MORAL SENSE AND MORAL IMPERATIVE:
An Analysis of the Biological Foundations of Morality and their Implications for Moral Theory

by

MICHAELA LEIGH WOOLLEY

B.A., Simon Fraser University, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

Department of Philosophy

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2003

© Michaela Leigh Woolley, 2003
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy
The University of British Columbia
Vancouver, Canada
Date September 26th 2003
ABSTRACT

I argue that morality is in significant part a biological phenomenon, and that this has implications for substantive moral philosophy. I begin the first chapter by arguing for the hypothesis that human morality is on a continuum with conflict reducing behaviours that have been extensively documented in some non-human primate species. These behaviours provide evidence of the presence of moral building blocks such as sympathy, empathy, and a sense of social regularity. In the second chapter, I take up the objection that morality must be conceptually distinct from social behaviour evident elsewhere in the animal kingdom because altruism is an essential component of morality, and genuine altruism cannot evolve in nature. I argue that the concept of group selection can be used to demonstrate that there is indeed room for biological altruism in nature. In the third chapter, I explain how moral building blocks such as sympathy, empathy, and a sense of social regularity provide elements from which moral systems can be constructed. In the fourth chapter, I investigate the implications that the biological nature of morality has for substantive moral philosophy. Here I argue that some highly exigent moral demands are disconfirmed on the grounds that they are not supported by our experience of moral phenomena.
TABLE OF CONTENTS

Abstract ........................................................................................................................................................................ ii  
Table of Contents ........................................................................................................................................................... iii

CHAPTER I Evolution, Moral Elements, and the Burden of Proof ...................................................... 1

1.1. Evolution and Moral Elements ................................................................................................. 2
1.2. The Burden of Proof ............................................................................................................. 8

CHAPTER II The Evolution of Altruism .......................................................................................... 14

2.1. Selfish Genes ........................................................................................................................... 15
2.2. Group Selection and the Evolution of Altruism ................................................................. 23
2.3. The Evolution of Altruism in Humans ............................................................................... 29

CHAPTER III Moral Elements and Moral Systems ........................................................................ 33

3.1. Morality and the Demarcation Problem .............................................................................. 35
3.2. Social Regularity and Prescriptive Rules ........................................................................... 40
3.3. Emotional Building Blocks .................................................................................................. 46

CHAPTER IV Moral Sense and Moral Imperative ........................................................................... 59

4.1. Exigency in Moral Theory ..................................................................................................... 63
4.2. Moral Sentiments and Feelings of Obligation ....................................................................... 68
4.3. Exigency and Disconfirmation ............................................................................................ 72

Bibliography .............................................................................................................................................................. 81
CHAPTER I

EVOLUTION, MORAL ELEMENTS, AND THE BURDEN OF PROOF

Historically, we have taken ourselves to be the only creatures capable of moral behaviour. However, recent research into the social behaviour of non-human primates is challenging the long-held and deeply rooted belief that humans are isolated as moral creatures. This new data suggests that, rather than being different in kind, human moral behaviour is on a continuum with behaviours observable in some non-human primates.

In their paper “Any Animal Whatever”, psychobiologists Jessica Flack and Frans B.M. de Waal present the results of behavioural research and observations of non-human primates which they argue provide confirming evidence for their hypothesis that the building blocks of moral systems can be found in the conflict-management behaviours of some non-human primates. Their work, however, has met with criticism from several commentators who have argued that Flack and de Waal’s research fails to provide evidence of anything moral, or proto-moral, about non-human primate behaviour. In what follows in this chapter, I will sketch Flack and de Waal’s argument, the experimental results and observations they appeal to, and the major criticisms levied against them. I will then argue that these criticisms of Flack and de Waal’s work ought to be rejected because they suffer disastrous flaws, most notably that of violating the principal of evolutionary parsimony. In doing so, they betray an untenable bias that places the burden of proof on those who postulate behavioural similarities between humans and non-human primates, rather than on those who postulate differences.
1.1 Evolution and moral elements

It is now commonly accepted as the best of available theories that life as we know it evolved from its modest chemical beginnings through the process of descent with modification. Each species has evolved as a modification on its ancestors, and those ancestors as a modification on their ancestors all the way back to life's beginnings. Although there are several processes that can explain evolutionary change, the mechanism of natural selection plays a large part in explaining the diversity of life observable today and in the fossil record. As Darwin's collection of finches from the Galapagos Islands has classically come to illustrate, speciation occurs when a splinter group becomes isolated from a general population. As mutations arise in the isolated groups, and are selected for or against according to whether or not they enhance the fitness (ability to survive and reproduce) of the organisms in their environment, the DNA of the two groups can become increasingly distinct until they are different varieties of the same species, and then two different species altogether.1

Because of the branching nature of speciation, we ought to expect that species that lie close to each other in the "tree of life" will, despite being different enough to be distinct species, manifest many more phenotypic commonalties then they will differences because of their shared ancestry. To put this another way, the "design work" that has gone into closely related species has largely just been copied over from their shared ancestors.2 The principle of evolutionary parsimony (PEP) captures this important fact.

---

1 Zimmer, pp. 88-92.
2 Dennett, p. 72. Dennett argues for the "Principle of Accumulation of Design." He maintains that, since "each new designed thing that appears must have a large design investment in its etiology somewhere, the cheapest hypothesis will always be that the design is largely copied from earlier designs, which are copied over from earlier designs, and so forth, so that actual R-and-D innovation is minimized. We know for a fact, of course, that many designs have been independently re-invented many times—eyes, for instance,
about the evolution of species. *PEP* demands that, when making comparisons, one ought not to postulate differences between closely related species unless compelled to by evidence. To put this another way, because of the way speciation works, *PEP* places the burden of proof on those who postulate differences between closely related species, rather than on those who postulate similarities.⁵

Humans are products of evolution just like every other organism we know of. This entails that our species’ phenotypic characteristics—our morphology, physiology, and behaviour—have evolutionary explanations.⁴ As members of the ape family, humans are very closely related to primates such as chimpanzees, bonobos, gorillas, and orangutans.⁵ Members of the ape family in turn are closely related to the monkey family, which includes primates such as macaques, baboons, and capuchins. Because of the degree to which these species are related, *PEP* advises us to expect a great deal more phenotypic similarities than differences, and places constraints on the postulation of differences between these species.

In beginning their argument, Flack and de Waal note that if human morality has biological origins, one ought to expect that elements of it will be present in other closely related social species.⁶ Flack and de Waal argue that such elements can in fact be found in the methods that some non-human primates use for resolving and preventing conflicts of interest within their groups. These methods, they argue, are very similar to those that humans use, and include reciprocity, food sharing, reconciliation, consolation, conflict
dozens of times—but every case of such convergent evolution must be proven against a background in which most of the design is copied.” Ibid. p. 72.

⁴ Sober, p. 663.
⁵ Zimmer, p. 263.
⁶ Flack and de Waal (2000a), p. 3.
intervention, and mediation. Flack and de Waal maintain that these methods of conflict resolution and prevention provide the raw materials from which moral systems can be constructed. They define morality as "a sense of right and wrong, that is born out of group-wide systems of conflict management based on shared values." Moral systems function by providing a set of rules and incentives to resolve competition and conflicts within the group in the interests of deriving the benefits of resource distribution and collective action. Flack and de Waal argue that the conflict management behaviours used by some non-human primate species contain the building blocks of moral systems because they reflect a concern for, and even understanding of, the predicaments of fellow group members. Further, they "are based on, and facilitate, cohesion among individuals, and reflect an effort by community members to find solutions to social conflict." Consequently, although non-human primates may not exactly be moral beings, the use of conflict-management behaviours in some of their societies suggests the presence of a sense of social regularity that parallels the rules and regulations of human moral systems.

Flack and de Waal argue that these methods used to manage conflicts of interest and preserve order appear to require or make use of such key moral components as the capacity for empathy, sympathy, community concern, and a sense of social regularity. They contend that these components are fundamental to moral systems because they help to generate connections between individuals despite inevitable conflicts of interest. By

---

7 Ibid, p. 3
8 Flack and de Waal (2000b), p. 69
9 Ibid, p. 69
10 Flack and de Waal (2000a), p. 3
11 Ibid, p. 1
12 Ibid, p. 3
13 Ibid, p. 3.
generating and reinforcing connections between group members, these moral components facilitate co-operative social interactions. They require individuals to make "commitments" to behave in ways that may not accord with their independent individual interests, but that further collective or shared interests.\(^{14}\)

Flack and de Waal appeal to numerous behavioural experiments and observations in arguing for these claims. Food sharing, for example, is a well-documented behaviour in several primate species including chimpanzees, bonobos, siamangs, orangutans and capuchin monkeys.\(^{15}\) Food sharing serves as a method by which adult members of a group distribute resources amongst themselves. Although typical resource transfers are passive, and involve the selective relinquishment of food rather than active giving, the majority of food-sharing behaviour observed appears to require fine-tuned communication about intentions and desires in order to facilitate the transfers.\(^{16}\)

Although competing hypotheses exist to explain food-sharing behaviour, Flack and de Waal find that the experimental data is most in line with the reciprocity hypothesis—that food sharing is part of a system of mutual obligations that can involve material exchange, the exchange of favours, or some combination of both.\(^{17}\) Studies of brown capuchins and chimpanzees have addressed the question of whether the reciprocity involved in behaviours like food sharing is calculated or symmetrical. Reciprocity is calculated when it is based on an ability to keep track of favours given and received, and symmetrical when it occurs as a by-product of frequent association and symmetrical

\(^{14}\) Ibid, Flack and de Waal use the term 'independent interests' to refer both to those interests that are truly independent and those whose pursuit requires engaging in competition.

\(^{15}\) Ibid, p. 4

\(^{16}\) Ibid, p. 4.

