not be immediately obvious, is the specific nature of the natural ends against which individual members of some kind of living thing are measured. According to the first possibility, which I have suggested is the less likely reading of Foot’s account, the end against which individuals is evaluated are common to all living things. The specific features of the life cycles of various kinds of animals, that is, how they go about reproducing, maintaining themselves, and so on, are instrumentally related to the natural end of all living things. That is, they are how the good for that kind of thing is brought

---

<sup>30</sup> I think that it is possible to also read these quotations as suggesting something akin to the earlier position I suggested, where living things share certain, abstract, natural ends. However, it is on balance *more difficult* to read these passages as suggesting this than it is to read the earlier passages as suggesting the other view that I have described. In addition, the first view has certain features which would make some of the other things Foot argues difficult. As a result, on balance, I think the second view is probably a better one to attribute to Foot.

about, or achieved by members of that kind, and are subject to the sort of natural evaluations in question as a result of this.<sup>31</sup> As a result *all* living things no matter how they characteristically live are directed at that end, in as much as they count as living things at all.

The reason that it would have to be tied to the notion of a living thing in general, that is, applying not only to all living things but to all living things in as much as they are living things, is because the sort of evaluations that this sort of account of natural ends would return are kind dependent. If the sort of evaluation that these natural ends support is that some particular member of a kind is good or bad in as much as they are a member of that kind of living thing, then the ends must be related to what it is to be a living thing of that kind. And if the ends are common to all living things it must be that what is especially important is not that some particular thing is a member of a specific kind, but that it is a living thing, or that it is a member of a *kind of* living thing. It is not difficult to imagine how this sort of justification could go, for Foot. The natural end, or the good for living things, would have to be justified by reference to the idea of a living thing in the first place, or by reference to what is essentially true of living things in as much as they are living things. In other words, the sort of end in question would be drawn from the

---

<sup>31</sup> It is not automatically clear why, according to this reason, it might be good for some member of a kind to achieve the good for living things in the way *characteristic* of members of that kind when it might, perhaps more easily, achieve it in some other fashion. It is difficult to see why, that is, the fact that members of some kind generally or characteristically achieve the good for members of that kind in one way, rather than another, is important. However this can be largely avoided, though in a slightly unsatisfactory way, by simply distinguishing between evaluating something as a living thing, and evaluating it as a member of some particular kind of living thing. Along these lines it could be that a free-loading social animal, for example, is a poor member of its species since it is poorly fit to achieve the good for living things in the way characteristic of that species, but an excellent living thing, since it is admirably fitted to achieve the good for living things (by exploiting other members of its species). It is not clear that this would be an easy fix for Foot's general account, though, since it might have unfortunate consequences when applied to human beings.

idea of something that maintains a certain sort of homeostasis, takes in energy in various ways and transforms it into other forms for various ends, is capable of making more things like it in appropriate circumstances, and so on (or whatever specific account of life is relevant here – Foot does not provide one.) The sort of natural ends that would most likely result from this sort of reasoning are at least similar to the ones that Foot suggests, and so the first possible account of the good for living things should be tempting here, though, as I have suggested, I take the second account to be a better reading of what she argues.

According to the second possible account of the good according to which individual living things can be evaluated, the natural ends for living things are contained within or identified with the form of life proper to specific kinds of living things, as opposed to being conceptually, separable from them. As a result the justification for this account of natural ends cannot rely on considerations separable from the actual facts about living things of a particular kind, in the way that the preceding suggestion's justification could. Instead the justification for attributing any particular natural end to some kind of living thing would be the justification for claiming that the life form of that type of thing contained that particular result as an end of some characteristic behavior or feature. To say that it contained that particular result as an end is, effectively, to say that the form of life contains things whose function is to bring about that result, and, in the case of final ends, to say that that result has no further function in the life form of the species. In effect, the justification for claiming that the good for a living thing of a certain kind includes some particular end is equivalent to the claim that the form of life



for living things of that kind is *that way*. It is not automatically clear how the specifics of this justification would go, but that is less important than it seems. The question here is equivalent to the question of how we could come to be justified about any Aristotelian Categorical claims. What is important here is that this sort of justification, whatever the details amount to, could only be in terms of empirical facts of some sort, and not conceptual claims. It also means that the natural ends proper to living things of different kinds *could* vary quite widely, distinguishing it in an important respect from the previous account.

Along the lines of the first possibility, then, the natural ends attributed to kinds of living things would have to be justified separately from the particular life forms of kinds of living things, and by largely conceptual claims. Along the lines of the second possibility, the natural ends attributing to kinds of living things would be justified as *part of* the life forms of kinds of living things, and by reference to facts about living things of that kind. I have also suggested that the second is a better reading of Foot's account than the first, since it can be more easily read into the passages that appear to suggest the first account than the first account can into the passages that suggest the second. There is also an additional reason to read Foot as suggesting the second, rather than the first account, which is that the first account has an immediate and significant problem attached to it. The problem is that the natural ends that Foot suggests, either as a general description of the sort of natural ends particular to each species, as in the second suggested account, or as the actual ends common to all living things, as in this account, are not the ones that would result from the sort of justification that could be given to them on the first account.

This is true for two different reasons. The first reason is that it is at least *possible* to imagine living things whose form of life lacked either some particular end of the ones that Foot suggests (though this is not a species that would be likely to last long in nature), or which lacked any characteristic way of achieving that end, which comes to the same thing. The problem here is that it would pose an example of a living thing which did not have as one of its criteria for evaluation what appears at first to be a criterion for evaluating any living thing at all. And to the extent that we can imagine such a thing whichever end we picture lacking from the form of life common to members of this hypothetical kind cannot be a natural end common to all living things in as much as they are living things (since this last condition requires that these ends would apply to all *possible* living things in addition to the actual ones that exist)<sup>32</sup>. The problem here is not necessarily that an account could not be set up in this way, though I suspect it would not work, but that Foot's account cannot be set up this way. The sort of natural ends that would result are at best unclear, and quite possibly not the ones she suggests. But more importantly there is a gap between attributing some set of ends to all living things insofar as they are living things, and to living things insofar as they are members of a kind of living thing that is problematic here.

The second reason that this way of accounting for natural ends would be problematic for Foot is similar to the first, but in the opposite direction. It is even easier to imagine a kind of living thing which had all the basic natural ends that Foot mentions, but also one or more additional natural ends built into its form of life. That is to say,

---

<sup>32</sup> By comparison, if we attempt to imagine some kind of lion which lacks as a natural end some natural end that we take to be common to lions in as much as they are lions – say, that they maintain themselves by predation, or that they are viviparous – what we are imagining is something that is *not* a kind of lion at all, but some other sort of living thing that resembles a lion.

there could easily be more ends for living things of some particular kind and, as a result, that it is very likely that there would not be one set of natural ends common to all living things, or that the ends that are in common would not exhaust the relevant ends involved in evaluating members of particular kinds of living things. This is especially relevant to Foot's project since she claims that something very much like this *is* the case when it comes to human beings, namely that human beings have additional natural ends compared to other sorts of living things. It would be problematic for her project were she committed to the claim that the natural ends common to all living things exhaust the natural ends in question when it comes to any particular type of living thing.

As a result I think it is best to attribute to Foot the second picture of natural ends: that the good for living things amounts to instantiating some ideal end point for things of that kind generally. The extent to which all living things that exist tend to have a similar set of natural ends, at least at the level of generality involved in listing reproduction as a natural end, allows the passages that appear to support the view that all living things have the same, abstract natural ends to be read as consistent with this view. The passages which seem to indicate this second view can to some extent be read as consistent with the first view, but it is more difficult to do so. Finally, and most importantly, the first view would introduce into Foot's account a substantial inconsistency, which it is best to avoid.

The complete account I have attributed to Foot, then, is as follows. The natural ends that play an important role are ones that are specified the good of that kind of living thing: in other words, the good for living things of some particular kind amounts to the good life for things of that kind. Once that natural end is established it is possible to

evaluate members of that kind by appealing to some description of a particular feature of a specific member of some kind, the claim that this feature is, generally, characteristic of members that kind (as expressed by an Aristotelian Categorical), and, finally, a proposition linking that characteristic feature of members of a kind to the natural end for members of that kind. However, despite this there is a serious problem for this way of accounting for natural evaluations, and that is that the good for living things of some kind or other could not possibly play the explanatory role that it is supposed to play in this account.

## 5.

According to this way of accounting for natural evaluations, it is the good for members of some kind that *explains* why some or other characteristic feature of that kind of living thing is *important* in the characteristic life of members of that kind. Those features are important or unimportant based on whether or not they are linked in the right way to the good for members of that kind, or to the good life for members of that kind, or whether or not they are needed to achieve some part of the good life for members of that kind. On this account, however, since it is precisely the teleologically significant features characteristic of members of some kind of living thing that together make up the life form of that thing, it is also those features that explain what the good for that kind of thing amounts to. This creates a significant problem since the good for members of some kind cannot *both* be determined by the form of life for members of that kind *and* determine the form of life for members of that kind, by establishing which features are and are not significant. Or, in other words, it cannot be true both that what it is to live well is

determined by reference to the expected life of an excellent member of the species and that what it is to be an excellent member of the species is determined by what is needed for members of that species to live well. And, according to Foot's account of natural evaluations, that is precisely what is being suggested: that the good for members of some particular kind is both constituted by the form of life for members of that kind, and what determines the form of life for members of that kind. As a result, Foot's account cannot succeed.

Either what it is to live a good life is determined by general considerations unrelated to the specific form of life shared by members of that kind, or it depends on that specific form of life. In the first case Foot's account would lack the flexibility which is needed to return the sort of evaluations which she takes to be entailed by her account. The other alternative has the desirable feature of being able to return these sorts of kind related evaluations that the first could not. However according to this account of the good for living things the ends at which living things are characteristically aimed are at least partly determined by the form of life shared by members of that kind, and hence cannot simultaneously determine it.

The problem comes from the combination of two important structural features of Foot's account. First the account posits an explanatory end, namely that of a good life or the good for members of whatever kind is in question. This end is the end towards which the life characteristic of things of that kind is structurally directed. As a result, it is this end that explains why certain features characteristic of living things of that kind are or are not teleologically significant for them. The features that are

teleologically significant are the ones that are related in the right way to that end (usually that they are necessary for achieving the end). Those features can support evaluations of the individual living things of that kind that are stronger than mere claims about normalcy: they can support evaluations in terms of natural goodness or defect. The second feature is that the sort of evaluations in question are tied tightly to an account of kinds of living things. The sort of goodness or defect at stake is goodness or defect *qua* membership in a certain kind, and not generally. Since these accounts are about, or provide for, evaluations which are relative to membership in some particular kind of living thing, instead of a more general or abstract kind, the accounts are responsive to actual facts about living things, and the world in general.

The fact that the evaluations in question are related to the good for actual kinds of living things, as opposed to some general conception, requires that any account of the good for living things of some kind or other be justified with reference to specific facts about the life characteristic of members of that kind. Since these facts can differ it is, at least, possible that the good for living things of various kinds can differ. As a result, the fact that the evaluations in question are kind related requires that the good for things of that kind, which grounds them, be explained by reference to the form of life characteristic of things of that kind. However, if the account of the good for things of some kind is justified by reference to the form of life shared by members of that kind it cannot simultaneously be the explanation for why certain characteristic features or behaviors are part of the life characteristic of members of that kind. This latter role is precisely the explanatory role claimed for it: that it determines what is and is not significant in the life

of members of that kind. And, as a result, the accounts has a serious problem. In fact, any account of evaluations of this kind that contains both of the above features is very likely to face exactly this problem.

The practical effect of this problem is that accounts of natural evaluations cannot contain both an emphasis on the kind relative nature of the evaluations *and* a strong explanatory role, in the above sense, for the good, or the good life, for members of some particular kind. Neither feature is problematic on its own, only their combination. But an account that starts from the notion of a good life, or the good for members of a kind, and which is independent of considerations about the form of life for members of that kind of living thing will not have at the results that Foot, and others pursuing a similar strategy, would like. As a result the incompatibility between the use of the good in an explanatory role and the strongly kind relative evaluations is not one where either alternative taken on its own is equally useful. This is not to say that a competing sort of account starting from the notion of a general or specific end for kinds living things, divorced from questions about the characteristic form of life for things of those kinds is impossible: it is a distinct possibility. It is not, however, one that is especially relevant here since it would be an entirely different type of account.<sup>33</sup>

As a result of this it is worth considering what it would mean for an account such as Foot's if one were to abandon the explanatory role for the good for things of some kind. The first point that is important to make is that this alteration would not involve removing the idea of a good for things of some kind, or a good life for members of that

---

<sup>33</sup> It would, for example, be the sort of account according to which we could evaluate various species in general, and compare them to each other to arrive at judgments like "Snails are better than panthers". If treated as the basis for a normative account this is unlikely to have results that are as intuitive as the alternative that emphasizes the kind relative aspect of the evaluations in question.

kind entirely from the picture. It merely means that the good life would not play the role in determining what features count as significant or important in the form of life characteristic of members of that kind, or which characteristic traits of members of that kind could support natural evaluations, in the way that it did in the accounts described above. The idea of a good life for members of some kind would be a result of the account instead. The forms of life of living things are characteristically directed at various ends, some with further ends and some without further ends attached to them, and as a result they can be read as determining some general idea of a good life for members of that kind.<sup>34</sup> This notion of a good life for members of a kind of living thing is not one which plays any significant explanatory role when talking about the form of life characteristic of that kind of living thing nor could it: it is a result of that account.

In place of a notion of the good it should be possible to rest an account of natural evaluations solely on the notion of a form of life characteristic of living things of a certain kind, or at least it would have to be as possible as giving an account of this sort in the first place. Since Foot's account already depended to at least some extent on the notion of a form of life characteristic of living things of a certain kind this should not be difficult to imagine. The main function of the good, or a good life for things of some particular kind, in the account that I suggested above was to distinguish between features of members of some kind of living thing that were characteristic of that kind of living thing but not significant or important for them, and ones which were both. This role is one which can be covered easily by an account of the functions of some of the features

---

<sup>34</sup> This is not necessarily much different in effect from Foot's suggestion that the good for living things amounts to instantiating the form of life characteristic of their kind.



characteristic of living things. What is important in picking out certain features as teleologically important in the characteristic life of members of that kind is not that the features are important for some *particular* final end: it is just that they are important. That is to say, what is relevant is that the feature in question has some function, and not what the eventual result of its performing that function might be down the line. As a result there is no need for a particular end to play an explanatory role here, merely the network of characteristic parts with their interlinking functions that make up, for living things, the life characteristic of their kind.

## **2. Functions**

I have argued that Foot's account of natural evaluations fails because it requires both that what it is to be a truly good member of a kind is to be the sort of thing ideally fit to live the good life proper to members of that kind, and also that what it is for some life to be the good life for members of that kind is the life that a truly good member of that kind is most fit to live. Since both of these cannot be usefully true it is necessary for one of them to be dropped from the sort of account that she proposes. I have also argued that it is the former that should be abandoned. To arrive at an account of natural evaluations useful for normative ethics the idea of a good life for things of some kind or other should be treated as a result of the account, following from the idea of a good member of some kind. This change makes understanding how some features or parts of living things have functions important. And it is especially important because there would not be some final end to justify claims about them. As a result, it is necessary to spend some time getting clear on what sort of thing is being appealed to here.

I will begin in section 1 by discussing what Foot has says about functions, but conclude that what she says will not be helpful for my purposes. I will then suggest that there is reason to refer to the literature on function attributions in biology. In section 2 I will consider two possible approaches to explaining function attributions, but argue that they cannot serve my purposes. In section 3 I will present in greater detail the best contender for a way to explain function attributions that can underlie claims about goodness and badness in human beings, but in section 4 I will argue that that approach lacks the flexibility required. In section 5 I will conclude by arguing that we can understand functional evaluations as grounded in evaluative criteria for membership in

certain kinds, namely functional kinds.

## 1.

Foot addresses the issue of functions relatively briefly in Natural Goodness, treating talk of functions as largely synonymous with talk of goods. In a footnote on page 32 she does spend some time distinguishing what she takes to be our common usage of function talk from discussions of adaptations in evolutionary biology, arguing that

It is easy to confuse these technical uses of words such as 'function' and 'good' [in evolutionary biology] with their everyday uses, but the meanings are distinct. To say that some feature of a living thing is an adaptation is to place it in the history of a species. To say that it has a function is to say that it has a certain place in the life of individuals that belong to that species at a certain time.

It is not entirely clear what sense of function is the everyday sense, but the most likely option given how Foot talks about functions is that some feature or part of a living thing has a function due to its relation to the good for things of that kind. This is likely given that the sentence which prompts the footnote just quoted equates the two: "In other words, 'What is its function?' or 'What good does it do?'"<sup>35</sup> If this is the case then the general view here can be taken, roughly, to amount to the view that a function is identifiable with some good that is brought about.<sup>36</sup>

---

<sup>35</sup> Foot, 32. These are both intended as equivalent to the question "What part does it play in the life cycle of things of the species?"

<sup>36</sup> There is another possibility here relating to how Foot's general account can be read. It is possible to read the link between the good life for things of some kind and Aristotelian Categoricals in a more limited fashion than I have: giving the account of functions suggested here and nothing more. However this would require a more than normal amount of interpretive charity given what she has argued, since it is not an especially plausible reading. More importantly this possible way of reading Foot would still leave her account susceptible to the critical argument I have made in the preceding section.

While this may be sufficient for Foot's purposes, however, it is not helpful for me. As a general account of what we mean by function, even in the relatively limited case of talking about the functions of parts of things, there is an obvious objection. The central problem here is that it does not seem immediately impossible that there be parts of things or things whose functions do not bring about some good but instead some bad, or perhaps simply something that is neither good nor bad but neutral. This may not be a problem for Foot's account, given the large role that the notion of the good for things of some kind plays. But it is a problem if there is no final good determining what is and is not a function, even though it might be the case that all the functions of parts of living things do bring about some good for the living things in question. I have just argued that appealing to the idea of a good for living things of some kind in order to cash out the notion of a function is problematic. As a result it is better not to resort to this way of characterizing functions.

Luckily there is an extensive literature on functions, and, especially, attributing functions to parts of living things within the philosophy of biology. So if there is a need for a more substantial account of functions the first place to look for one is clear. If a line can be drawn directly from some plausible account of functions of parts of living things within the philosophy of biology to the problem at hand there might be a convenient grounding for the sort of natural evaluations I have in mind, since it would be less likely to appeal to a notion like the good life for things of a particular kind. With an independent account of what it is for some part of a living thing to have some particular characteristic function it would be possible to arrive at a means for evaluating living

things as defective or excellent members of the kind to which they belong. And that could then be used to arrive at a general account of the good for things of that kind. Unfortunately, however, an account of the sort I would need cannot be found within that literature. The reason for this, however, will turn out to be helpful in arriving at an account of functions which will be useful.

Before examining the accounts of functions available within the general literature, though, it is necessary to make an important distinction. The discussion of functions is split into two very different projects. The first is concerned with a primarily descriptive account of the uses to which practicing biologists put the notion of a function, and the various different senses the word ‘function’ has in biological research.<sup>37</sup> The second project is concerned with the metaphysical status of functions generally, and especially with what can be called *proper* functions. While the two projects occasionally overlap, it is *only* the second one that is relevant here. Proper functions are distinct from other sorts in that they are attributed directly to the thing in question, as a result of the sort of thing that it is. They are not dependent on how the thing is being used, or what it happens to be doing. Something can have a particular *proper* function and be incapable of performing it: it can *malfunction*. Other ways in which we use the word do not have this feature, and are often referred to as derivative functions. For example we speak of something or other functioning *as* something of a particular kind even though it is not, like we say that a nutcracker is functioning as a paperweight. This is the sense in which these sorts of functions are derivative of proper functions: they are describing cases where something is

---

<sup>37</sup> Such as, for example, a description of the different ways in which term is used, or an application its use in addressing other questions. Wouters, A (2003) provides a good example of the former, and a truly excellent example of the latter is the brilliant analysis in Love, A. (2007).

used or operates as a thing of a different kind. This is clear from the sorts of evaluations that result: we would say that the nutcracker is a good paperweight. When speaking of proper functions, however, we do not just describe what some particular thing happens to be doing, instead we attribute to some object some particular role, use or purpose *independent* from what it actually is doing, or how it actually is being used. As a result of this the fact that something has a proper function does not necessarily mean that it does, or even could do the function attributed to it.