\(^{17}\) Ibid, p. 5
Flack and de Waal note that, although sophisticated and cognitively demanding, calculated reciprocity has been demonstrated in experiments involving female capuchins and chimpanzees. Chimpanzees, for example, appear to withhold favours from ungenerous individuals in future exchanges, but are more likely to relinquish additional resources to individuals who have previously groomed them. Additional evidence of calculated reciprocity is given by the phenomenon of third party conflict intervention. Third party conflict intervention occurs when an individual who is not directly involved in a conflict intervenes in support of one of the opponents. This type of behaviour has been observed in both chimpanzees and some species of macaques, where it has been shown that individuals are more likely to intervene in support of those who have intervened on their behalf in the past than those who have not.

The fact that chimpanzees, capuchins, and macaques behave as they do in these social situations, Flack and de Waal argue, implies that they have expectations about how both they and others are to behave in certain contexts. Further, they contend that observations of calculated reciprocity demonstrate how prescriptive rules can be formed when members of a group learn to recognize the contingencies between their own behaviour and the behaviour of others.

Flack and de Waal maintain that the existence of these rules and expectations reflects a sense of regularity that they venture may even be a precursor to the human sense of justice. The claim that a sense of social regularity is at work is further supported

---

18 Ibid, p. 8
by observations of individuals taking revenge on others for perceived violations of the group social code. Robert Trivers has boldly termed these behaviours instances of "moralistic aggression."  

Moralistic aggression most often manifests in the form of protest by subordinate individuals or punishment by dominant individuals. However, in several cases one member of a chimpanzee group has been observed to take on a control role. In such cases, the control member acts as a peace officer, often breaking up conflicts without taking sides. When sides are taken, as in cases of protective intervention, interventions by the control member regularly occur on behalf of the victims of aggression.  

Flack and de Waal suggest that moralistic aggression serves to help reinforce systems of reciprocity by increasing the costs of not co-operating, failing to return favours, or cheating, and in doing so contributes to social order.  

Flack and de Waal also contend that the presence of community concern is reflected by the existence of mediation in some primate groups. Mediation occurs when an individual who is not directly involved in a conflict helps to reconcile two former opponents who could not bring themselves to reconcile on their own. Group-wide celebrations have often been observed following the successful resolution of important chimpanzee conflicts. Flack and de Waal take these behaviours to reflect the presence of community concern because they provide evidence that non-human primates devote both time and energy to ensuring that their social group remains peaceful.  

Further, Flack and de Waal note that systematic data also exists in support of the hypothesis that chimpanzees engage in active consolation behaviour. Active consolation

---

21 Ibid, p. 9. The development of prescriptive rules in non-human primates will be discussed at some length in chapter three.  
22 Trivers, p. 46.  
occurs when a third party approaches and affiliates with a recent recipient of aggression. They argue that this type of behaviour suggests both that chimpanzees are able to recognize the distress of an individual who has been a victim of aggression, and that they have the capacity to be concerned enough about an individual to make attempts at relieving his or her distress. Accordingly, they take the practice of active consolation to suggest the presence of both empathy and sympathy in non-human primates.26

1.2 The burden of proof

Although I have only outlined a few of the experiments and observations that Flack and de Waal appeal to, it is enough to illustrate that they are arguing on the basis of significant and compelling data. However, as noted at the outset, Flack and de Waal's work has met with criticism from their commentators. It is to each of the most important of these criticisms that I will now turn.

I.S. Bernstein has charged that there are ways in which the behaviours that Flack and de Waal have described could be explained more parsimoniously. There is no need, he argues, to explain them in the manner that Flack and de Waal have chosen to—as evidence of community concern, or precursors to the human sense of justice.27

Peter Railton has objected that Flack and de Waal run the risk of anthropocentric projection by attributing anything like a rudimentary sense of justice or sense of community to non-human primates. It is enough, Railton argues, that we have inherited the necessary perceptual, cognitive, and motivational capacities from our pre-linguistic

---

26 Ibid, p. 18
27 Bernstein, p. 32.
ancestors that made the moral point of view possible. We neither need to discover a proto-form of the moral point of view in non-human animals, nor do we even have to have any idea what this would be.\textsuperscript{28}

Finally, Jerome Kagan has raised the objection that the behaviours that Flack and de Waal have described fail to include the most essential features of human ethics—the application of judgements of good and bad to events, the capacities of guilt and empathy for another's state, and the ability to suppress actions that would compromise the self's virtue. Because these components are lacking, he argues, the idea that human and non-human primate behaviour lies on a continuum is challenged.\textsuperscript{29} Whether or not one agrees with the features that Kagan takes to be essential for morality, or that Flack and de Waal have failed to provide evidence of all the ones he has listed, the general objection that there is something essential to morality that Flack and de Waal have provided no evidence for in non-human primates is one that must be taken seriously.

As noted at the outset however, each of these criticisms of Flack and de Waal can be rejected. Bernstein contends that Flack and de Waal fail to choose the most parsimonious explanation when interpreting the behaviours they describe. Bernstein points to the important distinction between motivation, and function—the reason that an individual performs some action, and the consequence that the action achieves. Bernstein argues that, in describing the non-human primate behaviours that they do, Flack and de Waal erroneously conflate the effect of an individual's behaviour with the motivation for performing it. Flack and de Waal indicate that there are types of primate behaviour that function reciprocally, result in the sharing of resources, reconcile antagonists, and

\textsuperscript{28} Railton, p. 58.
\textsuperscript{29} Kagan, p. 46.
console victims, but it doesn’t follow that primates perform the behaviours they do to achieve these consequences. An individual may perform an action that results in the reconciliation of antagonists simply because it “makes a loud noise go away.” As a result, Bernstein argues, it is not parsimonious to postulate that these behaviours require moral capacities such as empathy, sympathy, and community concern. There is no reason one needs to invoke such traits in explaining why individuals groom one another, seek one another out, and enlist one another’s help. Primate groups merely “establish and maintain social bonds with others and use these relationships to solve environmental problems.”

Although initially compelling, when one recalls that reconciliation, consolation, and the other behaviours in question, are being examined within an evolutionary framework, it becomes unclear exactly what grounds a call for parsimony like Bernstein’s. As Flack and de Waal argue:

If a person embraces a victim of attack, we assume empathy and perspective-taking to underlie her behaviour. On what grounds would we conclude otherwise when a chimpanzee embraces a victim of attack? To propose a different explanation in the case of these animals seems uneconomic, and from an evolutionary perspective, thoroughly puzzling, because identical actions under functionally similar circumstances are unlikely to be produced through entirely different emotional and cognitive channels.

PEP demands that one not postulate differences between closely related species unless compelled to by evidence. However, Bernstein’s appeal to parsimony rather unparsimoniously suggests that we do just that. Although we are certainly warranted in postulating that non-human primates do not do algebra, the available evidence is far from compelling that the behaviour of non-human primates fails to contain moral building blocks.

---

30 Ibid, p. 32.
It seems that it is unproblematic to assert what PEP entails about the phenotypic similarities between monkeys and non-human apes, and for the most part, between humans and their close evolutionary relatives as well—especially with regard to physiology and morphology. However, as Bernstein’s objection makes clear, there remains something unpalatable to some of us about what PEP entails about the behavioural similarities between humans and our close evolutionary relatives—at least when it is moral behaviour that we are considering. The capacity for moral behaviour has long been taken to be a distinguishing characteristic between humans and “brutes,” and has remained as one of the last bastions of human uniqueness and superiority in the post-Darwin world. As morality has historically been construed as uniquely human, the project of understanding it within an evolutionary framework is still often approached as one where the burden of proof is on those who hypothesize that human moral behaviour is different in degree but not in kind from that of our close evolutionary relatives. Yet, ironically, this placement of the burden of proof is the converse of that required by the principles of evolutionary theory.

This untenable burden-of-proof bias is also evident in Railton’s objection that Flack and de Waal risk anthropocentric projection by attributing such things as a rudimentary sense of justice or sense of community to non-human primates. In effect, Railton argues as Bernstein does, that we ought to assume that behaviours like mediation and arbitration are produced through different emotional channels in non-human and human primates. Railton goes further, though, in arguing that it is enough that we have inherited the necessary perceptual, cognitive, and motivational capacities from our pre-linguistic ancestors that made the moral point of view possible. There is no need to

hypothesize the existence of anything in any other species that is or was proto-moral. However, the very hypothesis that the moral point of view arose suddenly with humans and that no elements of it are present in closely related species is, although possible, exceedingly improbable given the way evolution and speciation work. Further, there is no reason to think that humans possess any other special capacities or abilities that lack precursors in our close evolutionary relatives.

Kagan has objected that the behaviours that Flack and de Waal have described fail to include the most essential features of human ethics. As I conceded above, this is a serious criticism of Flack and de Waal’s project. If indeed there are essential features of human morality of which no elements can be found in our close evolutionary relatives, the idea that human moral behaviour is on a continuum with behaviours observable in non-human primates is in peril. We would have compelling evidence for postulating a fundamental difference between these closely related species.

Kagan argues that the application of good and bad to events, the capacities of guilt and empathy for another’s state, and the ability to suppress actions that would compromise the self’s virtue, are essential to morality but are missing from Flack and de Waal’s account. It is not obvious that all of these features are essential to morality, or that Flack and de Waal have failed to show that elements of each of these are present in the behaviour of non-human primates. However, with the burden of proof in the right place, there is little reason to be concerned. It is up to Kagan, or any other critic, to show that particular features are required for morality, and to gather the necessary data to demonstrate that they cannot be accounted for in the behaviour of non-human primates. Only then will we have sufficient grounds to postulate that human morality is
incongruous with social behaviours observable elsewhere in the animal kingdom.
CHAPTER II
THE EVOLUTION OF ALTRUISM

In the previous chapter I argued that research and observations of non-human primates strongly suggest the presence of conflict-management behaviours that, when the burden of proof is properly placed, are correctly called proto-moral. These behaviours provide the building blocks from which moral systems can be constructed, and include reciprocity, food sharing, active reconciliation, consolation, conflict intervention, and mediation. As Flack and de Waal have noted, if human morality has biological origins, we ought to expect that elements of it will be present in closely related social species. Accordingly, the presence of these behaviours in non-human primates lends a great deal of support to the hypothesis that morality has a biological basis.