While both of these projects are generally considered part of philosophy of biology it is only the one focusing on proper functions that is important here. The first approach to the question would only be relevant to the extent that practicing biologists do use the notion of a proper function. It is certainly possible that they do this to some extent, or that they ought to do it. However it is only a proper function that would entail the sort of kind relative evaluations that are important. That is in part what sets proper functions apart from other uses of the term. As a result I will be focusing entirely on the accounts of proper functions.

## 2.

There are two main things that an account of proper functions needs to explain: how it can serve as the basis for evaluations, and how it can be used in explanations. Attributing a proper function to something entails a variety of evaluations. Even if we are simply saying that a pen is functioning as a doorstop, we can still ask whether the pen is or is not doing so successfully, and if so what about it makes it good or bad at being a

doorstop. With proper functions the evaluations are even simpler: they are simply evaluations of some thing relative to its proper function. This is why they allow for the possibility of malfunctions, and also why I am interested in them in the first place. Attributing a function to something has an important second feature as well. It tells us something about the thing in question, or is useful in explaining something about it. What that explanation is varies with the type of function that is attributed to it. Saying that a pen is functioning as a doorstop explains why the pen is on the floor, and what it is up to. Attributing a proper function to something is even more significant: proper functions can explain why things of that kind exist. The function of a hammer (hammering nails) is a crucial part of the explanation for why hammers exist. They exist because we made them to hammer nails. This is also true for things that are not artifacts, but without being able to appeal to some intention behind their creation it requires a more involved story.<sup>38</sup>

The explanation for how proper functions are related to those evaluations and explanations will have to be given in terms of kinds. The explanations that proper functions entail are ones that relate the kind generally to the specific object: hammers exist because of our intentions in creating a certain kind of thing. This is why attributing a particular function to a defective or malfunctioning hammer can still be informative. The evaluations that are entailed by proper functions are more obviously related to kinds: they are kind-relative evaluations. The function of a hammer is what allows us to judge whether or not some hammer is a good *hammer*. It is what allows us to talk about what

---

<sup>38</sup> The exact way in which this sort of explanation is cashed out, even speaking only of proper functions, differs in different accounts of what we mean by 'function'. This is not especially important here.



features of hammers make them better or worse *hammers*, and so on. Something like a hammer has its specific function when it is considered *as a hammer*, or to the extent that it is a hammer and not independently of that. After all, a hammer with a broken handle could simultaneously be a bad hammer and a good pestle. If something could have a (proper) function independently of being considered as a member of some kind, it would not entail the sort of evaluations that come from attributing a function to something. So the question is how to understand how things have a function as a result of the kind of thing that they are, rather than what they are doing or how they are being used.

In the discussion of proper functions there is relatively little disagreement about the best kind of account. A etiological approach to the question, one which is focused on the history, often the evolutionary history, leading up to the object in question, is the most common. Bekoff and Allen (1995) refer to it as 'the standard line'.<sup>39</sup> However it is still worth considering two other ways of making sense of proper functions before discussing etiological accounts. Both of these other sorts of accounts often show up in the separate and more descriptive project in the philosophy of biology that I mentioned earlier. However they are not as useful when it comes to dealing with proper functions. The first of these bases its explanation of the nature of proper functions on the effects they have, and the second on the systems in which they are found.

The simplest version of effect based accounts claim that for some object to have a function is simply for that object to bring something about, either on its own or when used appropriately. This is the simplest way of identifying a thing's proper function:

---

<sup>39</sup> Bekoff, M. & Allen, C. (1995). Wright, L. (1973); Griffiths, P.E. (1993); Neander, K. (1991a); Neander K. (1991b); Godfrey-Smith, P. (1994) are some examples of the way this general approach works out in practice.

referring to its consequences or its use. Because of this simplicity, though, it lacks the resources to distinguish between proper functions and any other functions, or even accidental results that might occur. As a result this might count as an explanation of certain uses of the word function, especially when we say that something functions *as* something else, but does not succeed as an account of proper functions, since those are non-accidental in character.

There is a more sophisticated, and more common, version of this account that adds that the effect based account takes holds that the result that is brought about must be a good one, or in the case of living things, that it must benefit the sort of living thing in question. This change has the benefit of making some distinction between accidental (non-beneficial) and non-accidental (proper) functions. The distinction here makes the most sense when it comes to parts of living things: their functions amount to whatever it is that they do that benefits their possessor, and whatever other effects they have, neutral or harmful, are accidental. However the distinction itself still fails to capture what is needed from an account of proper functions. Accidental results of some object or part of a living thing can be as beneficial as non-accidental ones, like in the stories in which a soldier's life is saved by an appropriately placed bible. It is also possible to imagine that some part or object's proper function could lead to a bad outcome in certain situations, or perhaps even generally. As a result this more sophisticated account does not fully solve the problem possessed by the less sophisticated version – though it is a little better. This sort of account is also likely to be what Foot has in mind when speaking about functions and so inherits the same problem.

Unlike effect based accounts which rely on the link between a function and some effect or good, system based accounts focus on what surrounds the thing in question and how it interacts with those other things. The context in which function attributions are most appropriate is one in which various simpler parts make up something that can be recognized as a sort of system set up to achieve certain ends, even if this is simply preserving some general state.<sup>40</sup> Living organisms are good examples of this sort of thing. According to this sort of account having a function amounts to being implicated in the system's general operations, and having some specific contribution to those operations or goals. For example, attributing a particular function like "pumping blood" to the heart is the same as saying that the heart does pump blood, that pumping blood occurs within an organism with a systematic organization, and that the motion of blood through that organism contributes significantly to the operations of the system.

Paul Davies, in Norms of Nature argues for a system based account . He argues that:

"Where "A" refers to the analysis of system S into components, and where "C" refers to the systemic capacity we wish to explicate, item I has systemic function F if and only if:

- (i\*) I is capable of doing F,
- (ii\*) A appropriately and adequately accounts for S's capacity to C in terms of the organized structural or interactive capacities of components at some lower level of organization,
- (iii\*) I is among the lower-level components cited in A that structurally or interactively contribute to the exercise of C,
- (iv\*) A accounts for S's capacity to C, in part, by appealing to the capacity of I to F,

---

<sup>40</sup> Probably the most important of which is found in Cummins, R. (1975).

(v\*) A specifies the physical mechanisms in S that instantiate the systemic capacities itemized.”<sup>41</sup>

An object can have a function, he argues, if it is embedded in the right sort of system. That system is one which is organized in a hierarchical fashion, with higher and lower levels of organization, and has some overall capacity or thing that it does. Within that system the object sits at a lower level of organization, and does something that lets the overall system do what it does. In this specific case the reason that the heart has a function, pumping blood, is because of how it is connected to the circulatory system. The circulatory system has a certain systematic capacity: it distributes necessary resources like oxygen or glucose through the body. The heart is part of that system, along with veins, arteries, capillaries, and so on. The overall analysis of the circulatory system that divides it up into veins, arteries, capillaries, the heart, and so on explains how it distributes those resources through the body. The fact that the heart pumps blood is a significant part of that explanation, and, finally, the heart in this case does actually pump blood throughout the body. As a result we can reasonably say that some particular heart has a function, namely pumping blood through the body.

Specifying the properties of the right sort of system in which this can be made sense of is tricky. Davies has relatively little to say about how one might go about identifying what systems are and are not organized in the appropriate hierarchical fashion. In general though this approach does have less trouble with the distinction between accidental and non-accidental functions than effect based accounts. A carefully placed bible may have some significantly beneficial effect for a soldier, but it is not part

<sup>41</sup> Davies, 89. In this book Davies explicitly argues for an account based on what Cummins (1975) argued.

of the function of the bible to stop bullets because, according to system based accounts, there is no general system of protection from bullets with strategically placed bibles as a part. If there were it would not be odd at all to attribute a function like “stopping bullets” to a bible, in precisely the way we do attribute it to Kevlar vests. This is not to say that these accounts are immune from worries of this sort, as parts of some system might in certain cases manage to bring about a benefit for the system in a way unrelated to their proper function.

System based accounts do have trouble dealing with a different feature of our talk about *proper* functions, though. An object can simultaneously have some particular (proper) function and yet be incapable of performing it. Since the systems accounts tend to focus on what an object actually does in the context of the system in which it plays a role, it is difficult for them to account for proper functions, since attributing a proper function to something entails normative claims. Some object can fail to perform its function while still having that function, or function well or poorly. Bibles may in some context function as a sort of body armor. However since bullets are not reliably stopped by objects like bibles, and usually go straight through them with very little trouble, this would be a case where an object’s function was something that it could not manage. Any account of proper functions which includes a requirement that the object in question actually does perform the function in question has significant trouble accounting for this feature of functions. This is usually a significant problem for effect accounts as well. This makes systems based accounts less effective when it comes to accounting for proper functions. It does not necessarily make them less effective when it comes to giving an

account of how function attributions are used in biology, though, which is where it is more common. Davies himself is clear that he is pursuing that latter project. He is not concerned with how to explain how proper functions might be used in biology, except to argue that they are not.<sup>42</sup>

### 3.

It is unusual to run across variations of these two views when considering proper functions: the most common accounts are etiological accounts.<sup>43</sup> The two previous sorts of accounts were primarily built around the intuition that a proper function for some object or other is something that object does or brings about. This is certainly something that is likely to be true of most of our function talk, which is why they are more common in accounts focused on the way the term is used by biologists. However it is not the only thing that can be said about proper functions. Etiological accounts are centered around a somewhat different idea of what is essential to proper functions.<sup>44</sup> These accounts start from the intuition that function attributions are *explanatory*, or, at the least, that they entail explanations. Functions are taken to explain the presence of the object to which that function is attributed, not simply to explain what it is up to. When we say that the heart's function is to pump blood, or the kidney's function is to filter the blood, or that the

---

<sup>42</sup> "The aim of a theory of functions is to conceptualize functional properties in ways that fit the methods and postulations of our best natural sciences. We thus should be guided by the role or roles that function attributions play in the course of inquiry." *ibid*, 91.

There is some debate on whether or not the second project excludes the first. Millikan, R.G. (1989a), Amundson, R. and Lauder, G (1994), and Godfrey-Smith, P. (1993) argue that accounts like Davies' and etiological accounts are consistent but pick out different senses of function.

<sup>43</sup> As per footnote 40.

<sup>44</sup> This usually but does not always involve taking the account to result from a form of conceptual analysis, as in Neander, K. (1991a), though Millikan (1989b) argues that it should be understood only as a theoretical definition.

carburetor's function is to aerate gasoline we are explaining why we *have* hearts and kidneys, and why our cars have carburetors.

The first philosopher generally credited with advancing this position is Larry Wright, in his article "Functions".<sup>45</sup> There he argues that

it is ordinarily supposed that a function explains why each of these things is the case. The function of the quills is why porcupines *have* them, and so forth.

Moreover, the kind of explanatory role suggested by both of these considerations is not the anemic "What's it good for?" sort of thing often imputed to functional explanations. It is rather something more substantial than that. If to specify the function of quills is to explain why porcupines *have* them, then the function must be the reason they *have* them. That is, the ascription of a function must be explanatory in a rather strong sense.<sup>46</sup>

In the case of artifacts like carburetors the source of this explanatory power is clear: the function of some artifact explains why it was made, and made that way. What we are expressing when we attribute a proper function to some object is our intention in making the thing, and in placing it into the context in which it is found. When it comes to parts of organisms, however, the source of the explanatory power of function talk is unlikely to rest in the intention of some creator.

This is not necessarily a problem, however. In the case of artifacts the source of the explanatory power rests in the intention of a creator because artifacts are created

---

<sup>45</sup> Subsequent proponents of the view have advanced slightly different versions of this account. For example Godfrey-Smith, P. (1994) argues that it must be the recent history involved rather than a broader historical notion, and Bigelow, J.B. And Pargetter R (1987) and Enc, B and Adams, F (1992) argue that rather than the past, recent or not, it is the future that is important. The latter is not *etioloical* since it is not backwards looking but the differences are unimportant *here* and Wright can serve as an example of the way this general approach addresses the question.

<sup>46</sup> Wright, 155.

intentionally. Natural selection is also capable of explaining the creation of living things, and as a result is also the reason for the explanatory power of function attributions: it is not necessary that the creation be mediated by specific intentions. Just as with artifacts that are created intentionally there is something about hearts that explains the fact that they evolved to be the way they are in living things, and that they are maintained in those living things. It is that they pump blood. The historical story is different in certain respects, but the sort of explanation is similar. What is important is that there be some procedure of selection to make function attributions explanatory, but not that it involve intentions. What is important is that the function of a particular heart is the historical explanation for why the creature itself has a heart.

As a result, Wright argues, we can take the difference between artifacts and organisms to be unimportant as far as an account of proper functions is concerned. It does matter when it comes to specific explanations, since the historical stories will be different, but not in what it is to have a function generally. Proper functions, Wright argues, can be understood as follows:

The function of *x* is *z* *means*

(2) (a) *x* is there because it does *z*

(b) *z* is a consequence (or a result) of *x*'s being there.

The first part, (a), displays the etiological form of functional ascription – explanations, and the second part, (b), describes the convolution which distinguishes functional etiologies from the rest.<sup>47</sup>

The first condition cashes out the explanatory force of function attributions, and the second condition specifies the right direction for the sort of explanation. The reason for

---

<sup>47</sup> Wright, 161.



the second condition is that a pair of things can often satisfy the first condition in both directions, but the sort of explanations that we get from attributing a proper function to something are not symmetrical in this way. To use Wright's example, one function of hemoglobin in blood is combining with oxygen, and it is true that the hemoglobin is present in the blood because it combines with oxygen.<sup>48</sup> It is also true that oxygen is present in blood because it combines with hemoglobin. But combining with hemoglobin is not the function of oxygen in the body. The sort of explanation given by the first sentence, that hemoglobin is in the blood because it combines with oxygen, is a different sort of explanation from the second sentence. Both claims are true, but they do not generate the symmetrical explanations: hemoglobin is not a result or consequence of oxygen in blood. And the function of oxygen in the blood is not to combine with hemoglobin.

Etiological accounts have a number of advantages over the other accounts when it comes to proper functions. Their main strength is that they have an effective means of drawing a clear distinction between accidental and proper functions. This distinction is both essential to an account of proper functions, and also an especially tricky one, as demonstrated by the difficulty the other accounts have with it. In etiological accounts this distinction is cashed out in historical terms: accidental functions lack the sort of history that proper functions have. It may be beneficial to us that our hearts produce a rhythmic sound for diagnostic purposes, but there is no story about the development or maintenance of hearts that includes the sound they produce as part of the explanation for their presence. The historical story we would tell about human hearts, and the

---

<sup>48</sup> Wright, 159-160.

evolutionary explanation for their presence would, however, be likely to make a great deal out of their tendency to pump blood. Because of this difference, proper functions generate explanations for the presence of the object in question and accidental ones, lacking that history, do not.

A result of this is that etiological accounts are very good at making sense of one of the central intuitions many have regarding biological functions, namely that they provide an important *sort* of explanation. If what makes some particular activity the proper function of some part of a living thing is that that activity explains, in terms of natural selection, the presence of that feature, then knowing the proper functions of parts of a living thing is very useful knowledge. It amounts to knowing why those features or parts are present, which is both useful in its own right and also directly relevant to further explanations. This link to the ancestry of the possessor of the trait also accounts for the fact that some object can both have some function and be incapable of performing it, which was a problem for both the effect and system types of accounts. According to etiological accounts, this is due to the fact that what sets the function for an object is not any actual feature of that object itself, but instead facts about the object's predecessors. As a result it is entirely possible that some particular heart could both have the function of pumping blood, as a result of its ancestry, and yet also be defective with regards to the pumping of blood.

Finally etiological accounts have relatively little trouble accommodating the intuitions about proper functions underlying the other two types of accounts. While there is no necessary link to some sort of beneficial effect for the organism in question, there is

at least very likely to be one. Evolutionary benefit is not always personal benefit, or at least the intuitive notion of what benefits an organism diverges somewhat from it on occasion. And there is always the possibility of a malfunctioning part. But evolutionary and personal benefits are nonetheless very close, and where they do diverge it is not automatically clear that an etiological account would give a worse answer. Along similar lines, the sorts of objects that could have a proper function at all on an etiological account are by necessity bound up in something approximating the sort of system that a system based account might require. In the case of organisms subject to natural selection this is a result of the functional organization necessary for something to reproduce. And in the case of objects where the selection is intentional rather than natural the addition of an intentional agent insures that there is a larger functional context in which the artifact plays some role. As a result the intuition standing behind the systems accounts is also, to some extent, borne out by an etiological account.

#### 4.

Etiological accounts tend to be the most common. However there is also a problem in appealing to them when it comes to functions that are important for my project. Etiological accounts are useful when it comes to the philosophy of biology, and I have no intention of attacking their use in that context. But they do not work when it comes to talking about the sort of functions that can ground a moral theory. The reason that they will not work is directly linked to the reason why function talk is important here. Particular objects only have proper functions as a result of being some kind of

object or other; that is why they can be said to have the function even when they are not capable of performing it. Applying to objects only insofar as they are objects of some kind is why proper functions entail kind relative evaluations. But it is not necessarily the case that all kinds of things which have functions have the right sort of history standing behind them, or that two objects with the same function have the same history. I will argue that this poses a problem for using an etiological account to explain the picture of human nature needed to ground claims about ways in which human beings can be good or bad.

The root of the problem is that any one object can be a member of any number of different kinds of things, depending on how it is described. An object can be falsely described as being a kind of thing that it is not but that does not mean it does not belong to multiple different kinds at the same time, especially since there are more and less general kinds of things. For example ballpoint pens and fountain pens have different (though similar) functions, since the way in which they are designed to put ink onto a page is different. But any two pens have the same function, even if one is a ballpoint pen and one is a fountain pen. In the second case the kind of thing, and the corresponding function, is more general and covers a broader range of differences than in the first case. And this can scale both upwards and downwards, since we could also talk about writing implements as opposed to other artifacts, or fountain pens that use cartridge fillers as opposed to piston fillers. Etiological accounts depend on the historical story behind artifacts and parts of living things to explain how they can be said to have functions. But difference in the way that fountain and ballpoint pens achieve their function makes the

extent to which they share a history more complicated than talking about fountain pens or ballpoint pens specifically.

There is a way to deal with this. It could just be that the historical stories behind both ballpoint and fountain pens are part of the general story of how we came to have pens<sup>49</sup>. The selection pressures on both are similar, and at the level of generality involved when talking about them as pens, identical. As a result, the function of any pen is identical to any other pen: to serve as a writing implement that applies fine lines of ink to pages. This is the same sort of story that we would tell to explain the biological functions of things like the heart. There are a large variety of different kinds of hearts: iguana hearts are different from human hearts, and function differently to some extent. But there is a unified historical story to tell about hearts in general, starting from the first recognizable heart and developing on from there into the wide variety of different things that we call hearts. So while the specific functions of specific kinds of hearts vary, there is still a story that we can tell generally about the development and maintenance of the heart. And as a result anything that is a heart has a particular function based on its being a heart. The explanation for why things that are human hearts exist is different from the explanation for why things that are lizard hearts exist. But the explanation for why there are things that are hearts is the same. The overall history of the heart includes both even though some of those hearts are lizard hearts and some are human hearts, it is simply one with a great deal of branching.