Richard Dawkins argues in *The Selfish Gene* that organisms are genetically programmed to be selfish. He argues that we cannot look to biology to explain the origins of morality, because morality involves something that defies our genetic programming—genuine altruism. A behaviour is biologically altruistic if, as a consequence of this behaviour, an organism increases the fitness of others at her own expense. A behaviour is selfish if, as a consequence of this behaviour, an organism increases her own fitness at the expense of others.¹ Humans, Dawkins maintains, are unique in the animal kingdom in that we are dominated by culture. As our biology is selfish, Dawkins argues, morality can only be a purely cultural phenomenon that exists despite our biology and not because of it. Dawkins argues that we must “try to teach altruism, because we are born selfish. . . we may then at least have the chance to upset
Quite simply, in Dawkins's view, genuine altruism cannot exist in nature. I will turn to his reasons for this claim shortly. However, if it is indeed the case that altruism cannot exist in nature, any argument for the claim that morality is a biological phenomenon will be significantly weakened. Many of the most morally praiseworthy qualities, such as honesty, charity, trust, and heroism, appear to benefit others at a cost to oneself.\(^2\) If there is no such thing as biological altruism, it would appear that, as there is something fundamental to morality that cannot be found in nature, there is a chasm between human morality and social behaviours found elsewhere in the animal kingdom.

It is the purpose of this chapter to argue that there is indeed a place for altruism in nature, as genetic strategies that direct an organism to behave altruistically can be selected for. In what follows in this chapter, I will sketch Dawkins's argument for the claim that biology is necessarily selfish. I will then show how the concept of group selection, defended in *Unto Others* by Elliot Sober and David Wilson, can be used to make a compelling case against Dawkins's position that altruism cannot evolve. In the final section I will relate this discussion specifically to the evolution of altruism in humans.

2.1 Selfish genes

Dawkins maintains that individual organisms are best understood biologically as survival machines constructed by the genes or "replicators" they contain for their promulgation. Organisms are best understood as mere survival machines because it is

\(^1\) Dawkins, p. 2.
\(^2\) Ibid, p. 3.
replicators and not individual organisms that are the real units of natural selection. In order for anything to be a unit of natural selection, it must have the properties of longevity, fecundity, and copying-fidelity. Because individuals exist briefly, in one exact copy, and fail to produce identical duplicates of themselves, individuals cannot be the units of natural selection.\(^4\)

Replicators can copy themselves indefinitely into as many identical copies as their circumstances permit. These replicators—molecules of deoxyribonucleic acid, or DNA—constitute the genetic material of almost every known organism. Each DNA molecule is constructed out of two chains of smaller molecules called nucleotides. Each individual nucleotide consists of one sugar molecule and one phosphate molecule bound to one of the four DNA bases of thymine (T), adenine (A), guanine (G), or cytosine (C). Together the two chains of repeating genetic letters wind around each other repeatedly in a clockwise direction forming a double helix. The DNA double helix of a particular organism spells out the set of instructions for constructing that organism. These instructions direct the manufacture of specific proteins by coding for particular combinations of amino acids.\(^5\) Each individual protein instruction is preceded by a “start” message and followed by a “stop” message given in the same genetic letters. A gene, in the narrowest sense, is a cistron, or the total sequence of nucleotides, from start message to stop message, that codes for all the amino acids that make up a specific protein.\(^6\)

---

3 Sober and Wilson, p. 4.
4 Ibid, p. 34-35. This is true of both sexually and asexually reproducing species. If an asexually reproducing organism like a female stick insect were to lose a leg, this change would not be passed down to her offspring. Thus, in the case of asexual reproduction, the entire genome can be the unit of replication—but the organism itself cannot. Ibid. p. 273.
5 Brooks, p. 12.
6 Dawkins, p. 28.
There is, however, no universally agreed-upon definition of a gene. Dawkins follows G.C. Williams in using this term much more broadly.\textsuperscript{7} In this broader definition, a \textit{gene} is any portion of heritable biologically encoded information that survives for enough generations to act as a unit of natural selection.\textsuperscript{8} This definition includes genes as understood in the strict sense above, but it can also include groups of cistrons, or only a segment of a single cistron.\textsuperscript{9}

There are several reasons in support of defining genes as Dawkins does. Among these reasons is the fact that \textit{crossing-over}, or the process of recombining heritable information during meiotic division (the formation of gametes in sexually reproducing organisms), pays no special attention to cistron “start” and “stop” messages, as recombination can occur anywhere on the nucleotide chain.\textsuperscript{10} The broader use of the term can also accommodate application to more than just \textit{DNA}. Recent research has confirmed that \textit{DNA} is not the only source of biologically heritable information, as such information can also be \textit{epigenetic}.\textsuperscript{11} As epigenetic systems also have the properties of

\textsuperscript{7} Ibid, p. 272.
\textsuperscript{8} Ibid, p. 28.
\textsuperscript{9} Ibid, p. 29.
\textsuperscript{10} Ibid, p. 28.
\textsuperscript{11} For example, in order to develop normally, a cell must contain structures such as basal bodies, membranes, organelles, and \textit{DNA methylation patterns}. These structures are not coded for in an organism’s \textit{DNA}, rather, they are resources required by the cell for normal operation that have their own methods of replication. However, just as in the case of \textit{DNA}, changes in these structures can cause heritable variation that appears in the descendents of a particular cell. A DNA methylation pattern, for example, is a series of additional chemical groups attached to a \textit{DNA} sequence. A methylation pattern serves to block the transcription of any cistrons that it covers, and is replicated by the methylation coping system that has been inherited from the parent egg cell. Differences in methylation play an important part in tissue differentiation during the lifetime of an organism, but can also pass between generations. Methylation patterns are often applied by parent organisms to their sperm or egg DNA. These methylation patterns serve to influence the development of their offspring. It is theorized that, in humans, methylation patterns are partly responsible for behavioural differences between male and female children. This is because adult females methylate (block transcription of) a sequence of the X-chromosome in their egg cells. Because males receive only one copy of the X-chromosome, they are unable to transcribe the cistrons in the blocked sequence. Adult males demethylate the blocked sequence of the X-chromosome in their sperm cells thereby providing their daughters with a fully transcribable version of the sequence. As a result, there
longevity, fecundity, and copying-fidelity, they can be accommodated under the broader definition of gene or replicator to which Dawkins appeals, even though—strictly speaking—they run in parallel with normal DNA inheritance, rather than as part of it.

Finally, and most importantly for my purposes here, the broader definition allows us to speak of genes “for” complex bundles of physical characteristics, and for complex types of behaviour. In many species, large clusters of replicators code for distinctive packages of physical properties that are almost always either wholly present or wholly absent in a particular individual.

For example, some species of butterflies have evolved the adaptive strategy of tasting very nasty to birds. Such species warn off predators through their distinctive shape, colour, spot-pattern, and rhythm of flight. Other species, however, have evolved the mechanism of looking like they belong to a nasty-tasting species, even though they do not, by emulating their physical characteristics in detail. In some “mimic” species, every member of that species will mimic one of two different nasty-tasting varieties. As a result, it appears that there is a gene that specifies which species an individual will mimic. Yet no one gene, in the sense of a cistron, could ever code for the variety of characteristics that effective mimicry requires. Rather, it is a cluster of replicators that have worked together so successfully that they have come to be selected for as a unit—as a gene for mimicking a particular species.¹²

Further, as the effect of any gene depends upon its environment, and the other genes in a survival machine are as much a part of an individual gene’s environment as temperature, food, predators, or companions, the cooperation cannot end here. Genes appear to be genetic products that are available to females but denied to males that may influence behaviour. Griffiths, pp. 399-400.
must work well together if any of them are to make it into the next generation. Building a successful survival machine is a complicated venture, and every gene requires the cooperation of thousands of other genes to complete the task. Genes must coalesce into individual organisms that function as adaptive units if they are to survive. 13

Dawkins, however, argues that the fact that individual genes must coalesce into organisms does not compromise his position that genes, and not organisms, are the units of natural selection. In order to explain why, Dawkins draws an analogy between building a survival machine and winning a boat race. One rower alone cannot win the race. A rower needs eight colleagues, each specialized to a particular position in the boat (e.g., cox, bow, or stroke). Dawkins asks us to imagine a scenario in which a rowing coach must put together her ideal team from an available pool of candidates. In order to make her selection, the coach randomly shuffles candidates for each position into three new teams each day, and then races the teams against each other. After some time, the coach will find that the winning boat in each race will most often contain particular individual rowers—those that are both good at rowing and can row well as part of a crew.

In this analogy the individual rowers are genes, rowing fast corresponds to building a successful survival machine, and the pool of alternative candidates corresponds to the gene pool. As far as the survival of any one organism is concerned, “all of its genes are in the same boat.” Just as rowers who work well with the others in their crew are more likely to find themselves in winning teams then those who do not, genes that cooperate well with genes they are likely to meet in successive bodies will have an advantage over those that do not. It is individual genes that are the units of

12 Dawkins, pp. 31-32.
selection, but one property necessary to successful genes is that of being able to cooperate well with others in building a vehicle for survival.\textsuperscript{14}

To the extent that genes control the behaviour of the survival machine they build, they do so not by controlling the organism’s day-to-day behaviour, but by building a fast biological “computer” (i.e., brain) into their survival machine, and programming it with general behavioural strategies.\textsuperscript{15} As a result, behaviour is under genetic control in an indirect but still very powerful sense.\textsuperscript{16} As Dawkins dramatically puts it, “genes are master programmers, and they are programming for their lives. They are judged according to the success of their programs in coping with all the hazards that life throws at their survival machines, and the judge is the ruthless court of survival.”\textsuperscript{17}

An organism’s decision about how to behave under particular conditions—to fight or not fight, to cooperate or not cooperate—should ideally be preceded by a complex, if unconscious, cost-and-benefit calculation.\textsuperscript{18} Genes influence bodies such that, by following genetically programmed strategies, they behave as if they have performed these complex calculations.\textsuperscript{19} Just as we can expect certain things of a person’s character when we know something about the conditions in which she survived and prospered, we can expect certain things of successful genes from the fact that they have survived, in some cases for millions of years, under competitive conditions. What we can expect, Dawkins argues, is that a predominant feature of successful genes will be ruthless selfishness, and that this gene selfishness will usually translate into selfish

\textsuperscript{14} Ibid, pp. 38-39.
\textsuperscript{15} Ibid, p. 52.
\textsuperscript{16} Ibid, p. 60.
\textsuperscript{17} Ibid, p. 62.
\textsuperscript{18} Ibid, p. 69.
\textsuperscript{19} Ibid, p. 97.
behaviour on the part of the organism.

An evolutionarily stable strategy, or ESS, is a strategy that cannot be bettered by an alternative strategy when most members of a population adopt it.\textsuperscript{20} Such strategies can be driven to fixation within a population by the dynamics of replication, whereas highly unstable strategies are quickly driven to extinction.\textsuperscript{21} The more successful a strategy is, the more genes for that strategy will be carried into the next generation; the less successful, the fewer. Dawkins argues that altruistic strategies cannot be selected for because they are highly unstable. Any genes for altruism that arise will be quickly driven to extinction.

To see why, consider the following simplified scenario in which dangerous ticks have infested a primate population.\textsuperscript{22} One strategy an individual could follow in this scenario is that of helpfully removing ticks from any member of the population that is infected. If every individual plays this strategy, the fitness of every individual will increase, and thus, more genes for playing this strategy will make it into the next generation.

However, where there is an opportunity, some strategy will evolve to exploit it. As a result, a group of “suckers” like these helpful tick removers is highly vulnerable to invasion by a “cheat” strategy that will gladly accept tick removal, but never reciprocate. Cheats will benefit enormously from having their ticks removed and not having to pay the price of removing ticks from others. Computer simulations show that in iterated pairings, cheat genes spread briskly through the population, and sucker genes are driven

\textsuperscript{20} Ibid, p. 69.
\textsuperscript{21} A population is at a state of fixation for a given strategy when 100% of the population uses that strategy. Skyrms, p. 55.
\textsuperscript{22} The following example is based on that found in Dawkins, pp. 184-185.
to extinction. This is because, no matter what their ratio in the population, as long as the benefit of being groomed exceeds the cost of grooming, cheats will always do better than suckers, until there are no suckers left to exploit. The cheat strategy is the ESS in this scenario, because a population consisting largely of cheats cannot easily be invaded by other strategies. A stable strategy is stable not because it is the best for all involved, but because it is immune from treachery from within.\textsuperscript{23}

Although the altruistic sucker strategy is not an ESS, a conditionally cooperative strategy such as a "grudger" can be. A grudger will groom anyone on first encounter, and those that have previously groomed them. However, if an individual cheats on them by failing to return the favour of grooming, a grudger will refuse to groom that individual in future encounters. A small number of grudgers would not be able to invade a population of cheats because they would expend a large amount of energy grooming each new individual they met, only to find that the favour was rarely returned. Accordingly, the average pay-off for grudgers would be low, and the grudger strategy would be driven to extinction. If grudgers become prevalent enough in the population, however, their chance of meeting others of their kind increases. When there are a sufficient number of individuals playing the grudger strategy in a population, the average pay-off for such individuals increases until it is higher than that of the selfish cheats. The cheat strategy, in turn, is then driven towards extinction.

However, although the fact that conditionally cooperative strategies can be stable may initially appear to bring us some of the way towards explaining how altruism can evolve, this promise is merely \textit{prima facie}. Recall that a behaviour is altruistic if the consequence of this behaviour is that an organism increases the fitness of others at the

\textsuperscript{23} Ibid, p. 72.
expense of her own fitness. The grudger strategy is selected for only when it is not altruistic. When a small number of grudgers are played against an established population of cheats, on average, their actions serve to increase the fitness of others at a cost to themselves. As a result, the grudger strategy is driven to extinction. However, when there are a sufficient number of grudgers in a population, on average, the fitness of a grudger is increased. It is only when the grudger strategy is the selfish fitness maximizing strategy that it is selected for. Accordingly, the fact that conditional cooperation can be selected for goes no distance in explaining how altruism can evolve.

On these grounds, Dawkins maintains that there can be no true altruism in nature as all biologically driven behaviour must be selfish at the gene level. That is to say, that all biologically based behavioural strategies—even those that appear altruistic—must serve to further the selfish ends of genes if they are selected for. Truly altruistic behaviour must then be the product of culture, and thus, almost certainly, a specifically human phenomenon. As Dawkins concludes, only "we have the power to turn against our creators. We, alone on earth, can rebel against the tyranny of the selfish replicators."  

2.2 Group selection and the evolution of altruism

As noted at the outset, Sober and Wilson argue that there is a place for altruism in nature, and that the evolution of altruism can be explained through group selection. Sober and Wilson define a group as "a set of individuals that influence each other's fitness with respect to a certain trait, but not the fitness of those outside the group."  

24 Ibid, p. 201.
25 Sober and Wilson, p. 92.
They argue that it is easy to show how a trait like altruism can evolve when more than one group is present, and the following necessary conditions are met: (i) the groups differ in the proportion of altruists that they contain, (ii) there is a direct, and positive, relationship between the number of altruists a group contains, and the number of offspring the group produces, and (iii) the groups are primarily isolated from one another, but the offspring of each group mixes or otherwise competes in the formulation of new groups.\textsuperscript{26}

To be sufficient for the evolution of altruism, the differential fitness of the groups must be strong enough to counter the differential fitness of the individuals within the groups.\textsuperscript{27} Sober and Wilson argue that individual selection favours traits that maximize relative fitness \textit{within} groups (i.e., the fitness of individuals in a group relative to the other individuals in that group), whereas group selection favours traits that maximize the relative fitness \textit{of} groups (i.e., the fitness of groups relative to the other groups in a population).\textsuperscript{28} As long as the process of group selection is sufficiently strong, altruism can evolve. For, although altruism will decline in frequency in each individual group, it can evolve nevertheless because groups that contain altruists are more fit than groups that do not.

Consider the following simplified example.\textsuperscript{29} Imagine that there are two groups, group A and group B. Group A contains 20 altruists and 80 selfish types. Group B contains 80 altruists and 20 selfish types. Let us assume that each individual has 10

\begin{flushright}
\textsuperscript{26} Ibid, p. 26.
\textsuperscript{27} Ibid, p. 26.
\textsuperscript{28} Ibid, p. 27.
\textsuperscript{29} This example has been simplified in the following ways: it assumes asexual reproduction, that offspring exactly resemble their parents, and that there is a one-to-one relationship between genes and behaviour. Although Sober and Wilson note that these assumptions are unrealistic, they have been made in order to keep the example simple and illustrative. Sober and Wilson, pp. 21-22.
\end{flushright}
offspring in the absence of altruism, and that every altruist bestows an additional five offspring on a single recipient at a cost of one offspring to herself. The altruists in each group will then average nine offspring each, but will have the opportunity to benefit from one of the other altruists in their group (not including themselves). Thus, their chance of benefiting is 19/99 in group A, and 79/99 in group B. The selfish types in each group will average a minimum of ten offspring each as well as having the opportunity of benefiting from any one of the altruists. Thus, their chance of benefiting is 20/99 in group A and 80/99 in group B. The average altruist in group A will then have 9.96 offspring\(^{30}\) and selfish type 11.01 offspring (for a total of 1080). The average altruist in group B will have 12.99 offspring and selfish type 14.04 (for a total of 1319.8).

Two important things become clear: the selfish types in each group do better, but the more altruists there are in a group, the more offspring that group produces as a whole. So, although individual-level selection will select for selfishness, group-level selection will select for altruism. Altruism will evolve as long as the selection pressures between groups are sufficiently strong to counter the forces of individual selection, the progeny of the two groups disperse and then physically come together before forming new groups of their own, and altruists and non-altruists come to be concentrated in different groups as a result.\(^{31}\)

As Sober and Wilson note, the idea of group selection has long been controversial, and has often been rejected as an important evolutionary force. This, they argue, is the result of extensive commitment of the *averaging fallacy*. As just seen, on

\(^{30}\) To calculate these averages, we take the probability of being the recipient of altruism (e.g., 19/99 = 0.191), and multiply this value by the benefit altruism provides (e.g., 0.191x5=0.96). We then add this value to the minimum number of expected offspring (e.g., 0.96 + 9 = 9.96):
Sober and Wilson's model, although altruists decline in frequency in each group, altruism still evolves because groups that contain altruists are more fit than those that do not. Another way to calculate whether or not altruism will evolve is to average the fitness of individuals across groups. On the averaging model, if the 20 altruists in group A average 9.96 offspring each and the 80 altruists in group B average 12.99 offspring each, altruists in general average 12.38 offspring each. If the 80 selfish types in group A average 11.01 offspring each and the 20 selfish types in group B average 14.04 offspring each, selfish types in general average 11.62 offspring each. As the average altruist in the overall population is fitter than the average selfish type, altruism will evolve. However, as this strategy is selected for because it is the selfish fitness maximizing strategy, it is no longer genuinely altruistic.

Sober and Wilson contend that, although this method of calculating fitness does not change any facts about the original example, it is fallacious for two reasons. Firstly, the averaging model encompasses everything that evolves, regardless of the type of selection process involved. In doing so, it makes individual selection synonymous with natural selection. The fact that other groups exist is "folded into the definition of individual selection, defining group selection out of existence." Secondly, and because of this, the averaging approach fails to identify the separate causal forces at work in evolution. When altruism evolves, it is because the forces of group selection have been strong enough to overcome the opposing forces of individual selection. Certainly, any satisfactory causal analysis should be able to describe what has occurred.