Since common ancestry is a feature of living things, and to some extent artifacts,

---

<sup>49</sup> Though it is still somewhat tricky: the original ballpoint pen was invented to replace the pencils used to write on leather or wood, and resembles a copying pencil as much as it does a fountain pen. J.J. Loud (1888)

then it seems plausible to suggest that when it comes to the functions of parts of those animals there should be a common history behind them to explain their functions, no matter how generally we are talking. But this is not necessarily true. The problem is that some parts of living things are the result of convergent evolution. For example, both hedgehogs and hedgehog tenrecs have spines for protection, and roll up into balls when threatened. But their common ancestor was quite a long time ago, and did not have spines; they are not even in the same taxonomic order.<sup>50</sup> Hedgehog tenrecs are more closely related to manatees than hedgehogs. The historical stories explaining the presence of hedgehog and hedgehog tenrec spines are unconnected: there is no common spiny ancestor that would connect the spines of a hedgehog with that of a hedgehog tenrec. But, just like pens, they are the same sort of thing: they are *spines*. And as spines they have the same function, and achieve it in the same way. When it comes to kind relative evaluations spines of hedgehogs tenrec and hedgehogs, as spines, are made excellent or defective by the same things. We can say that, as far as their spines go, some individual hedgehog is better off than some other individual hedgehog tenrec. And there is no common history to cash out that comparison

The closest that an account of functions based in common ancestry could say when it comes to spines would be to say that the hedgehog's spines are good at fulfilling their (hedgehog spine) function, and that the hedgehog tenrec's spines are not as good at fulfilling their (hedgehog tenrec spine) functions. This would be close enough to look like a comparison of the two sets of spines, and we could make this kind of evaluation if we choose to. But it is hardly what we are saying when we compare two sets of spines,

---

<sup>50</sup> Nikaido, Masata, Ying Cao, Norihiro Okada, and Masame Hasegawa. (2003)

both of which fulfill the same role in the same way in the respective animals' lives. Even if we were saying this, however, this would only resolve the problem if we could only attribute a function to *hedgehog* spines and *hedgehog tenrec* spines, but not to spines in general<sup>51</sup>. But there is no obvious reason why we could only ever attribute a function to an object under one specific description and no other one. And even if there were it is unclear how one particular description could be chosen at the expense of the others. The problem here is that while living things have a common ancestor their parts do not necessarily have one, and the functions are being attributed to parts.

The closest that an account of functions based in common ancestry could say when it comes to spines would be to say that the hedgehog's spines are good at fulfilling their (hedgehog spine) functions, and the hedgehog tenrec's spines are not as good at fulfilling their (hedgehog tenrec spine) functions. So the comparison would be equivalent to saying that a particular panther has better teeth than a particular snail's shell. We can make this sort of evaluation if we choose to, but it is hardly what we are saying when we compare two sets of spines, both of which fulfill the same role in the same way in the respective animal's lives. Even if we were saying this, however, this would only resolve the problem if we could only attribute a function to *hedgehog* spines and *hedgehog tenrec* spines, but not to spines in general. But there is no obvious reason why we could only ever attribute a function to an object under one description and no other one, especially if it was a similar but more general description. In other words, some particular thing can have one function under a certain description, and a different one

---

<sup>51</sup> We could argue that in this case the story about how spines came to be and were maintained is simply general enough to encompass both groups of species given the similarity of their ecological niches. But there are many other spiny creatures, even if the similarity between them is less striking than with hedgehogs and hedgehog tenrecs, such as porcupines, lion fish, stingrays, sea urchins, and so on.

under a different description. These two descriptions can be relatively similar, such as being more or less general, but they are still distinct and the functions attributed to the object under those descriptions differs.<sup>52</sup>

As a result, etiological accounts are not useful when trying to account for kind relative evaluations in general. While they can explain some of them, there are other kinds which are not the result of a single historical lineage. And for those an etiological account comes up short. As a general account of proper functions an etiological approach is less fruitful since it applies most naturally to kinds that are picked out by means of a story about their development over time, and not all functional kinds are this way. This is not to say that etiological accounts of proper functions are useless, just that they are accounting for a less flexible understanding of functions than is useful for talking about goodness and badness in human beings. Within the context of the philosophy of biology they may be quite useful and accurately represent how the specific functions of parts of living things work. The more general kinds that could cause trouble, E.G., 'eyes' or 'spines' are of relatively little interest compared to the specific types of eyes or spines being investigated. I do not intend to explore that question. However, etiological accounts end up being too narrow to account for proper functions in a way that would be useful for grounding an account of the virtues.

## 5.

None of these approaches can fully account for the sort of functional evaluations involved in moral evaluations of human beings. They are better suited for the projects in

---

<sup>52</sup> They could also be significantly different descriptions, and have significantly different functions, but that is less interesting and certainly less relevant to biology.



philosophy of biology. It seems reasonable, then, to start by asking what would be needed for the specific kind relative evaluations in question, and going from there. When we say that an object is a good screwdriver, or that one hedgehog has better spines than another, or that a person has excellent lungs we are not saying that the object is good *and* that it is a screwdriver. There is something about being a screwdriver, or a hedgehog's spines, or a person's lungs that makes the evaluation possible. Having the function (and evaluated accordingly) part of what it is to be one of those things, or to be something of that particular kind. And having the function is not incidental to being an object of that kind: things can fail to be objects of some kind or other if they do not function in the right way, or to live up to some standard. If having some functional standard against which objects can be evaluated is essentially part of what it is to be objects of that kind, then it is part of how objects are sorted into, and out of, that kind. Functional evaluations do not assert that some object itself is good or bad in some general or abstract sense. They assert that an object is a good or bad instantiation of some kind of thing in some respect. And that means that the standard against which they are evaluated is not simply added on after the object has been identified as an object of that kind, but that living up to the standard to some degree is part of what identifies the object as being of that kind. From the function of the hammer, for example, we can conclude that a particular hammer is a good hammer, or a poor hammer. We can say that having a hard head makes hammers better, and that having a soft one makes them worse. And the worse a hammer is the closer it is to not really being a hammer at all. A hammer sculpted out of butter is not much of a hammer at all, because of how poorly suited it is to hammering. There are

marginal cases, certainly, and hammers that are defective enough to be useless can still be hammers. But this ambiguity is characteristic of evaluations, and what we should expect from functional kinds. If being a certain kind of thing entails some evaluative criteria, which it does for functional kinds, then that kind has an essentially evaluative nature. And if it is essentially evaluative then its application in the world would be as well. Even if we might prefer otherwise or think that there might be some strict criterion for membership, we do not go about using that criterion when we call things hammers.

What I am suggesting is that we can understand functions, and the sorts of evaluations that they entail, as part of the criteria for membership in some kinds. And so an account of what it is for something to have a function is the same as an account of how that thing is a member of some (functional) kind. Attributing a function to something is equivalent to taking it to be a member of some general kind. Without more to say about what it is to be a functional kind, however, this is just naming a problem rather than addressing it. The word 'function' has simply been replaced by the word 'kind' in the question. When we divide up some kind of organism into a collection of various kinds of parts, with functions, what justifies our picking one way of dividing it up as opposed to another? This is especially pressing if this dividing up is to capture moral, and not strictly anatomical, evaluations.

When it comes to artifacts the question is easy. They were invented, intentionally. The different kinds of artifacts, with their characteristic functions, were intentionally invented for just those purposes. The intention behind the creation of any *individual* hammer can vary wildly: it could be mad with the intention of selling it for a

profit, or to demonstrate skill at manufacturing, or for revenge. But what it is to be a hammer, including the essential function of hammers, is something that we invented as well. So what justifies dividing up types of artifacts into different functional kinds does not require *discovering* what various kinds of artifacts there are in the world. The objects in question only show up after recognizing the existence of the kinds, instead of preceding them. Before someone invented the hammer there were no hammers. But we have not invented organisms in the world, or parts of them like livers or lungs. Instead those objects fall under functional kinds instantiated by things that preceded our taking them to be members of a particular, functional kind. It is unlikely that any lungs that exist at the moment existed before we recognized that lungs were a kind of thing. But there were lungs before we were taking there to be objects of that kind. So whatever justifies our taking there to be things like lungs, with their proper functions, it is not that anyone invented a new type of thing, with that particular function. And without some justification for taking there to be something like lungs, functions and all, it is hard to see what would justify calling something in the world a lung.

Whatever that justification is, though, it could not be in terms of truth or falsity. A way of dividing some kind of thing into different functional parts is not something that can itself be true or false, though it can be frivolous. Whether some actual thing is or is not a lung, or whether some particular organism has or does not have lungs are claims that can be true or false. But that there is a kind of thing, lungs, which applies to some things or parts of things in the world is a different matter. One way of dividing up some kind of object into parts cannot be more or less true than any other way of dividing it up

into parts. It is what we say with or about those parts that is true or false. Some ways of doing this may be better fitted to dealing with things in the world, or put us in a better position to talk about it. But we could, for example, say that instead of quite a few organs the human body only had two: the liver, and the not-the-liver. And as far as it goes this would be as true as saying that it had more than a hundred different organs. We could truthfully say, after dividing things up this way, that someone was having trouble with their not-the-liver, or that they had fallen and damaged one of the lower parts of their not-liver, and so on. The problem with saying that the human body has two parts is not that it is false, it is that this way of dividing up the body has little to no *use*. Our reasons for any division into parts, or to prefer one to another, are practical reasons. They are justified or unjustified by reference to some use we have for them, and not whether or not they correspond to the extent that they reflect some actual division in nature. There might be such a thing, and it seems plausible that kinds that are useful for some ends, such as scientific explanations, rather than others reflect it. But the possibility of this is irrelevant since what I am talking about are the kinds that we commonly use. What makes dividing up the body into the liver and the not-the-liver bad is not that it gets something wrong about the world but that it is dumb. Our ways of dividing up what we find in the world, including into functional kinds, are justified by their usefulness not just for the uses that might reflect a metaphysical distinction, but any number of other uses. Once we have divided things up into parts or taken some sort of thing to be made up of various other kinds of things, we can use those kinds to say true or false things no matter how poorly they divide up the world. Any metaphysical debates can be left aside here. What I am

talking about are the kinds we are using for practical purposes, and so the justifications for taking there to be various functional kinds like lungs or livers are also practical ones., relating to our ability to talk about things in certain ways, with certain goals.

There is an important worry here that this might compromise the ability to talk about human nature in a fully Aristotelian sense. It introduces a certain contingency into the account of what it is for there to be a specific kind of thing, human being, or for us to understand what it is to be a human being is one sense rather than another, which amounts to the same thing. But this is contingency is a limited one. It applies to the way in which we are talking about actual human beings, specifically the physical bits of the world that we are calling human. Our interest in dividing up the world into kinds is that doing so lets us say true or false things about the world, including true or false evaluations of things in the world, in the case of functional kinds. What we say using the terms in question is made true or false by the actual facts in the world, including, when it comes to functional kinds, the evaluations we make of individual human beings. So our interests in saying things certain kinds of things about the body in medical contexts make picking out some functional parts, hearts for example, useful. This is because, medically speaking, it is important sometimes to be able to accurately evaluate health in hearts, and especially to be able to do it in a straightforward way instead of having to talk about a problem with someone's not-the-liver. But whether or not we are right to say that someone has a weak heart is a not the result of our interest in being able to say that, it is a result of the weakness or strength of their heart.

The existence of the things that we are talking about when we talk about hearts,

and their characteristic features is still somewhat contingent, as is our reason in taking them to be strong when they pump in a certain way and weak when they do it differently. But this is a result of the fact that the way human beings work is itself contingent, on any way of dividing us up into functional kinds. The existence of human beings at all, whatever our nature, is a historical contingency, as is the fact that we have the distinctive features that we do. Evaluations of human beings qua human cannot be divorced from that, or taken to be a feature of the universe prior to our existence. This may cause a further worry regarding the moral status of these ways of evaluating human beings. But the importance of understanding human nature in one sense rather than another, and the distinctively moral features of certain evaluations of human beings is the topic of the following chapter, and will be discussed there in more detail.

As a result the difference between artifacts and living things is not as important as it seems. The objects that fall into one or the other general kinds differ, in that some are created by us and others exist independently. But in both cases the way the object is described, or what kind of object it is taken to be are what determine its function. And any object can fall under any number of different kinds, depending on how it is described, and be evaluated with reference to any number of different functions as a result. Our decision to divide things up into various kinds of parts is a practical one. Taking an organism to have a certain set of parts, with functions, is as intentional a process as creating a new kind of artifact. The fact that the objects that we take to fall under those general kinds existed prior to our dividing them up that way is not actually important. And so identifying parts of living things according to functional kinds is no stranger than

it is for artifacts. Objects, including parts of things, have functions by virtue of being objects of a certain kind. Specifically one where the kind in question includes a criterion for evaluating objects of that kind as functioning well, poorly, or not at all as part of what it is to be something of that kind.

As with the other sorts of accounts discussed, this account largely explains proper functions by picking some particular distinctive feature of proper function talk and building the account off of that point. In this case what I take to be the essential feature of proper function talk is that it is kind based: that proper functions are not things that *objects* do, but rather things that they ought to do, as a result of what *kind of thing* they are. And as a result of that feature I have suggested that what is important when talking about functions is that they are the result of taking objects to instantiate a functional kind which is part of some way of dividing up the world, which is justified by reference to one or more of our practical goals. But it is still worth looking the features that the other accounts I have discussed take as their starting points. The previous accounts all took some feature to be central and tried, and I have argued failed, to account for the sort of evaluations entailed by functions by appealing to those features. The central intuitions behind the systems and effect accounts are easy to account for. If the sort of evaluations in question are related to what an object does, brings about, or is used for this adequately captures both that functions are things that objects do, characteristically do, or bring about, and also that the sort of objects that have functions exist within the sorts of contexts that make it possible to speak of objects doing something, bringing something about, or being used in certain characteristic ways.

Accounting for the explanatory power of functions is more complicated. The explanatory power of function attributions in the case of artifacts is a result of the fact that we create artifacts to serve a purpose. So knowing that something is a hammer tells us something about why it exists, since that general kind of thing was invented with a specific purpose in mind. It is the kind of thing that is relevant for the explanation, since any individual hammer may have been created for a different reason. The function of the hammer does not explain the existence of any specific hammer, just hammers generally. This is technically true when it comes to the parts of organisms as well: we intentionally divide them up into different kinds of parts in ways intended to serve certain purposes, or some useful way of talking about them. But this does not entail the sort of explanations that we want from attributing functions to parts of animals in the way it does with artifacts. When it comes to parts of animals the object itself is what is explained, even if the explanation must be of the object under that description. It is not the presence of that description that we are interested in. While the way of describing parts of animals may depend on our intentions and goals that does not mean that the object under that description does. The explanations that function attributions provide when it comes to the biological functions that etiological accounts specifically address are not in terms of our intentions.

Even if all function attributions entail explanations, though, they do not all entail the same sort of explanations. I argued earlier that there are some functional kinds that do not have the right sort of evolutionary history to entail the sort of historical explanations suggested by etiological accounts. And this is even more true when it



comes to functional kinds that are not intended for use in biology. Not all ends towards which parts of human beings are directed are linked to or explained by natural selection. The parts of human beings which function to lead us to knowledge, or to the best actions are not: natural selection may care about their ability to keep us alive, but not in as much as they are directed at *knowledge*. The same general areas of the brain, or the psychological parts relating to how we reason might be taken to have a different function, which would entail an explanation relating to natural selection. But this would simply be a different way of dividing things up: a different set of parts with a different set of functions. Since the kinds that we appeal to are based in the uses we intend to put them to, and since we have a wide variety of different uses, we should expect to see any number of different and overlapping ways of picking out parts of human beings, each with their own distinct functions. In the case of the sort of biological functions that etiological accounts focus on that use relates to explaining the presence in living things of that part or feature, by reference to its evolutionary history. It is because of that intended use that the kinds in question are ones which entail those sorts of explanations and not the reverse. The reason that function attributions in biology do generate explanations, in roughly the way etiological accounts suggest, is because they are the right sorts of functions; namely the sorts of functions that belong to the way of dividing up organisms that is useful for explaining them in this way.

### **3. Virtue**

I have argued that appealing to the notion of a good life for human beings in order to explain kind relative evaluations of them is circular. Accounts that rely on both a notion of human nature and of a good life for human beings to pick out and explain the moral virtues should instead rely only on an account of human nature. An account of the good life for human beings can be taken to follow from an account of human nature, but not to be what explains why something is a part of it. Explaining the virtues of human beings in this way requires a more substantive account of functions than would otherwise be necessary, since it would be the foundation of the account. I have suggested a way to understand attributing functions of parts of human beings that does not require appealing to some final end for human beings. The criteria for taking things to be members of certain kinds, I have argued, include that things of that kind have a certain function. As a result taking something to be of one of these kinds involves taking it to be subject to evaluations regarding how good it is, as a thing of that kind.

But what I have said about functions does not claim that any particular way of dividing a person up into different parts has any priority over any other way. As it stands it is only an account of what makes functional evaluations true or false. It would include evaluations in terms of health, or strength, or perhaps even good looks, and treat them equally. This may be worrying, since we tend to think that moral evaluations of people have a special quality that evaluations of, for example, health lack. We treat them differently than other evaluations. They are important to us. We take them to be about their character, and not simply some physical problem or excellence. Moral excellence is praiseworthy in a way that differs from good health. Most distinctively moral evaluations

seem to have a unique relationship to voluntary actions, and reasons. Their relationship to voluntary actions and reasons are not as direct as moral evaluations of actions are taken to be, but how we go about evaluating actions, institutions, or other things is not the question here. The question here is solely about evaluating persons. Because of how we treat them it is tempting to think that moral evaluations are of a different kind than other evaluations, or that moral evaluations of persons relate to them in a way that evaluations of health do not.

I will argue that it is not necessary to take moral evaluations to be of a different kind of evaluation than other evaluations of human beings. Making sense of the distinctive ways in which moral evaluations are treated can be captured without taking them to be different in kind from other ways of evaluating people - as being made true or false in the same way as other ones. What is needed is something to make sense of why some evaluations of people are moral and why others are not, why we treat the ones we think of as moral differently from other ones, and why it is reasonable for us to do that. The lines we draw between moral and non-moral evaluations are not always clear. There are ambiguous evaluations and revisions are common, which is to be expected when talking about something we are doing. Since the major purpose here is to save the phenomena, the solution cannot be in terms of strict criteria, since any strict criteria would misrepresent precisely what is to be preserved: it would be a false clarity. What is needed is a way to make sense of what is characteristic of moral evaluations and sets them apart from other evaluations in a way that does not require that they be of an entirely different kind or work in a different way than other evaluations of human beings.

The goal, in effect, is to pick out what about them is characteristic, or characteristically moral without doing so in a way that makes them different in kind from, for example, evaluations of physical health.

In Section 1 I will go back to the account of moral virtue in Foot's account. Foot proposes that what is characteristic of moral evaluations of people is that they are evaluations of some specific part of human beings, namely the rational will. I will argue that it is too difficult to pick out the rational will independently for this to solve the problem, and that distinguishing between moral and other evaluations of human beings requires a different approach. In Section 2 I will argue that the best way to distinguish moral evaluations from other ones is by appealing to the characteristic use to which we put those evaluations, and in Section 3 I will argue that this way of picking them out makes sense of their distinctive, characteristic features. In Section 4 I will address the worry that other evaluations, which we do not typically take to be moral, might count as moral evaluations on my view. I will argue that my view is capable of distinguishing between moral evaluations and, for example, evaluations of health or aesthetic merit in most cases. I will also argue that my view leaves certain cases ambiguous, and that this is not a problem for my view since it accurately reflects a genuine ambiguity in those cases. And I will conclude in Section 5 by talking more generally about how moral evaluations would, on my account, bear a distinctive relationship to reasons for action despite being made true or false in the same way as other evaluations functional evaluations of human beings.

**1.**

Probably the most tempting way to approach this puzzle is to pick out one part of human beings, and argue that evaluations of that part are moral evaluations. This is the path that Foot takes when she suggests that moral evaluations are evaluations of the rational will: “to speak of a person is to speak of an individual not in respect of his body or faculties such as sight and memory, but as concerns his rational will.”<sup>53</sup> The range of different moral virtues are either evaluations of different parts of the rational will, or different ways in which the rational will can be excellent (the difference here is mostly terminological). In this sense moral evaluations of people are like evaluations of livers. While not being evaluations of some physically identifiable part they are nevertheless evaluations of something that is included in the makeup of human beings. The way in which we divide up psychological or moral features of human beings need not be in purely anatomical terms. But it is necessary to give some account of what the rational will actually is, and what distinguishes it from other things, and one that is largely independent of our standing intuitions about what is and is not a moral evaluation.

The rational will, according to Foot, is whatever is involved in acting, voluntarily, for reasons.

“A special connection with the voluntary is, then, the first of the conceptual marks of the special evaluations, picked out from others (such as speech defects) which have to do with goodnesses and defects *in* human beings but are not of the kind that I have indicated by saying that they are about goodness and defect *of the rational will*.”<sup>54</sup>

---

<sup>53</sup> Foot, 66.

<sup>54</sup> Foot, 72.