31 There are several ways in which altruists and non-altruists can come to be concentrated into different groups. I will look at this issue at some length in section three.
32 Sober and Wilson, p. 32.
33 Ibid, p. 32.
Problematically, though, the claim that the trait that evolves is that which has the highest average fitness cannot specify the causal factors at work. When fitness is averaged across groups as in the example above, the altruistic strategy must become selfish after all because it is the strategy with the highest average fitness. The fact that the “altruistic” strategy really is altruistic within each group is lost altogether.34

We can now begin to see how Sober and Wilson’s multilevel approach can answer the arguments given by Dawkins. Sober and Wilson note that Dawkins developed his selfish gene theory as a method of explaining the evolution of apparent altruism without resorting to group selection.35 The selfish gene model has often been thought to provide a decisive argument against group selection. However, as Sober and Wilson show, the selfish gene model suffers from a flaw that is obvious once it has been pointed out: the whole idea of genes as replicators is totally irrelevant to group selection.36

Recall that, in order to explain adaptation at the level of individual organisms, Dawkins introduced the concept of organisms as “vehicles of selection.” Genes in sexually reproducing individuals, Dawkins argued, are like the members of a rowing crew competing against other crews in a race. Unless each crew member rows well with the others in her crew, her team will lose the race. In the same way, genes are “trapped” together in their vehicle of selection, and can replicate only if the entire gene collective survives and reproduces. Because genes share their fate, they coalesce into individual organisms that function as adaptive units. The vehicle concept allows Dawkins’s perspective to accommodate adaptation at the individual level.

34 Ibid, p. 32.
The vehicle concept, Sober and Wilson argue, opens the door to the possibility that groups can function as adaptive units as well. Just as individuals function as adaptive units because their genes are in the same boat with respect to fitness, groups can also function as adaptive units when the individuals in them find themselves in the same boat with respect to fitness. The concept of genes as replicators then, is not relevant to group selection:

Selfish gene theory does not invoke any processes that are different from the ones described in multi-level selection theory, but merely looks at the same processes in a different way. Selfish gene theory calls the genes responsible for [altruistic] behaviours “selfish” for the simple reason that they have evolved. Multi-level selection theory, on the other hand is devoted to showing how these behaviours evolve. Fitness differences must exist somewhere on the biological hierarchy—between individuals within groups, between groups in the global population, and so on. Selfish gene theory can’t even begin to explore these questions on the basis of the replicator concept alone. The vehicle concept is its way of groping toward the very issues that multilevel selection theory was developed to explain. 37

That genes are the units of replication is not in dispute, and the standard definition of natural selection as a “change in gene frequency” is not under debate. In the end it really is genes that are passed on to the next generation. Selfish gene theory places special emphasis on this fact. However, the fact that it is genes that are the units of replication on its own says nothing about the causal forces at work in natural selection. Further, the selfish gene theory determines what will evolve by averaging the fitness of genes across individuals in the population. In doing so, it commits the averaging fallacy. Genes that evolve will always have a higher average fitness then the genes that don’t evolve. It is for this reason that successful genes are always selfish in Dawkins’s view. However, looking at genes alone, and averaging the fitness of genes across individuals, ignores the causal forces at work in natural selection. As a result, the selfish gene theory

cannot find genuine altruism in nature as it looks only at what evolves and never how it evolved.\textsuperscript{38}

4.3 The evolution of altruism in humans

As noted at the outset, if there is no such thing as biological altruism, it would appear that there is a chasm between human morality and social behaviour found elsewhere in the animal kingdom. We have seen however that genuine altruism can evolve by group selection under appropriate conditions. Yet, a significant difficulty remains. Although Sober and Wilson's arguments are theoretically compelling, it is not immediately obvious how they apply to the evolution of altruism in humans. It is easy to envision how altruism can be selected for at the group level in the case of ultrasocial insects such as bees. As each member of a social insect colony is a very close genetic relative of the other members of her colony, the colony functions like a superorganism.

However, the genetic structure of human populations (and those of other primates) is very different from that found in other ultrasocial species.\textsuperscript{39} In contrast to these other species, most human groups consist of a mix of both non-relatives and genetic relatives of varying degrees. Using modern hunter-gatherer societies as a guide, it appears that the average genealogical relatedness among group members was likely above zero for most of human evolution. However, the level of relatedness has never come even close to that of social insects. As a result, some theorists have concluded that we cannot expect group selection to have been a significant force in human evolution.

Sober and Wilson argue that the problem with such reasoning is that it treats

\textsuperscript{37} Ibid, p. 88.
\textsuperscript{38} Ibid, pp. 88-89.
genealogical relatedness as the only important variable in the evolution of group-level functional organization. What matters is that altruists settle with altruists. What does not matter is why they settle with other altruists. Sober and Wilson maintain that there are other mechanisms that can substitute for genealogical relatedness in the evolution of group-level organization. Humans and social insects are both highly group-selected, but for different reasons.\textsuperscript{40}

One mechanism that can substitute for genealogical relatedness is the ability to choose associates. If individual group members can have information about others, and have some control over group membership, this will create highly non-random groupings. As Sober and Wilson note, the idea that choosing associates favours the evolution of altruism is deeply intuitive. However, it has long been neglected because initial efforts to model assortative interactions encountered the \textit{problem of origination}. Altruism can evolve after discriminating altruists exceed a threshold frequency in a population. However, a model that requires discriminating altruists to already exist in a population at a frequency of 20 percent fails to address a fundamental problem about how altruism can evolve. This problem does not occur in the case of genealogical relatedness, because an initial mutant altruist that survives long enough to mate even with a nonaltruist will still produce altruistic offspring at a frequency of 50 percent.\textsuperscript{41}

The problem of origination is however not as serious as it may initially appear. Although it is convenient to model behaviours as discrete traits that enter a population at a mutation frequency, real behaviours are often continuous and have both a mean and variance. It is difficult to account for the evolution of altruism when altruists exist at a

\textsuperscript{39} Ibid, pp. 133.
\textsuperscript{40} Ibid, p. 134.
mutation frequency in a population of nonaltruists. However, this problem largely disappears when individuals vary in a continuous fashion. It may be "hard for a mutant altruist to find another altruist to interact with, but it is easy for an individual to who is above average to find another individual who is above average."\[^{42}\] This can be true, Sober and Wilson argue, even if the average degree of altruism in a population is very low.\[^{43}\]

Two further difficulties remain however, as not all groups are formed by individual choice, and it may often be difficult to detect the altruistic tendencies of others (especially if individuals have evolved to conceal their selfish tendencies). Although groups are not always formed by choice, this difficulty is less problematic if groups are able to expel or otherwise punish nonaltruistic members. Further, as long as group members have a certain minimal level of cognitive sophistication, they should be equipped to discern the altruistic tendencies of others.\[^{44}\] Humans have evolved powerful abilities to acquire information about others based on personal interaction, direct observations, and cultural transmission. This information can be used to seek out trustworthy individuals and avoid cheaters in social interactions.\[^{45}\]

It is certainly true that the ability to detect cheaters is likely to select for the ability of cheaters to avoid detection. However, such deception is unlikely to be completely successful. Even if a cheater is successful in initial interactions, as long as her history of past interactions can be known, her behavioural tendencies can be reliably discerned from

\[^{41}\] Ibid, p. 136.  
\[^{42}\] Ibid, pp. 136-137.  
\[^{43}\] Ibid, pp. 136-137.  
\[^{44}\] Ibid, p. 140. Sober and Wilson argue that evidence suggests that even with their limited cognitive abilities, guppies are able to accomplish nonrandom groupings. Ibid, p. 141.  
\[^{45}\] Ibid, p. 141.
prior behaviour. It is difficult to fake being an altruist or conceal selfishness in human social groups whose members have a long history of interaction. As Sober and Wilson note, the large anonymous societies of today are recent inventions. Human evolution took place in small groups where members had ample opportunity to observe and talk about each other. Information about a single antisocial act would have spread quickly throughout the group, soiling a person's reputation, and jeopardizing her future social interactions. Accordingly, Sober and Wilson conclude that human social interactions among unrelated individuals are anything but random. Our ability to learn, and to change our behaviour according to what we learn, provides a powerful mechanism for the evolution of altruism and other group selected behaviours.46

46 Ibid, pp. 141-142.
In the first chapter, I argued in support of the claim that, rather than being different in kind, human morality is on a continuum with behaviours observable in some non-human primates. Flack and de Waal argue that recent behavioural research and observations of non-human primates provide confirming evidence for the hypothesis that the elements of moral systems are present in the methods that some non-human primates use for resolving and preventing conflicts of interest within their groups. These elements or building blocks are found in methods similar to those humans use, and include food sharing, reconciliation, consolation, conflict intervention, protective intervention, and mediation. These methods appear to require or make use of such key moral components as a capacity for empathy, sympathy, community concern, and a sense of social regularity.¹

I have argued in chapter one that the most parsimonious explanation from an evolutionary perspective is that, when the conflict management behaviours Flack and de Waal describe occur under functionally similar circumstances, they are produced through similar cognitive and emotional channels in both human and non-human primates. Accordingly, as human methods of conflict management make use of such traits as sympathy, empathy, community concern, and a sense of social regularity, the best explanation is that these play a part when the same behaviours are observed in non-human primates.

However, it remains unclear exactly how the building blocks Flack and de Waal
identify provide raw materials from which moral systems can be constructed. In order to satisfactorily provide evidence of the biological foundations of morality, it is essential that the relationship between these building blocks and morality be elucidated. An acceptable working account of both what morality is, and the way in which the building blocks described can provide its components must be provided. Flack and de Waal define morality as “a sense of right and wrong, that is born out of group-wide systems of conflict management based on shared values.” They argue that components such as empathy, sympathy, community concern, and a sense of social regularity provide the building blocks of morality because they help to generate connections between individuals despite inevitable conflicts of interest.

However, the definition of morality Flack and de Waal appeal to is problematic for two reasons—it is too broad, and it renders their argument circular. It is too broad because it does not specify what demarcates a sense of right and wrong as particularly moral as opposed to belonging to some other normative category. Certainly the caveat that moral systems arise out of conflict-management behaviours based on shared values does little to specify them as particularly moral systems. Indeed, one could imagine amoral or even immoral systems arising in such a fashion. Further, building in an origin from conflict-management behaviours renders Flack and de Waal’s argument circular. Flack and de Waal discuss a variety of conflict-management behaviours that they argue evidence the building blocks of morality. However, defining morality as a sense of right and wrong that arises from conflict-management behaviours necessitates that conflict-

1 Flack and de Waal (2000a), p. 3.
2 Flack and de Waal, of course, do not commit themselves to the claim that these elements provide the sole basis from which morality springs. Rather, these elements play a role in the development of morality. I also take this view.
management behaviours contain the building blocks of morality. Thus, conflict-management behaviours evidence moral building blocks simply because they are conflict-management behaviours. As a result, we are left with no satisfactory account of why elements such as empathy, sympathy, community concern, and a sense of social regularity are moral building blocks, and how they provide elements for the construction of moral systems.

Although Flack and de Waal’s account is problematic, it is the purpose of this chapter to argue that a satisfactory account of the relationship between morality and the building blocks that Flack and de Waal discuss can be given. In giving this account, I will introduce an alternative conception of morality, and then show how the building blocks evidenced in non-human primates provide its raw materials.

3.1 Morality and the demarcation problem

When examining its potential biological foundations, failure to start with a suitable working conception of morality certainly raises one risk of making “spurious claims to ethical relevance.” To start, however, with a specific and sophisticated account such as Kantianism, utilitarianism, virtue theory, or social contract theory, and then look for evidence of the categorical imperative, the principle of utility, excellence of character, or either explicit or tacit contractual agreement in non-human primates would be to set the bar too high. Such accounts presuppose cognitive abilities (e.g., advanced forms of rationality), that are not attributable to non-human animals. Further, the principle import

---

4 Flack and de Waal, (2000a), p. 3.
5 D’Arms, p. 297.
of moral theories is prescriptive rather than descriptive. Such theories serve not so much
to tell us what morality is, but rather what it is moral to do. As such, they do not provide
an account of what makes a moral rule or judgement specifically moral, rather than a
member of some other normative category.\(^7\) On these grounds, accounts such as
Kantianism, utilitarianism, virtue theory, and social contract theory do not provide
appropriate starting points for investigating the potential biological foundations of
morality.

As Catherine Wilson observes, “it is difficult to specify the topic of moral rules,
[and] what they seek to regulate, in a way that is noncommittal between moral theories
and that does not import prescriptive considerations into what ought to be a descriptive
task.”\(^8\) As Wilson notes, few attempts to provide a truly descriptive account of morality
have been made.\(^9\) Justin D’Arms has recently proposed adopting a working conception of
morality as “a system of requirements and prohibitions enforced by sanctions (emotional
and punitive), imposed both externally by members of the community, and internally (to
varying degrees) through negative self-directed feelings.”\(^10\) However, as D’Arms does
not explore the precise nature of the requirements and prohibitions that mark a system as
moral rather than as falling into some other normative category, his proposed conception
may be useful for exploring the biological foundations of normative systems in general,
but not specifically for taking up the biological foundations of morality. In order to truly
avoid spurious claims to ethical relevance, an account of the nature of a system of

\(^6\) One would not, for example, set about looking for the origins of language in non-human primates by
specifically looking for evidence of Latin, French, or Esperanto rather than the linguistic structures or
features that comprise them.
\(^7\) Wilson, p. 2.
\(^8\) Ibid, p. 8.
\(^9\) Ibid, p. 8. Wilson discusses an attempt made by John Stuart Mill, but finds it lacking both because it
imports prescriptive considerations, and it fails to cover the scope of potential moral offences.
sanctions that demarcates it as a particularly moral one is required.

Wilson maintains that moral rules can be identified by their prohibitive nature. An analysis of the current moral literature, Wilson notes, yields many examples of allegedly objective moral truths. These examples most often concern the wrongness of harming non-human animals, harming children, torturing people, or engaging in genocide. What each of these purportedly objective moral truths has in common is a concern with actions that are forbidden. Wilson argues that this prohibitory form provides a clue to the usual content of a specifically moral rule. Moral rules appear to "command a sacrifice of opportunities for gratification; they deny permission to act in a careless, impulsive, overly-partial, or indifferent way in the pursuit of one's self interest."

That there is such a high degree of confidence in the correctness of moral judgements with regard to the harming of children and non-human animals, Wilson argues, can be explained by the hypothesis that they correspond to highly-presentable samples of moral rules. Children and non-human animals are easy to harm, and incentives to harm them are easily found as they are small, relatively helpless, under our control, and often provoking. However, we believe that the mere impulse to harm children and non-human animals must be restrained at all times, and that good reasons are required when they are harmed for the benefit of themselves or others. Although a social rule that encouraged the torture of children or non-human animals could exist in some society, it would not be a moral rule. As Wilson argues, "[t]he claim 'It is

---

10 D'Arms, p. 297
11 Wilson, p. 3.
12 Ibid, p. 3.
13 Ibid, p. 3.
permissible to torture children if you gain an advantage from doing so’ would not
ordinarily be regarded as an example of a moral judgement that happens to be false.
Whether the statement is regarded as false or lacking in truth-value altogether, it does not
seem to qualify as a moral judgement at all. 14

Wilson contends that these observations regarding the prohibitory and harm-
reducing nature of moral rules suggest that they can be characterized as rules that
“restrict, restrain, limit, or reverse the short or long-term advantage possessed by a
naturally- or situationally- favoured subject vis-a-vis another.” 15 On Wilson’s view,
moral rules are picked out by their concern with “the regulation of actions that tend to
secure or advance the aims of one agent or group at the expense of another agent or
group’s well-being.” 16 This feature serves to distinguish moral rules from the rules of
other normative systems such as appearance-regulating aesthetic rules; health-, wealth-, and
reputation-maintaining prudential rules; and the decorum-preserving rules of
etiquette. 17

Moral rules place limits on the pursuit of self-interest, or the interests of one’s
group, by restricting the physical and psychological injury that individuals can do to one
another. 18 Wilson argues that harms that result “from negligence and indifference, as
well as harms resulting from the desire to exploit or injure can also be understood as the
effects of self-interest in this sense.” 19 At bottom, moral rules appear to regulate
aggression and the approbation of goods or opportunities. In doing so, such rules protect
the physically weaker members of the group against the strong and agile. As Wilson

14 Ibid, p. 3.
15 Ibid, p. 4.
16 Ibid, p. 4.
17 Ibid, p. 4.
argues:

The existence of duties considered to form the core elements of morality, to avoid interfering with others and their goods, to refrain from exercising lethal force, tell the truth, keep promises, and perform contracts, presupposes that naturally or situationally favoured subjects can gain temporary or long-term advantages over others from interference, theft, violence, deception, and bad faith.\(^{20}\)

On this view, “Don’t steal” is a moral rule that serves to prevent those that are light-fingered from taking advantage of those who are inattentive, despite the fact that theft can offer a way to gain valuable resources; “Keep your promises” is a moral rule that aims to prevent stronger parties from breaking their word even when they could benefit very much from doing so.\(^{21}\) Moral rules function in the social economy as a brake “not just on our emotions or inclinations, where the latter are viewed as non-rational velleities, but—to some extent—on our intelligence, competence, and social forcefulness.”\(^{22}\)

Wilson’s descriptive account of morality is a semi-essentialist one. That is to say, it is an account with fuzzy boundaries. Some moral rules are better exemplars of their normative category, and there are rules that will straddle other normative categories. This is a strength, given that it does seem that some moral rules are more clearly moral in content than others, and some moral rules function dually as rules of prudence or etiquette. Wilson’s view has much to recommend it, but my purpose here is not to defend it at length. Wilson’s descriptive account meets the requirement of being both noncommittal between moral theories, and devoid of prescriptive considerations. Further, it does not rule out the possibility that elements of morality may be present in non-human animals. On these grounds I maintain that, among its other strengths,
Wilson's account can serve as a very adequate working conception of morality for the purpose of examining its biological origins.  

3.2 Social regularity and prescriptive rules

As noted at the outset, in order for the conflict-management behaviours that Flack and de Waal discuss to provide good evidence of the biological origins of morality, a satisfactory account both of what morality is and the way in which the building blocks they discuss provide its components is required. Wilson's descriptive account of morality as a system of rules aimed at the regulation of natural and situational advantages fulfills the first of these conditions. However, it remains unclear exactly how the components Flack and de Waal discuss can appropriately be construed as providing building blocks from which such a system can be constructed.

Flack and de Waal provide evidence that behaviours such as food-sharing, reconciliation, consolation, third-party conflict intervention, protective intervention, and mediation are present in some species of non-human primates, and they argue that these methods of conflict intervention appear to require, or make use of, such key moral components as a capacity for empathy, sympathy, community concern, and a sense of social regularity. What is required to complete Flack and de Waal's account, then, is an explanation of how these elements provide moral building blocks. Given our working definition of morality, we must see how these capacities provide building blocks for the construction of a system of rules that regulate natural or situational advantages. In what follows in this section, I will look at the primate sense of social regularity, and evidence

23 Wilson develops her account of morality as advantage-reducing imperatives in order to shed light on the problem of exigency, and provide a framework for discussing some central problems of distributive justice.
of prescriptive rules in primate societies. In the final section, I will look at sympathy, empathy, and community concern, and their role in moral judgement and action.

Flack and de Waal contend that a sense of social regularity is evident in non-human primate species, and that this sense provides a key component for the construction of moral systems. In “The Chimpanzee’s Sense of Social Regularity and Its Relation to the Human Sense of Justice” de Waal elaborates on the kinds of primate behaviour that appear to reflect a sense of social regularity, or how others should and should not behave. De Waal argues that prescriptive rules are present in non-human primate societies, and that these rules provide evidence of social expectations. De Waal defines a sense of social regularity as “a set of expectations about the way in which oneself (or others) should be treated and how resources should be divided, a deviation from which expectations to one’s (or the other’s) disadvantage evokes a negative reaction, most commonly protest in subordinate individuals and punishment in dominant individuals.”

This sense of social regularity is generally egocentric. However, because the interests of kin and even other individuals close to the actor are also sometimes taken into account, de Waal parenthetically includes others in its definition.

All animals conform to social rules in that their behaviour towards conspecifics is somewhat predictable. As de Waal notes, the complexity of these rules is related to the learning abilities of a given species. Insects and fish exhibit a limited set of interaction patterns, whereas mammals exhibit a wide range of social behaviour that is, although

---

24 The type of expectations are not specified because of differences in dominance “styles” between primate species. Some species are so hierarchical that dominant individuals claim all resources for themselves. Other species are more egalitarian, and subordinates are often able to obtain a share of whatever dominant individuals possess. Such dissimilarities would effect the kinds of expectations found in different species. de Waal (1991), p. 336.
regular, highly adaptable to new circumstances. In the case of maternal protection, for example, females with young either respond with withdrawal or aggression to conspecifics that pose a threat to their offspring. They may do so in different ways, or to different degrees, but protection of offspring is a general rule.