The way in which we identify the rational will is by means of what it does rather than, in the case of livers, where it is. This is not an important difference, though. The distinguishing characteristic of the rational will is that it is whatever in humans is responsible for voluntary actions, and moral evaluations are evaluations of it. Speech defects would not count as moral defects because while they affect the way in which our actions are performed they are not part of the process that leads to voluntary actions. They are not part of the process that distinguishes voluntary actions from other behaviors such as sleep walking or reflexes, even though they would affect actions. Moral evaluations then are evaluations of whatever in us is part of the process leading to voluntary actions (qua voluntary actions), or leads to some movement being a voluntary action as opposed to something else.

This approach to the problem is appealing since it appears to explain many of the distinctive features of the ways in which we morally evaluate people without taking them to be metaphysically distinct from other ways in which we evaluate people. Evaluations of the part of us that is responsible for doing things are evaluations of a very important part of us, since we do a lot of things. Evaluations of our rational will are best understood given the current limits of our understanding not as evaluations of specific bodily organs but of general traits about us, or in effect our character. This may change at some point as our understanding progresses, but we are not close enough to change our practices in anticipation of it. The way in which moral evaluations are important, and the way in which our idea of the rational will is picked out from other parts in our understanding of human nature, are different from the way in which evaluations in terms

of health are important and the way in which our organs and other bodily functions are picked out. The extent to which there is a unique relationship between moral evaluations of persons and voluntary actions and reasons is straightforward, since what is being evaluated is what brings them about.

Despite appearances, however, this is not sufficient to pick out moral evaluations or, as a result, to explain their distinctive features. Certainly not all evaluations related to voluntary actions are moral evaluations, and not all evaluations of the rational will either. Foot claims that there is a particular kind of evaluations of the rational will that count as moral, unlike other evaluations that are not linked to voluntary action in the right way. But cashing out what the right way is without appealing to an understanding of what evaluations are and are not moral ones is difficult at best. We can evaluate our capacity for voluntary action in ways that we do not typically consider moral. As a result while we can identify some part which is being evaluated, the possible evaluations of it would include moral and non-moral evaluations. Our rational will is affected strongly by moderately high fevers. We may have trouble picking out what we have most reason to do, and could deliberately act badly as a result of that confusion. But it would be odd to say that acting poorly or unreasonably while feverish reflects a defect in someone's character. We would certainly say that someone is reasoning poorly, or that there is a defect in their decision making process. But this does not seem like a *moral* defect as we normally use the word. What we want to say here is that a fever is a sort of affliction: it is an external influence on the rational will, and not a defect of the will itself. This distinction is harder to make than it seems, however. Determining what is and is not



external to the rational will could not be done by reference to what we take to be moral – that is what the distinction is supposed to explain.

One potential option is to appeal to the fact that a fever is a transitory state, caused by diseases, as opposed to a tendency to act poorly in certain ways. This is appealing because our moral evaluations are already linked to character and dispositions. But this is still problematic. Someone who becomes very irritable under stressful circumstances, or liable to overeat is also being affected by temporary (stressful) circumstances. Saying that they have a defect of some sort in their rational will as a result of this seems like a moral judgment. We could try to distinguish between these two cases by arguing that what we are evaluating in the case of someone who acts poorly under stress is their persistent tendency to act this way, and not simply their tendency to act this way at this moment or when under this particular stressful episode. But of course most of us also have a tendency to become confused or unreasonable when feverish as well, and not simply with regards to one particular fever.

There are also cases that do not depend on something like a fever. Mental health disorders look very much like the sorts of evaluations of the rational will that Foot takes to be moral. But it seems worrying to have to say that someone suffering from anxiety attacks in certain situations is displaying a type of cowardice, or that depression is a form of laziness and mania a severe intemperance. These are all persistent traits, however. Distinguishing them from moral flaws in someone's rational will or character would require some further criterion. It is hard to say what that might be, given their effects on decision making. This is problematic since the evaluations we now take to be ones

relating to mental health have been treated as moral evaluations in the past. It could be that we have simply gotten better at recognizing which of these evaluations are and are not ones involving the rational will. But where to draw these lines is not obvious right now, and if we were improving the way to pick apart moral evaluations from ones regarding mental health we should be able to point to the principles we are now appealing to rather than the ones we did before. These do not seem to be readily available, or at least we do not typically point to them as justifying those changes. Instead these changes have mostly followed our confidence in our ability to more effectively treat these disorders. It is strange to suggest that something could fail to be part of the rational will if it can be treated with a drug, since any number of moral failings can be affected by drugs. A person suffering from post traumatic stress disorder can treat their anxiety with anxiolytics or beta blockers, but so can someone who is simply excessively scared of public speaking. And the same is true if we are talking about evaluations of character, which is one of the advantages that appealing to the rational will supposedly had.

As a result the group of evaluations picked out by reference to voluntary action, or evaluations of the rational will contain more than moral evaluations. There are cases in which actions that do not result from any sort of reasoning process or deliberation are subject to moral evaluation, or are taken to express the sorts of character traits that are typically revealed through voluntary action. For example, we might say that someone who runs immediately from danger, before even having time to consider what that danger is or reason about the best action, is still acting cowardly. But in these cases the moral evaluations resulting from those actions are always in terms of features of the rational

will: there are no cases where we morally evaluate instinctive or immediate reactions in terms that we could not apply to voluntary actions. The problem is that there are ways of evaluating persons that are equally evaluations of their rational will, or of whatever results in voluntary actions in them, that we do not take to have the same moral weight.

That moral evaluations have a strong connection to our reasons for action is the best reason to think that they are evaluations of the rational will, even though not all evaluations of the rational will are moral ones. One final possibility to consider is that the distinction between moral and non-moral evaluations of the rational will just is in terms of the connection they have to reasons for acting in one way or another. It could be that moral evaluations of the rational will are ones that are more important, or which give us especially strong reasons for action compared to other ones. Any evaluation of a person could give them a reason to act one way or another, depending on the circumstances. But it is possible that what distinguishes moral from other evaluations of the rational will is that they give us stronger reasons than other evaluations as a result of a more direct connection to our reasons for acting one way or another.

But saying that moral evaluations have a more direct connection does not mean that they inevitably give us overriding reasons, or even that as far as our reasons to act evaluations of our rational will outweigh evaluations of other parts of us. This is only true to the extent that failing to act on our best all things considered reason might itself be a failure of the rational will. Moral evaluations aside from this are not necessarily more important than other sorts, though they do tend to be important, or more important for living a good human life. And there are even cases where someone might have good

reason to choose to be less capable of acting on their best all things considered reason. Someone, for example, who chooses to take a drug that reduces their capacity for self control when it comes to the pleasures of eating in order to improve their liver function would be preferring a non-moral excellence (of the liver) over a moral one (temperance, or even continence generally). If they would otherwise experience significant liver problems this would be the more reasonable choice. So the importance of moral evaluations to us is not a reliable way to distinguish between them and other evaluations, even of the rational will. Moral evaluations may be of significant importance to us, but this is not sufficient to distinguish between them and other evaluations of human beings.

Even compared to other evaluations of the rational will this is not a difference that could be used to distinguish between moral and non-moral evaluations. There is an important difference between moral and other evaluations of the rational will, but it is a difference in how we treat moral evaluations. The extent to which moral excellence seems praiseworthy instead of simply health is one that depends on our taking those evaluations to be moral rather than one of health. Praiseworthiness, when it comes to evaluating persons, simply means that we take the (positive) evaluation to be a moral one, rather than that it has an independent characteristic that can help distinguish moral from other evaluations. To make this distinction there needs to be a further feature that explains it, in order to justify our treating it that way. So this difference cannot be what explains what it is about moral evaluations that gives them this character, though it is a good way of picking out which evaluations we do take to be moral and which ones we do not. If this difference is taken to justify distinguishing some evaluations of the rational

will from others, it would amount to begging the question. Doing this would be taking the fact that we pick out some evaluations from others to justify the claim that the ones that we did pick out have a different character. It might mark an important difference between the two sorts of evaluations, but it does not identify some sub-faculty of the rational will or even some important (independent) distinction between the two groups.

As a result trying to pick out a distinctive faculty or part whose function is that faculty, like the rational will, and saying that moral evaluations are evaluations of that part of us is problematic. Picking out what part of us is involved, even by reference to some functional result, ends up with a collection of different evaluations, some moral and others unlikely to be. While we certainly can talk about a rational will, or a person's character, and even say that moral evaluations are evaluations of it, there are also going to be non-moral evaluations of that same thing. So what it is to be a moral evaluation cannot just be that it is an evaluation of that part of us. Instead we need some other way to pick out moral from other evaluations of persons. One that will both explain their distinctive character and importance to our practical reasoning.

## 2.

Even compared to other evaluations of the rational will this is not a difference that could be used to distinguish between moral and non-moral evaluations. There is an important difference between moral and other evaluations of the rational will, but it is a difference in how we treat moral evaluations. The extent to which moral excellence seems praiseworthy instead of simply health is one that depends on our taking those evaluations to be moral rather than one of health. Praiseworthiness, when it comes to

evaluating persons, simply means that we take the (positive) evaluation to be a moral one, rather than that it has an independent characteristic that can help distinguish moral from other evaluations. To make this distinction there needs to be a further feature that explains it, in order to justify our treating it that way. So this difference cannot be what explains what it is about moral evaluations that gives them this character, though it is a good way of picking out which evaluations we do take to be moral and which ones we do not. If this difference is taken to justify distinguishing some evaluations of the rational will from others, it would amount to begging the question. Doing this would be taking the fact that we pick out some evaluations from others to justify the claim that the ones that we did pick out have a different character. It might mark an important difference between the two sorts of evaluations, but it does not identify some sub-faculty of the rational will or even some important (independent) distinction between the two groups.

Trying to pick out a distinctive faculty or part whose function is that faculty, like the rational will, and saying that moral evaluations are evaluations of that part of us is problematic. Picking out what part of us is involved, even by reference to some functional role, ends up with a collection of different evaluations, some moral and others unlikely to be. While we certainly can talk about a rational will, or a person's character, and even say that moral evaluations are evaluations of it, there are also going to be non-moral evaluations of that same thing. So what it is to be a moral evaluation of a person cannot just be that it is an evaluation of that part of us. Instead we need some other way to pick out moral from other evaluations of persons that will both explain their distinctive character and importance to our practical reasoning. But it will require a different way

of identifying which sorts of evaluations of persons are moral ones. Since I have argued that the functional kinds that support evaluations of human beings are picked out and defined largely for pragmatic reasons, moral evaluations should reflect a way of understanding human beings that is suited to some particular use. Here I will argue that it is possible to understand the distinction between moral and non-moral evaluations of people according to how those evaluations are used. It is the fact that some evaluations of people are put to a specific, moral use that makes them moral evaluations as opposed to something else.

When it comes to thinking about human beings in functional terms there are a number of different general categories of evaluations. The two most immediately obvious are moral evaluations, and medical ones, though there are certainly others. Since the approach to understanding functions that I suggested in the previous chapter appeals to how we would use them to make evaluations, among other things, it makes sense here to start by thinking about what we use moral evaluations for when considering how to distinguish between moral and other evaluations of human beings. The link that moral evaluations of persons have to actions, and our reasons for actions is the most obvious starting place here, since it is what made taking them to be evaluations of the rational will appealing in the first place. Moral evaluations of persons only provide reasons for action to a limited extent: they do not have the force that evaluations of specific actions seem to. But I only intend to discuss evaluations of persons, and not other kinds or their relation to evaluating people. The best moral evaluations of persons can do is to initial, but not all things considered, reasons to do one thing rather than another. They can suggest that an

action reflects some positive or negative aspect of the agent. Knowing that an action reflects dishonesty, or even cruelty is a reason not to perform it, but one that could be outweighed by other more powerful reasons. And evaluating someone in this way does express approval or disapproval, but does not amount to a command. So however the moral evaluations that I am concerned with here are linked to actions it is not in a direct or immediate way but by virtue of expressing approval or disapproval of something about the person in question, by praising them or criticizing them.

Praising or criticizing someone is something that we do for practical purposes, and so when we ask about what we use moral evaluations for it is equivalent to asking about why we praise or criticize someone. And whatever practical goals we have in engaging in this sort of behavior are precisely the sort of things that would justify picking out one way of evaluating human beings as (characteristically) morally evaluating them. The use that we do typically have for moral praise or condemnation, especially public praise, is to give advice or encouragement. We say that someone is good in some respect in order to encourage the sort of behavior in question, to recommend them as a role model to others, or even just to encourage ourselves. We condemn or criticize someone for something that we want to discourage in others, or to prompt them to be better in that respect, or to remind ourselves to watch out for any weaknesses we might have in that respect. Whether the evaluations are accurate or not does not depend on whether it is useful for this purpose or not, anymore than the accuracy of a medical diagnosis depends on the extent to which a particular problem can be treated. Their characteristic use is what justifies picking out certain kinds of evaluations over others only. Entirely accurate



evaluations which are likely to be connected poorly to someone's actions, or otherwise fail as advice or encouragement, are generally taken to be *futile* (and not false), which is very good reason to think that this is their characteristic use.

Moral evaluations of people are not only thought to be futile or useless in situations where they would fail as advice or encouragement, but often we take them to be directly inappropriate. So it is reasonable to take this sort of advice to be the characteristic feature of these sorts of moral claims. It is the practical goals that we have in preferring some ways of dividing up persons into functional parts over others that makes some ways more appropriate than others. And it is also that goal or use that makes sense of distinguishing between different ways of understanding the kinds of parts that make up human beings. Moral evaluations human beings are evaluations that are characteristically useful for giving advice or for encouraging or discouraging certain kinds of actions, both when directed at the person in question or when praising or criticizing someone for the benefit of others. It is being used in this way that gives certain evaluations their distinctively moral character. Almost any advice or encouragement has a moral feel to it, and especially so if the evaluation involved it is commonly or usually used this way. And it is this that makes it important to be able to distinguish moral evaluations from other sorts of evaluations, which are less well fitted to this use.

This is not to say that the moral evaluations of people themselves are made true or false, or even justified or unjustified by their characteristic use. The goal of this chapter is to show that it is possible for them to be true or false in the same way that other

functional evaluations are while still accommodating their special character. Making or expressing entirely accurate evaluations, even the most characteristically moral ones, can be futile in some cases. But it is only their status as moral evaluations that is affected by their characteristic use. We can recognize moral evaluations of people, as opposed to other kinds, in practice by the fact that we use them in a particular way. We use them to encourage people to become better, both by praising the ways in which they are good and in praising others or holding them up as examples. We use them to blame or criticize people to shame them, or to present them as someone not worthy of trust, or to encourage others to avoid imitating them in that respect.

Using evaluations of people in the way that marks them out as moral does not necessarily have to be our direct intent in doing so, as it is when trying to teach someone. When I say to someone that Dick Cheney is a cruel person it is unlikely that they disagree, or are unaware of this, or that they need to be told not to be like Dick Cheney. If that was my intention in saying it it would be a foolish thing to do. If evaluating someone's cruelty is often or characteristically done for those reasons that does not mean that it is always done for those reasons. But the fact that I am not telling someone they should not be like Dick Cheney does not mean that the evaluation of his character is not being used in the same way as when I praise someone for doing a good but difficult thing. The link might simply be less direct. The evaluation could, for example, serve to express and reinforce a shared commitment to try to be one way rather than another, or an understanding of the importance of not being like Dick Cheney in that respect. The sense in which we are using the evaluations to do something can be in terms of how it is

functioning rather than just our direct intention in making the evaluation. This is not a unique feature of moral evaluations, but common when speaking about how things are characteristically used.

Even evaluations that we do not necessarily take to be moral ones can take on a moral character when used as advice or encouragement. For example, when we praise someone for a skill that they possess, like being good at chess, we are doing something that is very similar to praising someone for a more traditional virtue such as kindness or courage. If we are doing so for the direct purpose of encouraging effort from them or another then that evaluation begins to resemble a moral evaluation. But we might be using it in an entirely different way, such as arguing about whether a computer program could beat them in chess or to decline an offer to play chess or simply flattering them. In that, more common, case the evaluations do not feel particularly moral. And the possibility of taking on a moral character vanishes almost entirely when it comes to evaluating people in respects that do not lend themselves easily to this use, such as how effective their kidneys are at filtering their blood.

This way of using evaluations of persons does not exhaust the ways in which we encourage or advise people to act in certain ways or to attempt to become better. The fact that in this particular type of case the activity takes on an especially moral character, or that the evaluations that are being used that way do, does not mean that advising people or encouraging them is intrinsically moral, or that the evaluations inherit their moral status from some distinctive moral property of the activity. Moral evaluations of people are simply ones which are characteristically used, practically, in this way. This can be

recognized by the fact that asserting them, even to ourselves, when they would not be able to function or function well as advice or encouragement is usually futile. And if something seems futile when it cannot achieve a particular end, that is a strong indicator that we are using it to achieve that end, or at least trying to.

### 3.

Picking out the moral evaluations of human beings from other ones by referring to the way we use those evaluations is not useful if this way of distinguishing moral and non-moral evaluations cannot explain the distinctive features that moral evaluations seem to have. Moral evaluations of human beings are significant or important in ways that other evaluations are not. Being morally good is praiseworthy in a way that, for example, health is not, and it is an evaluation of some aspect of their character that is praiseworthy. Finally even though they do not have the direct relationship to voluntary actions, and reasons that moral evaluations of things like actions do they still do have an important relationship to how we have reason to behave.

The way of picking out moral evaluations that I have suggested can account for these characteristic features of moral evaluations. In many cases the way of accounting for them is the same as it was under Foot's account. Moral evaluations of human beings that we characteristically use for promoting certain courses of action, or encouraging people to do better are substantially related to how we go about doing things. The evaluations may not always give us very strong reasons to do one thing rather than another, but this is true of moral evaluations of people in general. This sort of use is related substantially enough to our actions that the importance of moral evaluations is

easily captured, for the same reason that Foot's account captures this feature. Since the most characteristic way in which we use these evaluations to promote or encourage certain kinds of behaviors, especially when it comes to ones that result in some improvement, take the form of praising someone for being good in some respect or criticizing them for being bad the link to praising and blaming is equally straightforward. To say of some trait that it is praiseworthy can be taken to mean both that having it means that one is good in some respect, and that it is appropriately used in the way that moral evaluations are characteristically used.

Accounting for the extent to which moral evaluations are evaluations of people's character is slightly more complicated. It is a bit empty to say that our character amounts to the things we can change through the behaviors we encourage with moral evaluations, but this is largely true. This would be to say that moral evaluations of human beings are evaluations of their character, because their character is whatever we happen to be morally evaluating. But it is a good way to account for what initially seemed like a characteristic feature of moral evaluations of human beings. Describing character in a way independent of moral evaluations is at best problematic, in the same way that independently picking out the rational will is for Foot. It is better to take the claim that moral evaluations of human beings are evaluations of their character to be equivalent to the claim that medical evaluations of human beings are evaluations of their health. When we think of things as subject to moral evaluations, that is when we take them to be evaluations of character, and not the other way around. When we stop thinking of things as moral evaluations, and start thinking of them as medical evaluations, as sometimes

occurs with mental health disorders, we also tend to stop thinking of them as evaluations of character. Instead we take them to be evaluations of something else, such as brains.

Moral evaluations also give us reasons for action. Moral evaluations of people specifically do not give us reasons in the same way or to the same extent that other sorts of moral evaluations or judgments might.<sup>55</sup> I will be talking only about moral evaluations of persons, which are in practice less direct when it comes to giving us reasons to behave one way or another. Nevertheless they do seem to be important to how we go about doing things in a way that other evaluations of human beings do not, and one which seems characteristic and distinctive. If moral evaluations are typically futile when they are not in a position to affect the actions of ourselves or others, then their link to giving people reasons to behave one way or another is important to them.