However, rules that pick out how an organism will typically respond to a specific situation are merely descriptive rules about behaviour. What is of interest are not rules that members of a species merely conform to, but rather rules that they have learned to follow because of active reinforcement by others. De Waal maintains that prescriptive rules are born when individuals learn to recognize the contingencies between their own behaviour and the behaviour of others, and to act in ways that minimize negative consequences. Returning to the example of maternal protection, one can see how a mother’s protective response could have an effect on the way in which infants are treated. Any individual that deviates from a mother’s norms regarding the treatment of her young will meet with a negative reaction. It has been observed that juvenile monkeys that have gained access to an infant for the first time will almost continuously throw glances at the infant’s mother, as the slightest sign of distress will mean the end of play contact. They recognize the contingencies between the mother’s behaviour and their own, and behave accordingly.

De Waal maintains that, in non-human primate societies, there are a large number of prescriptive rules like those governing the treatment of infants. There are, for example, rules governing the initiation of sexual contact in the presence of dominant

26 Ibid, p. 337.
27 Ibid, p. 337.
males, and rules governing friendly contact between the allies and rivals of the alpha male. The circumspect manner in which social rules are violated provides possibly the clearest indication that non-human primates are aware of them. Female chimpanzees, for example, have been observed to use surreptitious ways of arranging mating sessions with subordinate males by setting up “rendezvous” out of the sight of dominant males. Allies and rivals of a troop’s alpha male are normally threatened by him whenever they attempt to initiate friendly contact with one another. As a result, these males have learned to initiate contact only when the alpha male is out of sight or asleep. They may even take turns watching an alpha male sleep, and break off contact as soon as he opens his eyes.

Almost by definition, it is the dominant members of a group that most often put limits on the behaviour of the other group members. Dominants place restrictions on the mating opportunities of subordinates, forbid contact between their allies and rivals, and place a myriad of other controls on behaviour. However, de Waal notes that the handing down of social rules by dominants is not the only way that such rules arise. Social rules can also be established “from below.”

One example of how rules can arise from below is through a system of delayed retaliation, or revenge. Evidence suggests that chimpanzees are able to keep track of their social interactions, including those who have attacked or threatened them, and will often pay them back in kind. Although the revenge system encompasses all levels of social hierarchy, de Waal notes that this system has particular advantages for

---

33 Ibid, p. 338.
subordinates. Instead of risking injury by retaliating directly against powerful dominant aggressors, subordinate chimpanzees can follow the safer strategy of waiting for the opportunity to intervene against dominants when they are involved in conflicts with others. Dominants who behave negatively towards a large percentage of subordinates will then lose their position as a result of the coalitions that can be formed against them. Accordingly, de Waal holds that delayed retaliation is an effective way of putting pressure on dominants, and serves to establish rules of conduct within the community that respect the security of subordinates and provide them with access to resources.35

Thus, evidence is highly suggestive that a sense of how oneself and others are to behave is present in some non-human primates. It appears that non-human primate societies have prescriptive social rules that have developed because of active reinforcement by others. Many of these rules are simply imposed by dominants. However, as just discussed, mechanisms exist in chimpanzee societies that allow for the establishment of rules by subordinate members. As a result, chimpanzee societies are fairly egalitarian, and subordinates are often able to obtain a share of whatever dominant individuals possess.36 Although most rules protect the special privileges of dominant members of the community, there is evidence that rules are in place that protect the interests of subordinate members from those who are larger and more socially powerful. In a sense, then, there are rules in chimpanzee societies that appear to regulate advantage and the expression of dominance.

Despite the presence of such rules, one cannot jump to the conclusion that chimpanzee societies are moral ones. The majority of rules in chimpanzee societies

protect rather than restrict the special privileges of dominants. Further, the fact that dominants do respect the interests of subordinates with regard to certain resources and opportunities does not spring from any particular regard for, or cognitive awareness of rules \textit{qua} rules. Rather, it comes from an ability to recognize that disregarding the interests of subordinates too often meets with negative consequences. As de Waal observes, it is also not clear to what extent the distinction between accepted and unaccepted behaviour corresponds to any conscious awareness of "good" and "bad."\footnote{Ibid, p. 336.} A non-human primate likely does not come to the aid of a troop mate because she thinks she "ought to" or because it is "the right thing to do" nor does she avoid stealing the food of another because she thinks she "ought not to" or because it is "the wrong thing to do."

What is missing from the chimpanzee orientation towards the social world, Wilson argues, is "an interest in regulation as such."\footnote{Ibid, p. 347.} We cannot say of non-human primates that they "regard the whole field of social interactions as susceptible [to] molding and determination by themselves as agents."\footnote{Wilson, p. 3. Wilson doesn't make this claim specifically about chimpanzees, but rather about all non-human animals.} As de Waal notes, chimpanzee norms appear to operate almost exclusively at the level of private relationships, and arise only as a function of the contingencies between behaviours and consequences.\footnote{Wilson, p. 3. Wilson doesn't make this claim specifically about chimpanzees, but rather about all non-human animals.} Although one cannot conclude that chimpanzee societies are fully moral ones, a sense of social regularity clearly provides an important moral building block. In it we can see the beginning of prescriptive rules, and even specifically of rules that limit the expression of dominance, and thus regulate advantage.
3.3 Emotional building blocks

We have now seen how prescriptive rules arise when individuals learn to recognize the contingencies between their own behaviour and the behaviour of others, and to act in ways that minimize negative consequences. Further, we have seen how the primate sense of social regularity, and the prescriptive rules that inform it, can provide an important component in the development of morality.

As noted above, Flack and de Waal contend that emotional capacities such as those for empathy, sympathy, and community concern also provide important moral building blocks. In his recent paper "The Moral Emotions" moral psychologist Jonathan Haidt has compiled a natural history of the emotions comprising the moral sense. All emotions are responses to perceived changes, threats, or opportunities. However, the moral emotions are special in that they are "linked to the interests or welfare either of society as a whole or at least of persons other than the judge or agent."41 Moral emotions take us beyond the direct interests of the self.

Haidt argues that the moral emotions have two prototypical features: they can be triggered by disinterested elicitors, and they have pro-social action tendencies.42 That is to say, that the moral emotions can be triggered by events that do not directly affect the self. Further, emotions generally motivate some sort of action in response to the eliciting event. As Haidt notes, the action is often not taken, but the emotion puts the individual "into a motivational and cognitive state in which there is an increased tendency to engage

39 Ibid, p. 3.
42 Ibid, p. 3. Moral emotionhood is a matter of degree. Any emotion is a moral emotion to the extent that it has disinterested elicitors, and pro-social action tendencies. Ibid, p. 16.
in certain goal-related actions (e.g., revenge, affiliation, comforting, etc.). In the case of moral emotions, these action tendencies are pro-social, in that they either advance the interests of others, or uphold or benefit social order.

Haidt has found that a number of emotions meet these two conditions to varying degrees and classifies these moral emotions into four families: the other-condemning emotions such as anger, disgust, and contempt; the self-conscious emotions such as shame, embarrassment, and guilt; the other-suffering emotions such as sympathy, compassion, and empathy; and the other-praising emotions such as gratitude and moral awe. Haidt maintains that each of these families has a role to play in the functioning of morality. The other-condemning emotions motivate one to change relationships with those who do not respect the interests of others, violate their duties, or treat others unfairly. Anger in particular has been found to motivate one to repair the social order, and make violators change their ways. The self-conscious emotions help one to navigate the complexities of fitting into groups. These emotions motivate a regard for the interests of others so as to avoid triggering their other-condemning emotions. The other-suffering emotions motivate one to be moved by the distress of others, and to come to their aid. The other-praising emotions motivate directly pro-social behaviour towards benefactors and can even inspire benevolence towards society as a whole.

Haidt makes the strong claim that the moral emotions are essential to normal moral functioning, as they are in command of moral judgement and action. Moral

43 Ibid, p. 3.
44 Haidt's claim that particular emotions meet these conditions, and his account of the types of pro-social behaviour they motivate, is based on an extensive survey of research into these emotions. It is not possible to recount the details of all the cited studies here.
45 Emotions that lead to ostracism, shaming, and murderous vengeance are no less a part of our moral nature than emotions that lead to "nice behaviour." Ibid, p. 5.
46 Ibid, p. 4.
reasoning, he argues, is very much a secondary process to that of moral intuition. Moral judgement “involves quick gut feelings, or affectively laden intuitions, which then trigger moral reasoning as an ex-post facto social product.” However, research in moral psychology has traditionally focused on the primacy and importance of moral reasoning. As Haidt notes, moral psychology has long been dominated by rationalist models of moral judgement. Such models assert that moral knowledge and moral judgement are reached primarily through a process of reasoning and reflection. Although moral emotions such as sympathy may serve as inputs to the reasoning process, the moral emotions are not the direct cause of moral judgements on rationalist models. Rather, on such models, one weighs issues of harm, justice, and fairness before passing moral judgement.