If what sets moral evaluations apart from other ones is their distinctive use, however, then what makes them distinctive must also come from that use, and not from some special feature of what we are evaluating. Moral evaluations are ones that are characteristically used to encourage certain ways of acting, or improvement in ways related to them, especially in cases where that improvement involves practice or further actions. Whatever makes them *moral* then, or special in the way that moral evaluations are usually taken to be, must come from that use. This makes the explanation for their special link to reasons for action relatively straightforward. If a moral evaluation of some aspect of a person is correct, which is to say if they are good or bad in that way, and if

---

<sup>55</sup> Or possibly give us reason to believe that there reasons that we had not recognized. The question of moral evaluations generally, or details of the ways in which we have or do not have reasons for actions fall outside the scope of this chapter. I am only considering here moral evaluations of persons specifically, and reasons in the most general sense. The question of how moral evaluations of actions or other moral evaluations give us reasons is separate from, though related to, asking about why moral evaluations of persons would give us reasons to act.

that evaluation is being used appropriately, and not in a case where it would be futile or counterproductive, then directly or indirectly it amounts to good advice. As a result it would give us a substantial, if not automatically overwhelming, reason to act a certain way, because that is what good advice does.

If praising someone, or criticizing them did not give them or someone else a reason to act a certain way, after all, doing so would be futile. And it especially gives them a reason when the subject matter is one of significant importance to us, which is usually true in moral cases. Still, what makes someone better in some respect might not be something they would like to do. They might not even want to be better in that respect because of the work involved in becoming better or because of the consequences of being better for their life. Depending on the situation it might be that being a better human being in one respect might leave them worse off in another, either materially or morally. While moral evaluations and good advice in general may give us reasons to act a certain way this does not necessarily mean that we do not have better reasons to engage in other behaviors entirely, or think that we have them. But that does not mean that moral evaluations pushing us towards being better (or which are, at least, ones generally used for that purpose) are not *prima facie* good advice, or directed at benefiting us in *that* way. There are many ways in which someone can be better off, and surely being better is one of them, even if there are cases where we prefer some other benefit entirely.

#### 4.

Showing how moral evaluations of human beings can have these features without being metaphysically distinct from other evaluations is not enough to show that appealing

to how they are used is enough to capture them. It must also be the case that other evaluations of human beings do not have these features, or do not have them in the way that moral evaluations do. Saying that someone's kidneys are excellent at filtering their blood, or that they are especially tall or bald should not have the same characteristic features that moral evaluations have, since they make poor candidates for moral evaluations. They are more likely to be simpler evaluations: ones related to normality. So it is important to come up with some way of distinguishing between evaluations that have distinctively moral characteristics and ones that do not. However there is unlikely to be a clear line between moral and clearly non-moral evaluations, especially in practice. It is unlikely that evaluations of kidney functions are moral evaluations, as opposed to medical ones, but some evaluations are more ambiguous. For example, when we praise someone for a skill that they possess, like being good at chess, we are doing something that is very similar to praising someone for a more traditional virtue such as kindness or courage. And if we are encouraging them in their chess playing by doing so the effect is markedly different than if we evaluate someone positively in some respect that does not lend itself to this sort of encouragement, such as saying that their kidneys are especially good at filtering their blood. In both cases we are saying that they are good in some respect, but we are only encouraging them to act in some way in one of them.

This difference in character is true even when it comes to evaluations which are not directly functional in nature, though in a milder and more ambiguous way. Depending on the context saying that someone has particularly large or powerful biceps can have or not have a certain moral character to it, even though it is not



characteristically used in the way in which we use moral evaluations. If in the situation the evaluation involved has no link to any action or project, such as when a tailor says this to a client, it comes off as a simple evaluation of someone's general build. But if it does involve encouraging some behavior in the person, or praising them for something directly related to that behavior it might have an additional moral feel to it, such as when one body builder says it to another. An evaluation of someone's biceps is not generally taken to be a moral evaluation, and it is a marginal case. But in the right circumstances it can take on the general characteristics of more standard moral evaluations. It is a sort of praising or encouragement related to their behavior, and one which is related to their reasons for acting in certain ways. What is required is the same connection to things that are under our control, such as lifting weights, as in more standard cases like generosity or courage. Things that do not have this sort of connection, or cannot encourage someone in some sort of action, are different in this respect.

It is hard to imagine a context in which saying that someone was especially tall, or bald would feel like a moral evaluation, even if we would like it to be the case that we were one or the other. It is possible to be abnormal, even to a relatively significant degree, without being bad in that respect. But if these things were believed to be under our control to some extent, by stretching or spending time in the sun, then these evaluations would be more ambiguous. What is important is not whether they are or are not under our control in some unique way that distinguishes them from other evaluations. What is important is that we *use* them as if they were. We may simply be mistaken about our ability to control or affect them (we may be mistaken about our ability to affect

anything we subject to moral evaluations). There are ambiguous or marginal cases, where we sometimes use them as or like moral evaluations and sometimes or even mostly in other ways, such as physical gracefulness, large biceps, or even personality disorders. But these ambiguous cases are also, in practice, ambiguous. So it should not be problematic to have them turn out to be ambiguous here either. And, as happens, if certain things turned out to not be that way at all, then we would no longer take evaluations of it to be moral ones. Instead we might take them to be medical evaluations (if being bad in that respect was treatable), or simply an unfortunate state of affairs.

Aesthetic considerations, such as whether or not a person is good looking, are trickier since some of the factors involved are under our control in a way that our height is not. A large number of things we find appealing looking in people are not under our control, and there is also an overlap with medical evaluations as well. We tend to find the outward symptoms of poor health unappealing. But only awful people take these particular aesthetic failings to be personal failings, or to act like they do. There are things at least moderately under our control that are still subject to aesthetic evaluations, however, and we treat them differently than ones that are not. It is often the case that whether or not it is appropriate for negative aesthetic evaluations of someone to have a moral aspect to them is a matter of the extent to which it is under their control. And when it is not clear whether evaluations that appear to have a moral aspect are appropriate or not the debate is largely in exactly those terms, as in, for example, evaluations of peoples' weight.

This is especially true when it comes to aesthetic evaluations which are uncontroversially the result of our actions, relating to grooming practices or clothing choices, which we do treat as having moral value. This does not mean we treat them as being equivalent to things like truth telling or kindness, but smaller moral considerations are still moral considerations. The value underlying our taking these evaluations to have moral weight is likely to be something like considerateness, respectfulness or interest in presenting a pleasant exterior to others<sup>56</sup>. They are a part of etiquette, and there is good reason to think that as a result it involves genuine moral evaluations of people. These evaluations are less important to us than the most characteristic moral evaluations since, for example, good table manners are less central to our general lives than treating people cruelly or habitually stealing from them. But there is no reason to assume that all the aspects of people that are subject to moral evaluation are equally important or significant. In fact it would be just as strange to say that this was the case with moral evaluations as it would be to say the same about medical evaluations.

The comparison between moral and medical evaluations is a useful but tricky one. There is no simple line to be drawn between moral and medical evaluations. Moral evaluations make most sense in contexts where improvement is related to behavior and practice, whereas medical evaluations show up most distinctively when there is a treatment or cure involved. But there are cases where it is entirely ambiguous whether the evaluation is one, the other, or both. For example, we might suggest that someone,

---

<sup>56</sup> There is a small literature on this question (E.g., Stohr, K., 2006 and 2012, Buss, S., (1999), Calhoun, C., (2000) and Sherman, N., (2005)), but widespread agreement in it that good manners is a virtue, if not the biggest one, and that its goodness is tied to being considerate to and respectful of others. The agreement is most likely due to the fact that this is correct.

reacting to some situation with excess fear, is displaying a lack of courage. And by doing so we are encouraging a different pattern of behavior or state of character: that they act and respond differently to those situations, or are better inclined to react appropriately to them. Or instead we might suggest that they may have an anxiety disorder and recommend an anxiolytic or a medical procedure to help remedy the problem. In both cases, though, we are making evaluations of roughly the same thing. What is different is the context in which they are being made, and especially what we are using the evaluation to recommend.

The overlap is especially important when we consider the genuinely mixed cases: for example, when people take a drug in order to temporarily increase their courage in a scary situation like combat, or public speaking. Or, in the other case, when people learn to react differently to fearful situations by practicing reacting appropriately to them in a therapeutic context. And many ways of evaluating people have shifted from being treated as moral evaluations to medical ones, when we found a way to treat them medically. The transition from a moral matter to a medical one when it comes to what we now call mental illnesses is recent enough that it is an obvious example. Aquinas argued that the vice of sloth, or *acedia*, “is an oppressive sorrow, which, to wit, so weighs upon man's mind, that he wants to do nothing”<sup>57</sup>, which, to modern eyes, resembles something entirely different. Once we can cure some defect that we considered a moral one, or at least treat it relatively effectively there is, at least in practice, a push to treat it instead as a medical issue. This is not to say that there is always a complete transition in these cases, or that there should be, or even that this trend is a good thing on balance. But

---

<sup>57</sup> Aquinas, T. *Summa Theologica* II-II q.35 a.1

it does show how contingent these categories can be for us, and how closely related they are to our practical uses of evaluations based on them. A change in which category some evaluation falls under, however, does not mean there is a change in what makes it true or false, though it may mean abandoning one evaluation for another structurally similar but slightly distinct one. Even if it matches our practices, however, this sort of contingency can raise questions about the distinctively reason giving status of moral evaluations.

If what distinguishes moral evaluations of people from other ones, such as medical evaluations, is that they are characteristically used directly or indirectly as advice then those other evaluations are not characteristically used this way. There are overlapping or ambiguous cases where this may be less true, or messy, but if it is a way of distinguishing moral evaluations of human beings then it must be distinctive of them. This also means that moral evaluations of human beings are linked to reasons for acting in a way that other evaluations are not, since that link is an important feature of good advice and of moral evaluations. But this is tricky since other evaluations, medical ones especially, are important to us and they are usually connected to reasons we have for acting one way or another, and in ways directed at being better in some respect (related to our health). I have suggested that medical evaluations are typically used in ways relating to treatments or cures, but this does not necessarily mean that they are not also related to advice in the way that moral evaluations are. If moral evaluations can be picked out by reference to their characteristic use, then, there must be a way to distinguish between the characteristic uses of evaluations like “your cholesterol is unhealthily high” and “you are insufficiently concerned about your cholesterol levels.”

The difference between these two evaluations can be seen when we consider someone who has a medical problem, such as high cholesterol but decides not to treat it (through lifestyle changes or medication), for poor reasons. It would be unusual to say that a poor choice to accept a medical problem rather than treat it is itself a medical problem except perhaps in the sort of ambiguous cases mentioned above, such as an anxiety disorder relating to medical treatments. Instead this failure to properly weigh the good of health against other goods would reflect a *moral* failing on their part. And the normal response would be to advise the person to act differently, not to attempt to treat their refusal to accept treatment. This marks an important difference between medical and moral evaluations. A failure to act appropriately in response to a medical evaluation is a moral failing, not a medical one. But a failure to act appropriately in response to a moral evaluation is a moral failing. So there is an important difference between evaluations like “your cholesterol is unhealthily high” and “you are insufficiently concerned about your cholesterol levels.” The first is being used as an evaluation of something related to health, and the second is being used to advise a certain attitude towards that evaluation, but is not itself a medical evaluation even though physicians will usually provide both evaluations. As a result we can say that the medical evaluations themselves do not count as the advice, but the moral evaluations related to them, when used in their characteristic way, do. And their connection to reasons to act in some way or other is different, though both have one.

**5.**

One of the reasons that we ought to take medical evaluations to be important to take into account when acting, and that doing so in the wrong way indicates that we are bad in some respect, is that health is good for us, and desirable. And this is eventually grounded in a moral evaluation. But moral evaluations do not have the same relationship to some other evaluation: being a morally better person is something desirable because it is morally better to be a better person, not some further reason. But if this is true than one contingency has simply been traded for another one. Moral evaluations are most characteristically used to advise or encourage certain kinds of behaviors, specifically ones directed at being a better person. This means that the sense of good or benefit involved is fundamentally related to what it is to be a human being. But what we actually take being a human being to involve, or our specific understanding of how human beings *work* is only constrained to some extent by the world we live in, not determined by it. If we took cruelty to be something that made people better human beings, evaluations of them with regards to kindness and cruelty could serve exactly the same purpose as if we took kindness to be the virtue and cruelty to be the vice. Both of these are under our control in the same way, since we can make ourselves worse at least as easily as making ourselves better. Only in one case are we giving *good* advice, but how would we go about justifying claiming that kindness makes us better human beings and cruelty worse, and how can we go about figuring out precisely which things make us better human beings or worse ones? It is not enough to assert that our basic understanding of what it is to be human means that kindness is a virtue and cruelty a vice. There must be at least some reason to think that we do in fact understand human nature to be this way: to justify the

claim that kindness is an essential part of what it is to be a good human being rather than a marginal one, or that reacting kindly to situations is more often appropriate than not.

The easiest way to support the claim that kindness, rather than cruelty, makes people better is to appeal to the place that it has in human lives. As human beings we live together with other people, need close and caring relationships, and, as often as not, need help from others. Kindness makes us better suited to live a life like this than cruelty does. As a result we have good reason to think that it is kindness that makes us better at being human beings than cruelty. This sort of reasoning sounds dangerously close to suggesting that the *explanation* for why something makes someone a better human being is *because* that thing is needed to live a good human life, and that would be a problem. But the fact that the idea of a good human life is useful in justifying claims about whether or not something makes someone a better human being does not mean that that thing makes people better because they are needed for that life. Explaining something is not equivalent to explaining why we ought to believe something, or how we can come to know it. Appealing to the extent that kindness suits us to live well to justify the claim that it makes us better human beings is analogous to determining the existence of a cause by appealing to its effects. It is analogous to following a set of tracks through the snow in order to discover that they end at the local grocery store, and concluding that the person who made the tracks is inside the store. The tracks justify the belief that they have gone to the store, but they are not what explains why the person is in the store in the first place, or what justifies their being there. The idea of a living well as a human being, and what is required for it, and what would make us more or less well suited to live that life is



a good guide to what makes us better or worse as human beings, not because it determines what makes us better or worse but because it is determined by those things. As a result, we can evaluate whether advice is good or bad partly by considering whether following it is likely to make our lives better, as human lives, or worse, but what makes it good or bad is if it is directed at living up to our humanity or not.

As a result, appealing to the idea of living well is a useful way to get at an understanding of our general picture of what it is to be a human being. It helps us see how we add up to be the sorts of things that we take ourselves to be. This does not, however, and could not explain *why* some particular functional aspect is part of our idea of humanity. It only clarifies our already present, if vaguer, picture of what it is to be a human being by reference to more practical situations, ones where our intuitions are clearer or more focused. Specific questions about what traits benefit us as human beings, or which ones leave us worse off, are often easier to answer than general questions about what we take human beings to be on the whole. We could still imagine a different picture entirely, taking ourselves not to be human beings but some other sort of thing that worked, within our general biological constraints, entirely differently. We could easily point out a fundamentally decent person and say, accurately, that they are a bad Klingon. Or say that we should understand ourselves to be the sort of thing made better by cruelty rather than kindness, and make entirely different but equally accurate evaluations of people according to that set of standards. So there must be some reason for us to prefer to take being a human being to be a particular or specific sort of thing, namely the one that tends to show up in our moral practices.

Our reason for preferring our way of thinking about human nature to another cannot simply be that our current understanding is the one which leads us to the most enjoyable lives, or something along these lines, since there is no obvious guarantee in life that being a better human being has that result. Even if it did it would be possible to prefer some other kind of life, perhaps irrationally, and we would not say that a person who did was not a human being but some other kind of thing entirely. Appealing to the idea not of the most enjoyable or pleasant life but the best life for human beings would be an empty claim at best: it would be true of any possible way of understanding the kinds of things that we are. There is some reason to prefer some pictures of what it is to be human over others if they are less likely to entail that being a good human being results in dying very young, or being in great pain at all times (especially with regards to what it is to be a healthy human being). But this is a weak constraint given that we do tend to think of human beings as the sorts of things that are often in pain, and that do tend to die earlier than they would prefer. A deeply unpleasant picture of what it is to be human can be subsumed under the more general constraint provided by the physical sorts of beings that we are, in the same category of including an ovipositor in the list of human organs.

What we need from general picture of what it is to be a human being is an understanding of how people are made up, and how they function as a largely unified whole. Specifically we need one that lets us talk sensibly about human beings and how they work (or do not work) in functional terms. And, as a result, we need one that mostly hangs together, especially within the constraints provided by our physical makeup. And for the most part the stranger possibilities, such as taking cruelty to be a virtue and

kindness a vice, fail on these grounds since a great deal of other things in human lives relates to how kindness and cruelty work for us. It would be harder to give a significantly consistent picture of human nature that was even remotely similar to what we have now but took cruelty to make us better as human beings. This does not mean that what we have managed to cobble together into a general idea of what it is to be a human being, and what makes us better or worse, is the best we could have. It almost certainly is not. Revisions to our understanding are possible, probably a good idea, and certainly something that happens regularly and results in a shift in how we think good people are and behave. But we do have reason to think that what we currently have is moderately serviceable. It is better than taking cruelty or irresponsibility to make us better, or dropping the picture entirely in favor of an understanding of ourselves as Klingons, because we do have a picture of the good life resulting from it that more or less hangs together within the basic physical constraints involved when dealing with human beings. That is the reason *why* it is so useful in asking about what individual things make us better, and how.

What we do have is, while imperfect, a picture of human beings as complex creatures made up of many different functional kinds, which support evaluations. This results in an understanding which allows individual human beings to be evaluated in various ways, or with regards to different aspects of our makeup. It allows us to talk about ways in which human beings can be said to be better or worse off, or good or bad in some respect. And we need it to be as responsive to the basic physical constraints provided by the way we generally are, as internally consistent, and as comprehensive as

we can manage. An understanding of what human beings are that largely ignored the physical features of most human beings, such as one that included an ovipositor in its list of internal organs, would be frivolous: there is no value to a more complicated view without any additional advantage. An understanding of human nature which is internally inconsistent would be difficult to use in most practical situations. And an understanding of human nature is better if it is more detailed or of a relatively fine grain since that increases the amount and variation of what we can say about human beings – this is why something like dividing the human body up into the liver and the not-the-liver would not be a good idea.

The uses to which we put this understanding, and the ways in which we can evaluate people vary, and the easiest way to distinguish between kinds of evaluation, such as medical or moral, is by the goals we have in using them. Medical evaluations are ones that we use primarily or characteristically in cases involving treatments or aspects of us which are typically improved by medical procedures (when necessary). Moral evaluations have a similar character but are used characteristically in cases where we are advising someone towards a certain action or behavior, especially by means of praising excellence of some sort, or criticizing a flaw or defect. As a result, they characteristically involve aspects of human beings that improve with attention, behavior, and practice, and by virtue of their relationship to our behavior are also characteristically linked to medical and other evaluations. Their connection to advice gives them an especially normative character and as a result they also tend to stand above other evaluations, pushing us towards being good in those respects as well as in the case of health. And as a result they

take on a particularly important character when compared to other evaluations, despite resulting from the same general picture.

Connecting moral evaluations both to the way in which we use them, and taking them to be the result of an understanding of human nature that is itself evaluated according to criteria of usefulness and subject to change, gives moral evaluations a somewhat contingent character. This does not mean they are not true or false: that is a separate question. But it does mean that the understanding of human nature that allows us to make them in the first place is subject to change, as is whether they are or are not moral evaluations specifically (independent of their truth or falsity). But our understanding of human nature as a functional whole, and our taking certain evaluations to be or not be moral does change over time, in response not just to increases in knowledge but also changes in our circumstances, as does the understanding of what a good human life that follows from it does. As long as our general picture of what it is to be a human being is flexible enough to accommodate the way we use it, and comprehensive enough that it entails a reasonable picture of what it is to live a good human life, its contingency should not be concerning. If that is true there is no reason to worry about whether any substantial revisions to our conventional picture of the moral virtues would be necessary. I have argued that the account of virtue presented by Philippa Foot in Natural Goodness contains a dangerous circularity: it entails that the natural end for human beings, or the good human life, both determines and is determined by an account of what makes human beings good or bad. In order to avoid this circularity the idea of the good life for human beings can only be a consequence of human goodness,

and suggested a way in which to understand how we could evaluate human beings in functional terms without reference to some final natural end. I have also argued that the distinction between moral and other evaluations is largely a practical one. There are indefinitely many ways of dividing human beings up into various parts, each with its own function. As a result it was important for me to give an account of how to distinguish the specifically moral evaluations from other sorts of functional evaluations. I argued that there is no substantive distinction between different sorts of evaluations, and that what remains to distinguish the evaluations we take to be distinctively moral from other evaluations are pragmatic concerns. Instead of a substantive category, different in kind from other evaluations, moral evaluations are a practical category, identified by their characteristic uses. And, I argued, the typical aims involved in moral evaluations have to do with giving advice, including to ourselves, and also praising and blaming behaviors or characteristics, often in a way intended to recommend them to others. This use is what marks out moral evaluations as different from other sorts, such as medical evaluations, and also allows us to distinguish between more or less moral sorts of evaluations based generally on their usefulness.



#### **4. Defect**



In this chapter I will argue that there is an additional advantage to what I have suggested, namely that an account based on a notion of the good life for human beings is vulnerable to the possibility that there is no such life. I will argue that there could be no such thing as an ideally good life for human beings, which amounts to claim that human beings are defective by our nature: not simply in practice but in principle. In section 1 I will argue that it is possible to imagine something that is by its nature defective in the first place. In section 2 I will present one morally important way in which human beings are incapable of real excellence, as a result of a characteristic cognitive bias related to our self knowledge. In section 3 I will present a second morally important way in which human beings are incapable of real excellence, as a result of a characteristic psychological tendency usually referred to as the belief in a just world. And in section 4 I will present a third morally important way in which human beings are incapable of real excellence, due to the relationship between in-group and out-group biases. Finally in section 5 I will argue that the fact that our current understanding of how to morally evaluate human beings entails that we are naturally defective does not justify revising that understanding in favor of one that does not have this result.

## **1.**

Up till this point most of the arguments that I have made had to do with the theoretical structure standing behind the account of virtue. And I have spent some time arguing that the results of the changes I have argued for would not diverge too significantly from the sort of Aristotelian account that I have discussed. Many of the features of those accounts are also features of what I have suggested, some directly and in

other cases through a slightly different path. But the extent to which this is true should be a little worrying. If there is no interesting difference at the end, but only a minor reshuffling of the technical terms, why bother at all? If there's anything useful to this it has to have some additional advantage to it. And I haven't shown yet what actual changes result from focusing on functions of parts rather than a general end for human beings, when providing an account of the virtues. I want to argue here first that the change really does make an important difference, and second that it is one that gives the account I have suggested a significant advantage.

There are two different reasons for taking the account I have suggested to have a significant advantage over accounts in which an ideal end for human beings plays an important role. The first reason is that those accounts depend substantially on an empirical (and contingent) claim, namely that it is possible for someone to live an ideally good human life, or be fully virtuous. This should be worrying: if there were no such thing as an ideally good human life or there were one but it was not possible to live that life the accounts would effectively deny the possibility of moral evaluation. And whether or not it is true, it is certainly possible that there could be no such life available to human beings. So for accounts in which a notion of an ideal life for human beings is a substantially determining factor in what counts as a virtue at all, our ability to make moral judgments is dependent on an entirely contingent fact about the world, namely that our nature is coherent enough that an ideally perfect human being is a logical possibility. If human nature does not *add up* in the right way, or if we lack the right sort of teleological coherence, the result would be a sort of nihilism. This is an awkward

situation, especially given that human nature is largely shaped by the process of natural selection, which is almost certainly capable of producing creatures that are incapable of living well in any sense.<sup>58</sup>

The account I have suggested, however, is entirely free of this worry. It would certainly be better for us if it turned out that our nature was coherent enough to allow for the possibility of a fully virtuous person or an ideally good life. Even though both of these are vanishingly unlikely, and almost certain never to have occurred, their possibility would certainly be reassuring. But given that there is at least a possibility that this might not be the case it is an advantage of the account I have suggested that it does not require this. The final end for human beings, according to what I have suggested, can simply be a result of the account and not required for it. If human nature is not coherent enough to add up to an entirely consistent final end that would not make moral evaluations impossible, which is a significant advantage since that is at least a possibility.

The second and more significant reason to take the account I have suggested to have an advantage over accounts in which an ideal end for human beings plays a significant role is that human beings are the sorts of things whose nature does not add up

---

<sup>58</sup> The most important sources of this worry for accounts of virtue like the one Foot gives can be found in Bernard Williams (1995), in the form of a commentary, and in Williams (1983). In the latter he suggests that our current scientific understanding of human origins casts significant doubt on the possibility of a fully good human life, since:

“human beings are to some degree a mess, and that the rapid and immense development of symbolic and cultural capacities has left humans as beings for whom no form of life is likely to prove entirely satisfactory, either individually or socially. Many of course have come to that conclusion before, and those who have tried to reach a naturalistic morality which transcends it have had to read the historical record, or read beyond the historical record, in ways that seek to reveal a partly hidden human nature which is waiting to be realized or perfected. The evolutionary story, to the extent that it can now be understood (and to the much more modest extent to which I understand it myself) seems to me to give some support to the view that in this respect the historical story means much what it looks as though it means.” (109)

in a coherent way. We are *naturally defective*. As a result an account which can handle this fact about us has a decided advantage.

There is something odd about suggesting that some thing might be defective by its very nature, or that there could even be kinds of things which were defective simply by virtue of what they are. It is tempting to assume that if there could be no excellent examples of some kind of thing then it must be the case that there could be no defective examples either. If not our evaluations would seem *unbalanced*, and maybe even poorly defined. But I think this temptation is misleading. Defect and excellence are not mirror images of each other: there are many more ways in which something might be defective than excellent. And for some kinds of things there may simply be no way at all to be fully excellent, or more specifically no way to consistently satisfy the evaluative criteria proper to things of that kind. It would be a kind of thing with some significant evaluative conflict built into its nature.

The most intuitive way to imagine this is to imagine a kind of thing which cannot consistently be excellent in every respect at any given time. There is a basic trade-off for it when it comes to the functions of its various parts, to the extent that excellence in one part would entail a defect in another. But this is not the only way to understand how something might be naturally defective. The explanation might not be a conflict between different ways in which something could be excellent or between two different parts which could not both be excellent. Instead it could be that the conflict is internal to some more general kind of thing whose evaluative criteria are unfulfillable. In most cases this would amount to the same thing as the previous possibility. If something cannot be good

in one respect without being worse off in a different one, which is common, we could view this as either a conflict internal to that particular thing or view it as two separate parts which could not both be excellent at the same time. The difference has largely to do with how specific we take the function of the part to be: if we identify it in a broader sense the conflict is likely to be internal to how it operates. If we divide up the mechanism involved and treat it as two separate parts then the conflict internal to the more general part is a conflict between the two separate, more specific parts. As a result of either conflict, however, there would be no way in which the more general kind of thing could be excellent. For the same reason it could not be excellent, it would always be defective. And this is possible despite the lack of any specific way in which it could be excellent. I will be focusing on conflicts internal to specific parts for the remainder of this chapter, rather than conflicts between two separate parts, because it is easier to see the necessary defect when presented that way.

It is important to give an example of how this sort of natural defect tends to work. The claim that the human back, and especially the lower back, is not very well fitted to its task is not, I think, an especially controversial one. Lower back pain, and impaired function is a common experience in later life, usually stemming either from strain on the muscles of the lower back or the pressure exerted on the (gradually deteriorating) components of the spine itself. The reason that this is an especially common feature of aging is largely that the human lower back is simply not well constructed to support the upper body through its characteristic range of motion.

“When humans stood upright, they took a spine that had evolved to be stiff for climbing and moving in trees and rotated it 90 degrees, so it was

vertical—a task Latimer compared to stacking 26 cups and saucers on top of each other (vertebrae and discs) and then, balancing a head on top. But so as not to obstruct the birth canal and to get the torso balanced above our feet, the spine has to curve inwards (lordosis), creating the hollow of our backs. That's why our spines are shaped like an "S." All that curving, with the weight of the head and stuff we carry stacked on top, creates pressure that causes back problems—especially if you play football, do gymnastics, or swim the butterfly stroke. In the United States alone, 700,000 people suffer vertebral fractures per year and back problems are the sixth leading human malady in the world. "If you take care of it, your spine will get you through to about 40 or 50," said Latimer. "After that, you're on your own."<sup>59</sup>

I do not take this to be a particularly controversial claim to make about the lower back. What I do think we see here is a direct conflict between two very important features of how we go about evaluating lower backs in the first place. It is not enough when evaluating something like the lower back to only say that its function is to support the upper body through a certain range of motion. What is also important is that it do that in a *specific, characteristic way*. There is more to a function than simply a result, after all. How specifically we characterize what that way is depends substantially on what sort of thing it is, how general or particular the kind of thing in question is, and so on. But there still must be some characterization of how exactly that kind of thing *works*. We might say that a wrench is functioning *as* a hammer if we use it to drive a nail into a board by striking it sharply from above. But it would not be right to say that if we use it drive a nail into a board by gripping the nail and slowly applying a steady downward

---

<sup>59</sup> Human Evolution: Gain Came With Pain by Ann Gibbons on 16 February 2013. <http://news.sciencemag.org/sciencenow/2013/02/human-evolution-gain-came-with-p.html> (retrieved July 10 2013)

pressure to it. Only one of these ways of driving a nail into a board is *hammering*. The problem with lower backs, however, is that the characteristic way in which they go about supporting the upper body is not, in the long run, very good.

This situation is fairly common when it comes to the sorts of things we find around us. It is unusual to find a truly perfect kind of thing, or even one capable of being that way at all.<sup>60</sup> So it is probably unreasonable to expect either the way in which kinds of things achieve the result at which they are directed to be, when the thing is working ideally well, the optimal one. There are bound to be any number of things which, while not optimally designed, are still good enough that they do not count as defective, or at least not to anything but the most useless standards. But this should not stop us from recognizing that many things are especially poorly suited for their functions, and sufficiently so that they are what we could quite reasonably call naturally defective.<sup>61</sup> As with most things there is unlikely to be a sharp line that can be drawn here, and there are almost certainly cases where it is difficult to see how to classify them. Lower backs, however, do not seem ambiguous enough for that, given the frequency and results of their failings.

There is a natural objection to bring up at this point. Isn't the more appropriate response to this sort of situation to simply lower the evaluative criteria involved? After

---

<sup>60</sup> Bicycles maybe.

<sup>61</sup> Perpetual motion machines, as a general category, seem to fit this description nicely. We can recognize that kind of thing, with its specific function, while also being entirely aware that every one of them will turn out to fail miserably. The way in which they achieve the end towards which they are set will not achieve that end, since nothing will.

It is worth noting, however, that it would be necessary to talk more specifically about a sub-type of perpetual motion machines, with particular characteristic mechanism, before talking about any tradeoffs between parts of that mechanism. The value of the example is that we can be certain that there will be some inconsistency between any mechanism and what it serves to do since nothing will serve to do it.

all, we can easily pick out things that make lower backs *better* at doing what they do, such as strength, flexibility, and so on. And as a result we can say that some lower backs are better lower backs than others. After all, nothing about lower backs prevents us from imagining one that was as good as a (human) lower back could be, even though it would not be as good as other sorts of lower backs would be. Why should we take the range of lower backs to be between very bad lower backs and lower backs that are not as bad as most of the others? We could just as easily say that a lower back that was better than any other human lower backs *just is* what it is to be an ideally good lower back. We could simply take 'good' to be the same thing as 'best' and be done with it.

There is certainly a range of lower backs, with some being quite a bit better than others. But there is a difference between a *better* lower back and a *good* lower back, even if there is a back better than all the others.<sup>62</sup> In the second case, which I am arguing for, the appropriate standard for evaluating lower backs is the one set by the kind of thing that lower backs are, or their function. But in the first case the standard for evaluation includes both the standard set by the kind of thing that lower backs are, and also the other existing lower backs. It is still necessary to include the basic standards for evaluation since there must be some way to determine which lower backs are better, and which are worse. Simply comparing backs to each other cannot do this on its own since there would be no way to pick out which features to compare or how to tell which are better and which are worse. So taking 'good' and 'best' to amount to the same thing could only

---

<sup>62</sup> I will return to this general question further on when it comes to moral questions, because it is important in that context as well.



be a sort of rhetorical shift, or a way to signal a different practical attitude towards our evaluations of things.

If we did try to adopt this sort of attitude, and shift our way of using these terms accordingly, though, there would be some odd results. After all, the actual extent to which things like lower backs fulfill their functions is important to us! There is something strange about the idea of a doctor cheerfully reassuring an older patient with significant back pain that there is nothing particularly wrong with her back, since this is exactly how lower backs, even the best ones, feel once they have started to deteriorate with age. It would be strange to care a great deal about how our lower backs are doing compared to other lower backs, though we might enjoy boasting about ours if it is particularly good. What is more important is how well it is fulfilling its actual function. And since this is a matter of evaluating a back in the way I have suggested, shifting our focus from 'good' to 'better than the rest' would just distract us from what we really should care about. It may be that this is still a good idea, reflecting a more positive and less judgmental attitude towards the world. But we should not take it to involve a difference in the actual evaluations that we are making. (And that attitude could, I think, be adopted without shifting the way we talk in any important way.)

I have been talking only about parts of human beings here, and mostly lower backs. But it is fairly straightforward to extend the point out to whole human beings. If one or more parts of a human being were incapable of being excellent, or if that part was *always* defective, then it would be impossible for human beings themselves to be free from defect. They could never be free from a defective version of that part. So if parts of

human beings are structurally incapable of excellence, then so are human beings in general. In the case of the lower back the best way to understand the sort of evaluations I have been discussing is to take them to be medical evaluations (concerning pain, difficulty moving, and so on). So the argument here would amount to claiming that since there could be no ideal human lower back, it would equally be impossible for there to be an ideally healthy human being.

The claim that it would be impossible to have an ideally healthy human being is probably not a particularly controversial one. The many weaknesses and failings that are common to human bodies do seem to be distinctively characteristic of human beings. And some imperfections seem fundamental to being human at all. But what is less obvious is that this should matter when talking about *moral* virtues. Just because a person who is ideally healthy in all respects is like to be an impossibility does not mean that the same goes for an ideally morally excellent human being. We are nearly as likely to run across one as the other, of course, but that does not mean that they are both equally impossible. Claiming that human beings are naturally *morally* defective is a much stronger claim, even given the available examples of human beings.

The claim that human beings are naturally morally defective is an even stronger claim if, as I have argued, moral evaluations have a link to specific pragmatic goals, or are distinguished from other evaluations by their usefulness to us in certain circumstances. If this is the case there is at least a *prima facie* reason to think that moral evaluations of human beings might be different than other evaluations, such as medical ones. This does not mean that they might be different in kind, or another sort of

evaluation entirely. But there is some reason to suspect that they might be evaluating aspects of human beings which are *not* naturally defective: virtues which were at least theoretically achievable might be more useful, for the simple reason that we could at least aspire to them. And even if it did turn out that what we currently take to be moral virtues (as opposed to some other sort) were fundamentally unreachable, we might simply take that as a reason to abandon those particular virtues and replace them with different ones that we could achieve. It would be as strange if human beings were naturally defective in *every* way as if we were entirely perfectible, though what resulted might look very different from the way things are now.

As a result, arguing that human beings are naturally morally defective will require arguing for two separate points. First I will argue that we are naturally defective with regards to our conventional understanding of moral virtue, or that we are morally imperfectible. I will not argue that this is the case for every morally important aspect of human beings. This may or may not be the case. Instead I will pick out three significant ways in which we are naturally defective, which is sufficient to support the claim that we are fundamentally morally imperfectible. Secondly I will argue that our moral imperfectibility is not a sufficient reason to revise how we go about morally evaluating human beings. In fact, I will argue, we are better off taking ourselves to be naturally morally defective than adopting a different, potentially achievable set of moral virtues.

Demonstrating that we are, as far as our conventional understanding of the virtues goes, fundamentally imperfectible is relatively straightforward. It can be done in roughly the same way as the earlier argument about lower backs. The argument will have to be a

little more complicated: it would be impossible to change the structure of the back but leave everything else around it unchanged. The various muscles, internal organs, and so on all depend on the lower back remaining the way that it is. As a result, having a back that did not share the characteristically defective nature of human lower backs would still result in problems overall, since other bodily parts depend on the structure of the back being what it is. Showing that what may appear to simply be a common challenge faced by human beings is a characteristic defect is trickier since the extent to which morally relevant aspects of our nature depend upon each other is less clear. The flaw in question cannot simply be characteristic of human beings, or it could be supplanted. It must also be necessary for us to function properly overall: it has to be the case that even if we are harmed by our imperfection, which is inevitable, lacking that same imperfection would *also* harm us. Whether the defect or its absence is *more* harmful to us is less important.

## 2.

The first example that I want to consider involves what is generally known as the egocentric bias, which picks out several related ways in which we typically make mistakes in judgment. We have a strong tendency towards “unrealistically positive self-evaluations, exaggerated perceptions of control or mastery, and unrealistic optimism”.<sup>63</sup> In other words, human beings almost universally take themselves to be both better than they are at things and better than they have reason to think they are, and better in comparison to other people in similar situations. For example, a shockingly large percentage of people believe themselves to be better than average drivers.<sup>64</sup> We

---

<sup>63</sup> Taylor, S.E. and Brown, J. (1988) p. 293

<sup>64</sup> Svenson, O. (1981). Among subjects from the United States it was a stunning 93 percent.

commonly take ourselves to have control over events that we cannot actually control, or did not actually control. And in general we expect things to turn out better than is likely given what we know. This is not universally true of people: it does appear that people suffering from some (moderate) forms of depression tend to avoid this specific bias in reasoning.<sup>65</sup> But otherwise it is almost universal, even though the extent to which it affects people's judgments varies.

This flaw in our reasoning is important because the extent of our self knowledge has direct and significant consequences for our behavior. We tend, quite naturally, to take credit for the results of what is often blind luck. And, even worse, this bias also leads us to take credit for the work of others. We incline towards making false predictions about how we will act, since we take ourselves to be better than we are. We act on the assumption that we have more control over future events than we do, and so on. When someone is abnormally lacking in self knowledge the poor consequences are obvious. In more normal cases the effect is much smaller, and often relatively harmless. But something that is often or mostly harmless still leaves one worse off as a person, and is like to be harmful, eventually. And while someone might not display any significant effects for some time this is at best a matter of luck. There is an inevitable cost to this sort of failure, even if it is only taking larger risks than are reasonable or secretly looking down on one's peers or coworkers.

What is genuinely worrying here is that these biases are not just unfortunate but typical failings. Our tendency to reason badly in these ways is not just a matter of being too self interested, or short sighted, or something else for which real virtue might be a

---

<sup>65</sup> Alloy, L. B. and Abramson, L. Y. (1979)

corrective. We cannot simply take it to be the sort of vice that is generally found in people, to varying degrees, and which a truly virtuous agent would lack. Our egocentric bias almost certainly comes with a substantial moral cost, but its absence does as well. These biases are strongly linked with a sense of contentment and happiness, which is not especially surprising. But they are also linked with creativity, productivity, and general success in life.<sup>66</sup> Even leaving aside the only reliable cases in which people appear to lack this sort of cognitive distortion (the moderately depressed), it is within certain limits beneficial to suffer from a stronger, rather than weaker, bias. Not only does this bias enhance our lives, but lacking it can (and does) harm people to some extent, both in terms of how they feel their life is going and also when it comes to other virtues like the sort of productivity that leads to success in life, or creativity. As a result this meets the criteria that I set out earlier: this is a substantial flaw in our ability to accurately judge certain morally important features in the world; it is characteristic of human beings (its absence indicates a defect, even); and, most significantly, important goods *depend* on our being defective in just this way. As far as this particular aspect of our nature goes, we must end up acting badly in order to do well, morally and otherwise.

### 3.

Even so, the harms that tend to result from a normal susceptibility to the egocentric bias are probably somewhat minor. The fact that some of them are morally relevant harms is enough to establish my point, but at this level it seems only a little

---

<sup>66</sup> “Overall, then, research indicates that self-enhancement, exaggerated beliefs in control, and unrealistic optimism can be associated with higher motivation, greater persistence, more effective performance, and ultimately, greater success.” (199)  
Taylor, S.E. and Brown, J. (1988)

important in practice. But this is not the only morally relevant aspect of human nature to have this sort of problem. There is also what is referred to as the belief in a just world. This phenomenon was first studied in detail by social psychologist Melvin Lerner, and presented in detail in his book The Belief in a Just World: A Fundamental Delusion<sup>67</sup> Lerner demonstrated through a series of experiments that people are strongly inclined to blame victims of injustices for their own suffering, even when it was senseless to think that the victim was less deserving than others who ended up better off. The explanation for this, Lerner argued, was that people have a deep seated need to believe that the world is basically just: that people get what they deserve, in general, and do not suffer unjustly. This motivates people to act when they see injustice around them in order to correct it and restore the world to what they see as just. But when, as is often the case, that option is not available to them they will resort to other ways of preserving their belief, such as denying that the injustice was particularly bad or claiming that in the long run the experience will benefit the victim in ways that outweigh the unjust cost.<sup>68</sup> And the ability to reduce the injustice, or mitigate the suffering of the innocent victim to some extent is not typically enough to prevent this reaction. It shows up even when people have the ability to correct the injustice, but only at significant cost to themselves.

In these cases, people will also resort to blaming the victim for the injustice, and claiming that the victim brought it upon themselves by their actions. And, when this is unavailable, they may simply take the victim to be an inferior sort of person who, as a

---

<sup>67</sup> Lerner, M. (1980)

<sup>68</sup> For example, they will argue that severe poverty will lead people to appreciate the simpler things in life, or that persecution will strengthen their character, or that succeeding in the face of injustice is somehow

result, deserves the treatment that they receive. This is not a conscious or explicit motive – in fact the people involved had no idea they were doing this, or that their judgments were being affected. But when supposed experimental subjects received painful electric shocks, entirely as a result of random selection (others subjects were assigned to groups that would not receive the shocks), observers still rated them poorly across a wide range of criteria – even saying that they were less attractive. And the worse their suffering was the more severely they were condemned. Even more interestingly, if they were presented in the experiment as especially good people, by having them agree to undergo the painful treatment in the service of others, their evaluations were even harsher.

This same effect turns out to work in the opposite direction as well. People tend to think more highly (and evaluate in more positive terms) people who are randomly benefited, or enjoy an unjust advantage.<sup>69</sup> Also the more responsive people are to the idea that the world is basically just, the more they are susceptible to this effect. People who more readily agreed with sentences like “It is rare for an innocent man to be wrongly sent to jail”, “In any business or profession, people who do well at their job will rise to the top”, or “Although evil men may hold political power for a while, in the general course of history good wins out” were both more likely to engage in this sort of victim blaming, and did so with greater severity.<sup>70</sup>

There is a certain irony here. We have a deep and fundamental psychological need for justice. Injustice makes us deeply uncomfortable. And the result of this is that we have a need to believe that the world is basically just. If we cannot maintain that

---

<sup>69</sup> Ellard, J.H. and Bates, D.D. (1990) In this case the example involved lottery winners.

<sup>70</sup> This is the Rubin and Peplau BJW scale. The results discussed by Lerner are from: Rubin, Z. and Peplau, L.A. (1973)



belief in the face of injustice, by intervening effectively to correct it, we will maintain it by changing our other beliefs about the world, and the people in it. We will come to believe that victims of injustice brought it upon themselves, or are not actually suffering, or that they are simply inferior people who do not deserve better treatment. And usually we will do this without even recognizing that we are doing it. We even lie to ourselves and claim to believe that we do recognize the injustice of the world: that we see the world in a hard eyed, realistic fashion, as a tough and often unfair place where people primarily act for their own self interest. But this is just a useful fiction; a way to seal ourselves off from our very real desire for justice, and to protect our deeper belief that the world is just. It is largely a charade, or a social fiction that we maintain to avoid having to consider the contradictions between our experiences and our beliefs.

This, for example, underlies the effectiveness that many charities find in soliciting donations by selling inexpensive goods for higher than normal prices – it maintains the fiction that the exchange is taking place in the market, so to speak, rather than being a case where someone is motivated by a concern for justice. In a fascinating but unpublished study a group of researchers, including Lerner, tested the extent to which charitable donations are actually affected by pretending that people involved are purchasing things (in that case, decorative candles) being sold in support of a cause rather than donating money directly.<sup>71</sup> In cases where the need for help was moderate the difference between offering an exchange and asking for a donation was minimal (in this case, they were raising money for a training program for disabled children). But in cases where the need was great, and the demands of justice were correspondingly stronger there

---

<sup>71</sup> As described in Lerner, 1980.

was a substantial difference (in this case, they were raising money for a program to help handicapped and emotionally disturbed children recover and gain the skills needed to cope with normal activities). Offering an exchange resulted in over three times the amount received when asking for a donation – and roughly three times the amount received in any other case. There was also a much higher rate of response among people asked to give money (by purchasing a decorative candle). Somewhat ironically, it allowed people the option of pretending that they were remaining uninvolved with an attempt to remedy an injustice, and so preventing them from having to fully acknowledge its existence. Instead it allowed them to pretend they were merely engaging in a normal exchange of goods, something that takes place in a context where people are socially licensed to act purely on self interested grounds, gave them the space in which to actually respond to their desire for justice, and resulted in a far more effective fundraiser.

As before, though, it is worth asking if this is anything more than a common vice. It could just be a normal if unfortunate tendency that we have to avoid hard choices, or an attempt to avoid having to think in non-self-interested terms. It could even just be a normal sort of human irrationality, changing the wrong beliefs in response to evidence that contradicts what we'd prefer to believe about the world. Like the egocentric bias, however, our tendency to go wrong in this way cannot be easily dismissed since important goods hang on it. The people most susceptible to going wrong in this way, blaming victims, denying injustices, and even denigrating people because they are suffering unjustly, are also the ones most motivated to help when they *can* correct the injustice.<sup>72</sup> Even more than that, and almost exactly like the egocentric bias, people who

---

<sup>72</sup> Lerner, 1980.

are more responsive to the idea that the world is basically just are better off as a result. A confidence in the basic justice of the world is a critical part of undertaking long term projects, or at least ones which involve acting in socially acceptable ways.<sup>73</sup> People who score higher when tested on how strongly they believe in a just world are better at dealing with adverse circumstances<sup>74</sup>, report more positive affect and feel bad less often<sup>75</sup>. They report less stress, lower rates of depression, and higher life satisfaction overall compared to people who score lower than they do.<sup>76</sup> And though this is specific to the belief that the world is just with regards to one's self (and not necessarily for others), it “may ultimately contribute not only to emotional wellbeing but also physical health.”<sup>77</sup>

As a result we have a pretty serious problem. The extent to which we care about justice in the world seems, itself, to lead us towards bad, or even unjust actions at the same time that it motivates us to promote justice. In fact, it seems to do so simultaneously, through the same mechanism. While there is still a temptation to ascribe this to a normal, or even distinctive human weakness, that is, to our tendency to refuse to accept problems that we cannot solve, this is not enough to get us off the hook. The same failing that drives us to make these mistakes is part of other very important abilities of ours. Our ability to cope with life's stresses, and to make and pursue long term goals depend, to varying extents, on our succumbing to this sort of irrationality. So there is not just a failing here, there is a tradeoff between different failings. We need to fail with

---

<sup>73</sup> Hafer, C.L. (2000)

<sup>74</sup> Ritter, G., Benson, D.L., and Snyder, C. (1990)

<sup>75</sup> Ibid.

<sup>76</sup> Lipkus, I.M., Dalbert, G., and Singer, I.C. (1996)

<sup>77</sup> Ibid. (675)

regards to justice, in this particular way and to a reasonable extent at least, in order to succeed in other ways. And this conflict cannot be resolved in any ideal way. To argue that this, like some tradeoffs in life, simply admits of a mean, or has an ideal balance point that a truly virtuous agent would see would be unreasonable. There is little sense in suggesting that true virtue would involve blaming innocents suffering unjustly for their suffering in the right way, or to the right extent. And similarly a picture of the ideally virtuous agent who does not go wrong in this way would be a picture of a person who simply did not care about justice, or at least did not care about it in any recognizably human way. Or it might be a picture of an agent whose ability to pursue long term goals was severely impaired, or whose life was uniformly miserable. And none of these options are particularly appealing.

#### 4.

Finally, I would like to discuss a group of biases that are usually referred to as inter-group biases. We have a strong tendency as humans to form groups, even for the most minimal of reasons. Furthermore once we do we tend to favor members of our group over members of other groups, not just in affection but also in distributions of goods, or even basic evaluations of them. As with many biases like this, it is something that we do without intending to, and often without noticing. We simply experience “more positive feelings towards others who are classified as part of the in-group and are more responsive to their needs.”<sup>78</sup> Similarly, though more problematically, the reverse applies to members of other groups. We often evaluate members of other groups less favorably, and are significantly less responsive to their needs. In fact we are substantially

---

<sup>78</sup> Fiske, S. T., Gilbert, D. T., and Lindzey, G. (2010) P.1112

less capable of empathizing with people who we take to be members of a different social group than ourselves.<sup>79</sup> In one study participants even preferred to benefit their own groups over benefiting both groups to the same degree equally. When offered another choice, between significantly benefiting both groups equally, and giving their own group more than the other, they chose the second, even though in that option their own group received less benefit than it would have received had the distribution been equal!<sup>80</sup> Our tendency to think better of people whom we see as within our social group, or whom we see as close to us in some significant way, and to prefer them to others when it comes to distributing goods is to some extent distinct from our tendency to think poorly of others or to be less responsive to their needs. At least they are conceptually distinct. But it is unlikely that the two phenomena are unrelated to each other, and given their common cause there is a significant chance that a bias in favor of in-group members is strongly linked to or even causes higher out-group hostility. This is something that should be worrying.

The reason that a possible connection between the two is worrying is that in-group bias seems relatively unproblematic. It certainly represents some sort of failing on our parts, but the failing seems mostly to take the form of being overly nice when

---

<sup>79</sup> “Taken together, our research provides direct evidence that people are less likely to show prefrontal right alpha asymmetry when observing sad outgroup others. We take this as evidence that people generally do not vicariously feel the emotional and motivational states of those they categorize as outgroup members.”

Gutsell, J. N. and Inzlicht, M. (2012) p. 601

<sup>80</sup> “In other words, when the subjects had a choice between maximizing the profit for all and maximizing the profit for members of their own group, they acted on behalf of their own group. When they had a choice between profit for all and for their own group combined, as against their own group's winning *more* than the outgroup at the sacrifice of both of these utilitarian advantages, it was the maximization of *difference* that seemed more important to them.”

Tajfel, H. (1970) p. 101-102.

evaluating members of our in-group, or overly responsive to their needs, or overly willing to consider their interests in choosing actions. And as far as failings go that does not seem that bad. It could even be argued that this is simply a result of a general tendency to think less of people than they deserve, or to fail to take their interests seriously enough, and one which is corrected in close relationships. The appearance of a bias could simply be the result of having a clearer view of people one takes to be closer to one's self than others. This is not what I intend to argue: in-group bias does seem to represent some failing on our part, if a minor one, which could lead us astray at times.

Out-group bias, on the other hand, seems more directly worrying, since it seems to lead us astray in significant and occasionally harmful ways, even preventing us from recognizing the emotional states of others at times. So there is a two-fold worry here: the first is that in-group and out-group biases are interconnected, so that having one involves having the other; the second is that one or both might be unavoidable. And I think there is good reason to believe both of these claims: that in-group biases and out-group biases are fundamentally linked, and also that both are a result of one of the most important goods in life. Being close to others, or bonding with other people, can result in significant harms to others.

The way that in-group and out-group biases seem to mirror each other is suggestive of a connection, but not necessarily conclusive. It is possible to imagine someone whose attitudes towards members of other groups is not affected by belonging to a particular group, for example. And even if they are connected it is not clear why a truly virtuous agent would not simply lack both biases. If either could be avoided, or

simply not linked to any significant good for human beings these would simply be characteristic vices of human beings, like any number of other ones. The best evidence to think otherwise, to think that in fact these biases are both connected to each other and necessary, is in a slightly different area of psychology, specifically one concerned with the direct operations of the brain itself. So here I think it will be profitable to move away from social psychology, and to talk about the brain, and specifically the neural peptide oxytocin.

Oxytocin is somewhat famous for its central role in human bonding, especially maternal bonding. But it is also involved critically in love in general, and trust generally. As with all science relating to the brain, our understanding of precisely what is going on is shaky at best, but we do have good reason to take it to be significantly involved in caring relationships generally. Oxytocin, perhaps most famously, is associated with breast feeding in mothers. Studies have also shown that couples with higher oxytocin levels are more supportive and affectionate, and lower levels of oxytocin were associated with more negative interactions.<sup>81</sup> Oxytocin levels rise among dog owners after petting their dog for several minutes – and they rose in the dog as well.<sup>82</sup> The effect even shows up as a result of eye contact alone!<sup>83</sup> People dosed with oxytocin also display distinctive changes in behavior. It seems to increase the extent to which people behave trustingly. In one study subjects were presented with a betrayal of trust, and ones with artificially increased oxytocin levels continued to act with trust long after the control subjects had

---

<sup>81</sup> Churchland, Patricia S. (2011) p. 56

<sup>82</sup> Handlin, L., Hydbring-Sandberg, E., Nilsson, A., Ejdebäck, M., Jansson, A., and Uvnäs-Moberg, K. (2011)

<sup>83</sup> Nagasawa, M., Kikusui, T., Onaka, T. and Ohta, M. (2009)

stopped.<sup>84</sup> People who have a particular genetic mutation that affects oxytocinergic regions of the brain, especially in the hypothalamus, have trouble forming caring relationships with other people, or acting in normal social ways with others.<sup>85</sup> This is not to say that oxytocin is responsible for caring or bonding in humans, but there is good reason to believe that whatever the mechanism really is, deep down, oxytocin is significantly involved.

Unfortunately oxytocin also has a distinctly meaner side to it, and appears to play a significant role in some decidedly negative behaviors as well. When it comes to dealing with members of an out-group, increased oxytocin levels decrease the extent to which subjects felt bound to follow certain social rules. It decreased their tendency to follow basic norms of fairness, and to consider the perspectives of others.<sup>86</sup> The positive effects of oxytocin appear, in fact, to be limited to what the subjects perceive as belonging to their particular social group at any given time. One study concluded that “There is no doubt that oxytocin is implicated in the development of trust, empathy and prosociality, but these tendencies appear to be limited to individuals belonging to one's in-group”<sup>87</sup> And worse than that, there is evidence that increasing oxytocin levels in people *increases* their level of hostility towards out-group members.<sup>88</sup> In one study administering oxytocin to subjects increased the extent to which they felt envious of other subjects who gained more money than they did. When they gained more money

---

<sup>84</sup> Churchland, Patricia S. (2011) p. 79.

<sup>85</sup> Tost, H. et al. (2010)

<sup>86</sup> Radke, S. and de Bruijn, E. (2012)

<sup>87</sup> De Dreu, C. et al. (2011) p. 1265.

<sup>88</sup> De Dreu, C. et al. (2010).



themselves it also increased the amount of pleasure they took in having gained more than the other subject, and even increased gloating behaviors.<sup>89</sup>

These unfortunate effects of oxytocin indicate that there really is a serious problem here. A chemical that plays a significant, if not central role in basic human bonding and care, in love, friendship, trust, cooperation, empathy, and so on, is also involved in some very ugly behaviors as well. It is involved in hostility towards members of other social groups, ignoring basic norms of fairness when it comes to outsiders, and even increased gloating. In effect, it is significantly involved in many of the nastiest elements of out-group bias. And these two different sorts of effects are both affected in the same ways by oxytocin levels. What this indicates should be troubling: that the very thing involved in drawing us together, in caring relationships, simultaneously seems to set us against each other. The effect of oxytocin on the brain suggests that there is less difference between 'us' and 'us *versus them*' than is comfortable, especially if it indicates a tradeoff between two valuable features of our lives.

In this case we see the same pattern as before. The effects of caring for people, to forming bonds with other human beings, include treating other people less well than we otherwise would have, or even being directly hostile to them on some level. And this is not an unfortunate vice that we have, but linked to the mechanisms involved in caring for others. Caring for people seems to lead us towards taking ourselves to be in competition, or even conflict with other people, and we can often act badly towards them as a result. To suggest that an ideally virtuous agent would not experience this effect is a little strange: at the least we would need to imagine them responding to neurotransmitters in

---

<sup>89</sup> Shamay-Tsoory, S., et al. (2009)

ways that we do not appear to respond. If our idea of ideal virtue requires structural changes in the limbic system the extent to which it is *human* virtue is suspect. But without forming these bonds we would have deeply, profoundly impaired lives – if we had any at all! It is difficult to imagine a more significant and necessary good in human lives. And it is a good that we must act badly to achieve. As a result this is yet another case in which we are simply intrinsically, naturally defective. We cannot operate in the world without being deluded about our place in it, despite needing just that information to make fully informed choices. We cannot care about justice without being unjust – and without caring about injustice we do not act well, and do not do well either. And caring for some people seems to set us, unreasonably, against others.

At this point, though, there are two possible worries that arise. First it is important to consider the question that came up earlier when talking about human spines. If there is a necessary defect of some kind here, or an unavoidable tradeoff when it comes to how we react to situations calling for, for example, justice, then why not just say that there must be some ideal balancing point maximizing the likelihood of acting well and minimizing the likelihood of acting poorly? The ideally just human being might still react imperfectly in some circumstances, but would still represent the optimal balance when it comes to the tradeoff between caring about injustice and unjustly blaming victims of injustices. They would care enough about addressing injustice that the benefits of addressing it would outweigh the harms of treating people unjustly, and be disinterested to the extent of the harms is outweighed by the benefits of addressing unjust

circumstances. In other words, why not simply take the best we could manage while still being recognizably human to be the ideally virtuous agent?

Allowing our picture of an ideally just agent to represent an ideal balance point between the tradeoffs involved in human psychology seems an appealing and humane approach. It resembles how we act when dealing with people whose actual virtue was formed under the normal constraints of the world, rather than in idealized cases. But while it is appropriate when interacting reasonably with decent people this same approach would have unfortunate consequences when dealing with the possibility of ideal human goodness. Taking the standard for human goodness to still involve a disposition to, occasionally, act badly would make it difficult to take the badness of, for example, blaming innocent victims for their suffering seriously. At the very least we would be forced to say that even if doing so was inappropriate in that case it was nevertheless admirable, that doing it reflected well on the person involved. A truly virtuous person could reasonably, or would be disposed to look back on cases where they had done so and feel a sense of satisfaction or pleasure at their actions. And the other side of this result is even more disturbing. In a case where the ideally virtuous agent inaccurately or unfairly blamed the victim of an injustice failing to do that would reflect poorly on someone: responding justly to a situation would indicate some moral failing on their part. In practical situations it is possible to imagine a situation where someone would act worse as a result of a praiseworthy disposition, and where a worse person would end up acting better as a result. But this is a result of the virtues of normal human beings. It should not

be true of the ideally virtuous agent that sets the standard according to which people are better or worse.

Second the empirical nature of psychological research and neurology means that their conclusions are by their nature defeasible. It is possible that further research may reveal that the previous evidence for one or more of the cases discussed, no matter how significant, was the result of other factors and does not reflect a case in which perfection is conceptually impossible. If this is the case, however, it does not mean that this kind of natural defect is impossible or even unlikely in human beings. And this means that accounts which start from a picture of a good life for human beings specifically are still faced with the initial problem mentioned above. They rest on an empirical assumption about how human beings work and an assumption which, given the evolutionary history that shapes us, is at best unlikely to be true.

## 5.

But this raises an important question. It seems as if our psychology prevents us from being able to be ideally good with regards to these things: we face a tradeoff between different possible flaws. As a result it is hard to see how someone could, as a matter of logical possibility, get things right with regards to how they think and feel about them. And yet these seem like places where moral evaluations are significant and important to us. If true (human) virtue is impossible in these cases, or if these are ways of thinking about what it is to be human that entail that we cannot be truly virtuous, isn't this a reason to revise that understanding? It is not even clear if there *are* virtues related to justice, given what I have argued. There are certainly ways of being better or worse

off with regards to justice, but even the most just person is still, to at least some significant extent, unjust. And, more importantly, I have already argued that moral evaluations are not different in kind from other ways of evaluating human beings. They are only separated from others ones by our goals and the uses to which we put them. As a result moral evaluations are open to revisions, especially for those pragmatic reasons, just as we have done in the past when it becomes clear that some vice or other is better addressed in other contexts. If the kinds of things that we are used to evaluating end up giving us an account of naturally morally defective human beings, then it could just be that we have a good reason to look for a different group of evaluations which would not end up this way. This might be a difficult thing to manage, or take a very long time. It would be a substantial shift. But it would not be impossible, since these things do change over time, as our capacities and understanding of human nature changes.

We might have practical reasons to try to change our ways of thinking about morality, as well. If our motives in evaluating things are to help advise people on their actions then it might be better to avoid cases where genuine excellence is impossible. And if we are praising or blaming people, it might be better to avoid cases where some blame would always be appropriate, even if we would have no reason to actually criticize them. A less discouraging account could simply be more effective at inspiring actions, or helping people to do well. And if so then it would be irrational not to try to change some of the ways we think about moral evaluations, since that was exactly the purpose that set them off from other evaluations in the first place. An understanding of humanity as perfectible, if only theoretically, or as having the possibility of an ideally good life might

be more appealing to us. Even if one understanding of moral evaluations, or humanity is consistent with physical constraints internal inconsistencies make it more difficult to use them. Any trade offs or internal inconsistencies make it much harder to identify what we should aspire to be.

The alternatives to an understanding of ourselves as naturally morally defective are worse. An understanding of what it is to be a just person that we could achieve, one without the tradeoffs inherent in our current understanding, would be incapable of serving the purposes to which we put our current understanding. And the same is true for understanding how we should go about caring for other people in our lives, or the role of knowing who we really are and what we are like. It would be a hollow view, emptied of some of the most important things in our lives. In a similar vein an understanding of humanity removed far enough from the messy details of how we go about things in reality that there was no logical problem with the idea of a ideal human being, or human life, would be useless. An understanding of what it is to be a just person abstract enough to avoid dealing with the constraints imposed by our psychology could count as an account of a just thing. But it would not be a just *person*, any more than an understanding of ideal health as involving gills instead of lungs would be. It would be an understanding of ethics for a different sort of thing entirely, and any commitment to the idea of what it is to be a good human being is inconsistent with it.

Most importantly, revising our understanding of which evaluations count as moral, could only come at the cost of our ability to handle the situations in which our defects are important in the way we do now. This is what revising our sense of what

count as moral evaluations amounts to. By no longer treating ways in which we cannot truly excel as subject to the sort of advice or blaming that marks out moral evaluations, we would no longer be able to treat them in the way that we do now. This is not to say that we could no longer make these evaluations, just that they would not have the significance that they do now, or the bearing on what we take to be important in morally evaluating people. While this might make giving advice easier in general, since we would no longer be forced to consider some very hard cases, it would only do so by removing our ability to talk about those same hard cases in the way that we currently take to do. And some of the ways in which we are naturally morally defective are very significant in thinking about how our lives and the lives of those around us go, and for how we understand ourselves as human beings. This would amount to a sort of surrender: a refusal to address the ways in which we could improve, but never be truly good. In effect, revising our general understanding of the boundaries of morality out of a worry that people would give up on addressing the hard problems would simply be to give up on addressing those hard problems at all. And that is too high of a cost for the whatever benefits we would get from doing so.

Looking at medical evaluations instead of moral ones can help make this trade off clearer. At the moment some of the evaluations, and some of the problems we face, are considered under the general category of medical evaluations or problems. And in general the grounds on which we pick out those particular evaluations and problems are practical ones. We might have an account of important differences between medical and moral problems, for example, or worry about various marginal cases. But in general if

some defect can be treated or managed medically then we take it to be a medical problem, and if it cannot, or cannot effectively be treated that way we do not.<sup>90</sup> But treating or managing a problem is not the same as curing one, and many genuinely medical problems cannot be *fixed*. We might be able to manage the symptoms of, for example, schizophrenia or lower back pain. But that is the best that we can do, and even there our abilities are limited. We could argue that we might be better off separating those problems off and no longer treating them as generally medical problems. If we did so we might take doctors to be more trustworthy or reliable, or we might think people would be more likely to go see doctors in general. It is easy to see how we might get exactly these results. We would, after all, be eliminating many of the messy, problematic and often disappointing medical experiences that we currently have to deal with. We could, with one stroke, dramatically increase the health of the entire population! But we could only do this by deciding to no longer address the suffering those health problems can cause. And it is better to accept even a necessary imperfection than to remove the ability to address its hard, messy, or tragic results.

The result is that there is an unavoidable tension between an understanding of what it is to be a human being that gives us a consistent, logically possible picture of a fully good human life and an understanding of what it is to be human that can genuinely address important parts of our lives. The richness and power of our understanding of humanity and the extent to which we can see our nature as, even theoretically, perfectible

---

<sup>90</sup> This may seem an odd or strong claim, but there are active debates about how far we should go in medicalizing or pathologizing various problems people have, and they are generally in just these terms. Some people argue that some problems that we may be able to treat (to some extent) are important parts of the human condition and that the suffering they cause is a part of life that we should not interfere with, for various reasons. Others argue that our ability to address them even in a limited way means that we should. I take these debates to be a general working out of precisely this sort of tension.



are in conflict. The question is only how to balance them against each other in our moral practices: which evaluations to take as moral, and how best to understand what it is to be human or to live a human life. We have good reason to weight that balance in favor of our ability to address important things in our lives, and to choose difficulty over easiness in application, because the cost of avoiding that difficulty is our ability to do what we care about in the first place. And so we have more reason to prefer an understanding of ourselves as naturally, morally defective to one in which we are not, intrinsically, bad.

## **Conclusion**

I have focused here on accounts of human virtue fitting within the Aristotelian tradition. These are usually considered eudaimonistic accounts. They start by taking there to be a good life for human beings. This good life is one proper to the species, and not necessarily what we might want or enjoy in life, and contains certain characteristic elements, including things like loving relationships, fulfilling projects, health, and so on. Living well in this respect requires both particular external goods such as the resources needed for health and moderate pleasures and also being good in certain respects. It is tied directly to our understanding of what it is to be a human being, and not set by some standard external to that. It is constrained by empirical facts about human beings, and what is characteristic of us. Moral virtues are a part of what make us good human beings, or good examples of the species.

The virtues are things, among other goods, needed to live well in the human way. Instead of talking generally I focused on a specific characteristic example of this kind of picture: that of Phillipa Foot. I started by arguing that the account of human virtue proposed by Phillipa Foot in Natural Goodness has a problem. She argues that what it is to be a good human being to be determined by what is needed to live a fully good, and fully human life. But the only way to make sense of what it is to live a fully good human life is in terms of an understanding of what it is to be a human being in the first place, and that is exactly the account of what makes human beings good and bad that was supposed to follow from it. I suggested that instead Aristotelian accounts should focus on what it is for the functions that parts of human beings have independently from a general or final end for human beings as a whole.

I followed, in the second chapter, by arguing that it was possible to talk about functions without referring to a final end, or an ideally good life, for human beings. I first discussed the literature regarding attributing proper functions functions to things in the philosophy of biology. But I argued that none of the accounts that are useful in that context have the resources needed to explain the kind of moral evaluations present in Aristotelian accounts of moral virtue. Instead I suggested that we can understand functional evaluations, of the kind needed, as bound up in an understanding of the membership criteria of some kinds of things, in this case the parts of human beings that we evaluate in moral, and some other, ways. Then, in the third chapter, I went on to talk about moral evaluations in particular, because the way of talking that I suggested in the second chapter is not sufficient to distinguish between moral evaluations and some other kinds of functional evaluations of parts of human beings on its own. I argued that we can pick out moral evaluations of human beings from other ways of evaluating human beings, such as medical ways, while still taking them to be fundamentally the same general kind of thing, without any special metaphysical status. Moral evaluations can be picked out from functional evaluations of human beings generally by the characteristic use we put them to and, specifically, by their place in encouraging certain behaviors and discouraging others.

Finally, in chapter four, I argued that starting from an understanding of how parts of human beings function rather than a good human life has a further advantage. Even if the problem in chapter one did not exist taking an ideally good human life to determine what is and is not a virtue in human beings requires assuming that that life is theoretically

possible. Whether or not human nature is coherent enough to allow for this is an empirical question: the things that make us better might be inconsistent, or being good might conflict with having some of the things necessary to truly live well. And I argued that this is a significant problem because what makes human beings better or worse comes with certain innate tradeoffs that makes it impossible for there to be a human being good in all respects, and fitted to live a fully good human life. In actual practice of course it is not reasonable to expect any such person to exist, but what I have argued is that it is impossible even in theory for such a person to exist. And as a result starting from a notion of such a thing when constructing an account of the virtues turns out poorly. The account of moral evaluations that I suggested in place of accounts starting from an ideally good human life, however, has no such problem. I concluded by arguing that an understanding of human nature which is inconsistent in this way is valuable in its own right.

This does not mean that we cannot be better or worse, or that it is not valuable to be a better person than a worse one. The moral evaluations that we make of ourselves and others are still true or false. And we can still be better or worse even if there is no unreachable endpoint to strive towards. After all simply because human spines are poorly suited to their task, and seem inevitably problematic, this does not mean that some are not healthier than others. And it does not mean that having a healthier back does not make our lives better, or that it is not a good idea to avoid injury. However taking seriously the idea that we are fundamentally imperfectible does complicate how we address certain questions when it comes to health generally. And it affects the extent to

which we should value improvement in some respects rather than others. Beyond a certain point there is very little point in trying to improve the state of one's back, given its eventual degeneration.

Taking seriously the idea of moral imperfectibility also has consequences for how we go about doing things. Even though it does not make it reasonable to disregard questions about being more or less virtuous it should still affect how we think about cultivating virtues in ourselves and others. Many virtues come, in practice, with risks: honest people may tend towards tactlessness, courageous people towards rash actions. But being reasonably good in some respect can come with a cost in its own right. It can lead us to act poorly, not simply because of our tendency to mistake what virtues may be called for in some situation or other but because the virtue itself comes with an innate tradeoff proper to itself. Similarly being better in some respect may not be, on balance, a good idea. It may leave us worse off in our lives, not simply when it comes to getting what we want but potentially as far as living a fully human life. It may even, given the tradeoffs involved, leave us a worse person overall.

There is, in other words, a genuine cost to becoming a better person, as much as there is to becoming worse in some respect. This is an important, and significant feature of how we go about thinking of our own characters. It is something that the generally Aristotelian views I have been talking about can accommodate without too much trouble. In practice there are unlikely to be any ideally virtuous agents and, even if there were, living an ideally good life as a result of their virtue depends on any number of other contingencies. So it can easily be the case on those views that the virtues that we can aim

at in life, as opposed to the ideal ones, could leave us worse off as well. Taking virtue to be an important part of a good life is sensible; taking it to be necessary on a theoretic level for an ideally good human life is tempting; but taking minor or moderate improvements in our character to always leave us better off in practice is foolish. When it comes to our actual lives character is a blunt instrument even with the mediation of reason. It is tempting, especially if we approach questions of virtue from a generally eudaimonist position, to take it to only apply to very abnormal or extreme cases, but I think doing this is a mistake. On balance it may be the case that for the most part we are left better off by becoming better, but that in itself is a dangerous assumption. And there is value in foregrounding the risk of being left worse off, and making it an explicit part of our moral outlook rather than leaving it as an awkward possibility. This is what I take to be the advantage of an understanding of human nature that leaves us defective by nature, rather than simply by happenstance.

## **Bibliography**



- Alloy, L.B. And Abramson, L.Y. (1979) "Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?" *Journal of Experimental Psychology*, Vol 108 No. 4 pp. 441-485.
- Amundson, R. and Lauder, G (1994) "Function without Purpose: The Uses of Causal Role Function in Evolutionary Biology" *Biology and Philosophy* Vol. 9 pp. 43-69
- Annas, Julia (2009) Intelligent Virtue. Oxford University Press; Oxford.
- Aquinas, T. (1947) Summa Theologica, trans. Fathers of the English Dominican Province. Benziger Brothers, New York, NY.
- Bekoff, M. & Allen, C. (1995) "Teleology, function, design, and the evolution of animal behavior." *Trends in Ecology and Evolution* 10(6): 253-255.
- Bigelow, J.B. And Pargetter R (1987), "Functions" *Journal of Philosophy* Vol 84 pp. 81-97
- Buss, S. (1999) "Appearing Respectful: The Moral Significance of Manners" *Ethics* Vol. 109 pp. 795-826.
- Calhoun, C. (2000) "The Virtue of Civility" *Philosophy and Public Affairs*. Vol. 29 pp. 251-275.
- Churchland, P. (2011) Braintrust: What Neuroscience Tells Us About Morality. Princeton University Press, Princeton N.J.
- Cummins, R. (1975) "Functional Analysis" *The Journal of Philosophy* Vol. 72, No 20, pp. 741-765.
- Davies, Paul. (2003) Norms of Nature: Naturalism and the Nature of Functions. The MIT Press, Cambridge MA.
- De Dreu, C., Greer, L., Handgraaf, M., Shalvi, S., Van Kleef, G., Baas, M., Velden, F., Van Dijk, E., and Feith, S. (2010). "The Neuropeptide Oxytocin Regulates Parochial Altruism in Intergroup Conflict Among Humans". *Science*, Vol. 328 No. 5984 pp. 1408-1411.
- De Dreu, C., Greer, L., Van Kleef, G., Shalvi, S. and Handgraaf, M. (2011). "Oxytocin promotes human ethnocentrism". *Proceedings of the National Academy of Sciences*. Vol 108, No. 4, pp. 1262-1266.
- Driver, Julia (2001) Uneasy Virtue. Cambridge University Press, New York NY.

- Ellard, J.H. and Bates, D.D. (1990) "Evidence for the role of the justice motive in status generalization processes." *Social Justice Research* Vol. 4. pp. 115-134.
- Enc, B and Adams, F (1992) "Functions and Goal Directedness" *Philosophy of Science* Vol 59, pp. 636-654
- Fiske, S. T., Gilbert, D. T. and Lindzey, G. eds. (2010) Handbook of Social Psychology, Vol 2. John Wiley & Sons, Hoboken, NJ.
- Foot, P. (2001) Natural Goodness. New York, NY.
- Godfrey-Smith, Peter (1994) "A Modern History of Functions" *Nous* Vol 28, No. 3 pp. 344-362
- Godfrey-Smith, P. (1993) "Functions: Consensus Without Unity" *Pacific Philosophical Quarterly* 74, pp. 196-208
- Griffiths, P.E. (1993), "Functional Analysis and Proper Functions" *The British Journal for the Philosophy of Science* 44: 409-422
- Gutsell, J.N., and Inzlicht, M. (2012). "Intergroup Differences in the Sharing of Emotive States: Neural Evidence of an Empathy Gap" *Scan* Vol. 7. pp. 596-603.
- Hafer, C.L. (2000). "Investment in long-term goals and commitment to just means drive the need to believe in a just world" *Personality and Social Psychology Bulletin*, Vol. 26. pp. 1059-1073.
- Handlin, L., Hydbring-Sandberg, E., Nilsson, A., Ejdebäck, M., Jansson, A., and Uvnäs-Moberg, K. (2011) "Short-Term Interaction Between Dogs and Their Owners: Effects on Oxytocin, Cortisol, Insulin and Heart Rate – An Exploratory Study." *Anthrozoös*. Vol. 24 No. 3 pp. 301-315.
- Hurka, Thomas (2001) Virtue, Vice and Value. Oxford University Press, Oxford.
- Hursthouse, Rosalind (1999) On Virtue Ethics. Oxford University Press, Oxford.
- J.J. Loud, assignee. Patent 392,046. 30 Oct. 1888.
- Lerner, M. (1980). The Belief in a Just World: A Fundamental Delusion. Plenum Press; New York NY.
- Lipkus, I.M., Dalbert, G., and Singer, I.C. (1996) "The Importance of Distinguishing the Belief in a Just World for Self versus for Others: Implications for Psychological Well-being." *Personality and Social Psychology Bulletin*, Vol. 22. pp. 666-677.

- Love, Alan. (2007) "Functional Homology and Homology of Function: Biological Concepts and Philosophical Consequences." *Biology And Philosophy*, Vol 22. No. 5 pp. 691-708.
- MacIntyre, Alasdair (1999) Dependent Rational Animals: Why Human Beings Need the Virtues. Open Court; Chicago IL
- Millikan, R.G. (1989a) "An Ambiguity in the Notion 'Function'". *Biology and Philosophy* Vol. 4 pp. 172-176
- Millikan, R.G. (1989b) "In Defense of Proper Functions", *Philosophy of Science* 56: 288-302
- Nagasawa, M., Kikusui, T., Onaka, T. and Ohta, M. (2009) "Dog's gaze at its owner increases owner's urinary oxytocin during social interaction." *Hormones and Behavior*, 55(3):434-41.
- Neander K. (1991a) "Functions as Selected Effects: The Conceptual Analyst's Defense", *Philosophy of Science* 58: 168-184
- Neander, K. (1991b) "The Teleological Notion of 'Function'", *Australasian Journal of Philosophy* Vol 69 No. 4, pp. 454-468
- Nikaido, Masata, Ying Cao, Norihiro Okada, and Masame Hasegawa. (2003) "The phylogenetic relationships of insectivores with special reference to the lesser hedgehog tenrec as inferred from the complete sequence of their mitochondria genome." *Genes and Genetic Systems* 78, p. 107-112.
- Radke, S. and de Bruijn E. (2012). "The Other side of the coin: oxytocin decreases adherence to fairness norms." *Frontiers in Human Neuroscience*. Vol. 6, Article 193 pp 1-7
- Ritter, G., Benson, D.L., and Snyder, C. (1990) "Belief in a Just World and Depression" *Sociological Perspectives*, 33. pp. 235-252
- Rubin, Z. and Peplau, L.A. (1973). "Belief in a just world and reaction to another's lot: A study of participants in the national draft lottery." *Journal of Social Issues* Vol. 29. pp. 73-93.
- Shamay-Tsoory, S., Fischer, M., Dvash, J., Harari, H., Perach-Bloom, N., and Levkovitz, Y. (2009) "Intranasal Administration of Oxytocin Increases Envy and Schadenfreude (Gloating)" *Biological Psychiatry* Vol. 66, Issue 9, pp. 864-870.
- Sherman, N. (2005) "Manners and Morals" in Stoic Warriors. Oxford University Press, Oxford.

- Slote, Michael (2001) Morals From Motives. Oxford University Press, Oxford.
- Stohr, K. "Manners, Morals, and Practical Wisdom." (2006) in Values and Virtues: Aristotelianism in Contemporary Ethics. Ed. Timothy Chappell. Oxford University Press, Oxford. pp. 189-211
- Stohr, K. (2012) On Manners Routledge, New York, NY.
- Svenson, O. (1981) (Are we all less risky and more skillful than our fellow drivers?" *Acta Psychologica* Vol. 47 No. 2 pp. 143-148.
- Swanton, Christine (2003) Virtue Ethics: A Pluralistic View. Oxford University Press; Oxford.
- Tajfel, H. (1970). "Experiments in Intergroup Discrimination" *Scientific American* Vol. 225, pp. 96-102.
- Taylor, S.E. and Brown, J. (1988) "Illusion and Wellbeing: A social psychological perspective on mental health". *Psychological Bulletin* Vol. 103 pp. 193-210.
- Thompson, M. (1995) "The Representation of Life" In Hursthouse, R., Lawrence G., and Quinn, W. eds., Virtue and Reasons. Clarendon Press, Oxford.
- Tiberius, Valerie (2010) The Reflective Life: Living Wisely With Our Limits. Oxford University Press, Oxford.
- Tost, H., Kolachana, B., Hakimi, S., Lemaitre, S., Verchinski, B.A., Mattay, V.S., Weinberger, D.R., and Meyer-Lindenberg, A. (2010) "A Common Allele in the Oxytocin Receptor Gene (OXTR) Impacts Prosocial Temperament and Human Hypothalamic-Limbic Structure and Function" *Proceedings of the National Academy of Sciences* Vol. 107, No. 31 pp. 13936-13941.
- Williams, Bernard. "Evolution, Ethics and the Representation Problem" (1983) reprinted in Making Sense of Humanity. (1995). Cambridge University Press, Cambridge. pp. 100-110.
- Williams, Bernard. World, Mind, and Ethics (1995). Cambridge University Press: New York, NY.
- Wouters, Arno. (2003) "Four Notions of Biological Function." *Studies In History And Philosophy Of Biological And Biomedical Sciences*, Vol. 34C No. 4, pp. 633-668.

Wright, Larry. (1973) "Functions" *The Philosophical Review*, Vol. 82, No. 2. pp. 139-168