Lawrence Kohlberg developed his influential cognitive-developmental theory of moral development by building on the pioneering work of Jean Piaget, and originating an interviewing process designed to bring out the way in which people think about, and resolve, moral conflict. In his studies, Kohlberg presented both children and adults with dilemmas that featured moral and non-moral claims on each side of an issue, and then studied how individuals resolved the conflict. Kohlberg’s best known dilemma involves a man named Heinz who must decide whether or not to break into a pharmacy in order to

---

47 Ibid, pp. 5-16.
48 Haidt defines moral intuition as “the sudden appearance in consciousness of a moral judgement, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of search, weighing evidence, or inferring a conclusion... One sees or hears about a social event and one instantly feels approval or disapproval. Haidt (2000), p. 6.
49 Ibid, p. 18. Haidt defines a moral judgement as an evaluation (good or bad) of the actions or character of a person, that is made with regard to the set of values that are held by a group to be obligatory. Ibid, p. 6.
50 Ibid, p. 2.
51 Ibid, p. 2.
steal a drug that he cannot afford, but might save the life of his dying wife.\textsuperscript{52} Kohlberg found that the way in which people resolve such dilemmas conforms to a six-level progression of increasing cognitive sophistication. Children begin as egoists and judge actions based on a conscious assessment of the good or bad consequences that they bring to the self. However, as cognitive abilities develop, and children acquire the ability to see situations from other people’s perspectives, they resolve conflicts through a reasoned assessment of what community standards demand, and finally through a reasoned assessment of what universal moral principles demand.\textsuperscript{53}

As Haidt notes, Kohlberg trained or inspired most of the leading researchers working in moral psychology today.\textsuperscript{54} The most widely used approach at present—the social-interactionist model—falls well within the established rationalist tradition. On this model children are interviewed about rule violations of various sorts, and after providing their initial judgment, are asked to respond to a series of questions designed to determine how they think about the rule in question (e.g., Can the rule be changed? Would the action be O.K. if there were no rule about it?). Participants are also asked to provide their reasons behind the judgments they have made.\textsuperscript{55}

Based on the results of their research, social-interactionists maintain that people think specifically about the consequences of an action prior to judging whether or not it

\textsuperscript{52} Kohlberg, p. 12.

\textsuperscript{53} In more detail, Kohlberg’s model includes two stages at the pre-conventional level (where one follows rules to avoid punishment, or to gain reward), two stages at the conventional level (where what the majority deems right is right, and what preserves social order is what is right), and two stages at the post-conventional, autonomous, or principled level (where one has a commitment to internal principles of personal conscience, and concern with universal ethical principles). Ibid, pp. 17-19.

\textsuperscript{54} Haidt (2000), p. 4.

\textsuperscript{55} Ibid, p. 4.
constitutes a moral violation.\textsuperscript{56} Researchers have found that it is actions that result in injustice, harm, or the violation of rights that are recognized as falling within the moral realm, and that such actions are treated differently than other types of rule violations. They have also found that rules prohibiting moral violations are judged even by small children to be unalterable and universally applicable. Actions that do not result in injustice, harm, or the violation of rights are treated as violations of social convention, or as just matters of personal preference.\textsuperscript{57}

It is, of course, undeniable that people engage in moral reasoning. However, cognitive-developmentalists, social-interactionists, and other rationalist moral psychologists have failed to show that reasoning about rules and consequences is the cause of moral judgment rather than a result. In a recent study, Haidt and his colleagues asked participants to consider the following story:\textsuperscript{58}

Julie and Mark are brother and sister. They are travelling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep the night as a special secret, which makes them feel even closer to each other.\textsuperscript{59}

After considering the story, participants were asked: “What do you think about that? Was it O.K. for them to make love?”\textsuperscript{60} Haidt and his colleagues found that most subjects immediately remarked that it was wrong for the siblings to make love. However,

\textsuperscript{56} A significant part of the social-interactionist model is that moral development occurs in a social context through behaviours like turn-taking, sharing, harming, and responding to harm. I will not be discussing this part of the social-interactionist view here.

\textsuperscript{57} Ibid, p. 4.

\textsuperscript{58} This story is one of several utilized by Haidt and his colleagues. Each story is relevantly similar in that it includes an “offensive yet harmless” act. Other examples include eating one’s dead pet dog, or eating a chicken carcass that one has used for masturbation. See Haidt, J. S. Koller, and M. Dias (1993) “Affect, Culture and Morality, or Is it Wrong to Eat Your Dog?” in Journal of Personality and Social Psychology. Vol. 65. No. 4. pp. 613-628.

\textsuperscript{59} Haidt (2000), p. 2.
individuals then struggled to find reasons to support their judgement. They would point out the dangers of inbreeding, only to remember that two forms of birth control were used. Participants argued that Julie and Mark would be emotionally damaged, only to remember that no harm resulted. Haidt and his colleagues found that eventually subjects would admit something like: “I don’t know, I can’t explain it, I just know it’s wrong.”

61 These results suggest that, contrary to rationalist models, people do judge acts to be wrong without knowing why they condemn them. A model of moral judgement, Haidt argues, must allow for this phenomenon of “moral dumfounding.” Rationalist models cannot accommodate a person’s knowing that something is wrong without her knowing why it is wrong, as conscious reflection about rules and consequences is supposed to provide the very grounds for moral judgement.

There are a variety of reasons, Haidt argues, to doubt the causal role of reflective, conscious reasoning in moral judgement. Further, there is a lot of evidence to support the claim that it is the moral emotions that are in command of moral judgement. It is now widely accepted in social and cognitive psychology that there are often two processing systems at work when a person makes judgements or solves problems: reason and intuition. These two systems typically run in parallel, but are capable of reaching differing conclusions.63 It has been demonstrated, however, that intuition has primacy over reason, as it is the system that is triggered more quickly in real-time judgments. It

60 Ibid, p. 2.
61 Ibid, p. 2.
62 This term refers to the stuttering, laughing, and expressions of surprise exhibited by participants at their inability to find supporting reasons for their judgement. Inability to find supporting reasons, however, did not lead participants to alter initial judgements of condemnation. Ibid, p. 2.
has also been shown that when intuition and reason yield conflicting judgments, it is the
judgements of the former that are stronger and less likely to be revoked.\textsuperscript{64}

As Haidt notes, it is the emerging view in social cognition that most of our
behaviours and judgments are made automatically through intuition, without intention,
effort, or awareness of process.\textsuperscript{65} Evidence suggests that moral judgments work like most
other types of judgements, and arise primarily out of an intuitive process. Current
research on attitude formation, addressing the central question “How do people form
attitudes about other people?” indicates that attitude formation is better described as a set
of automatic processes than as a process of deliberation and reflection about a person’s
traits. It has been found that people form impressions at first sight, and that impressions
formed from observing as little as five seconds of behaviour are almost identical to those
formed from long and leisurely observation and deliberation.\textsuperscript{66} These first impressions
have also been found to alter subsequent evaluations, as positive evaluations of non-
moral traits such as attractiveness lead to beliefs about corresponding moral traits such as
kindness and good character.\textsuperscript{67} It has also been found that people categorize others
automatically, applying stereotypes that include morally evaluated traits such as
aggressiveness in the case of particular ethnic groups.\textsuperscript{68} These findings, Haidt contends,

\textsuperscript{64} Ibid, p. 8. It is of course possible that a person could just reason his way to a moral judgement, even if
this judgement contradicts an initial intuition. However, literature on everyday reasoning suggests that
“such an ability may only be common among philosophers, who have been extensively trained and
socialized to follow reasoning even to very disturbing conclusions (as in the case of Socrates, or the more
recent works of Derek Parfit and Peter Singer).” Ibid, p. 22.
\textsuperscript{65} Ibid, p. 8.
\textsuperscript{66} Ibid, p. 8.
\textsuperscript{67} Ibid, p. 8-9.
\textsuperscript{68} Ibid, p. 9.
illustrate how perception of a person or event leads instantaneously to moral judgement without conscious reflection or reasoning.\textsuperscript{69}

The automatic nature of moral judgement, Haidt argues, is also demonstrated by current research into persuasion. According to the recent heuristic-systematic model of persuasion, people are guided in part by the “principle of least effort.”\textsuperscript{70} People have limited cognitive resources, and heuristic processing (e.g., I agree with people that I like) is easy and adequate for most tasks. Research suggest that this intuitive process is generally used unless there is a special need for systematic processing.\textsuperscript{71} The fact that one’s friend has made a negative moral judgement about a person will generally affect one’s own intuitions directly. However, when the use of a heuristic leads to conflict (e.g., the person that has been negatively assessed is also a friend), effortful systematic processing may then be triggered. When no such conflict is present, feeling a strong identification with the source of a persuasive message leads to a shift in attitude. This suggests that the intuitive process is the default process, making everyday moral judgements in a quick, easy, and holistic manner.\textsuperscript{72}

Haidt maintains that standard moral judgement interviews like those used within the cognitive-developmental and social-interactionist models may create an unnaturally reasoned form of moral judgement. In an interview setting, participants are questioned by a stranger (the research psychologist) who challenges their judgement at every turn. This forces the participant to engage in extensive and effortful central processing that would normally not be activated. Such artificial judgements then lead researchers to the

\textsuperscript{69} Ibid, p. 9.
\textsuperscript{70} Ibid, p. 9.
\textsuperscript{71} Ibid, p. 9.
\textsuperscript{72} Ibid, p. 9.
erroneous conclusion that moral judgments are primarily the result of a reasoning process, when in fact this process occurs after a judgement has been made in order to defend it against the scrutiny of others. 73

Although the discussion thus far has been focused on the role of the emotions in moral judgement, Haidt maintains that the debate between rationalism and intuitionism can also be carried out using moral action as a dependent variable on moral judgement. Haidt notes that much work has been done to examine the relationship between moral reasoning and moral action, and between the moral emotions and moral action. Studies on psychopaths—intelligent and rational individuals for whom reason has become dissociated from the moral emotions—suggest that moral action is much more closely linked to the moral emotions than it is to moral reasoning. Psychopaths know the rules of social behaviour, and understand the harmful consequences that their actions can have for others. However, they do not care about these consequences. Such individuals fail to display normal affective reactions, particularly those that are triggered by the suffering of others (e.g., remorse and sympathy), condemnation by others (e.g., shame and embarrassment), and attachment to others (e.g., love and grief). There are well-documented cases of psychopaths dismembering live animals, stealing from people who care about them, and even murdering their own parents to collect insurance money—all without any remorse or shame when they are caught. 74

Haidt notes that several lines of research are converging on the conclusion that psychopaths differ from normal people in the functioning of their pre-frontal cortex—an area of the brain associated with affective processing. The role of the pre-frontal cortex

73 Ibid, p. 9.
74 Ibid, p. 15.
in moral behaviour has been extensively explored, and it has been found that there is a consistent pattern of change associated with damage to the ventro-medial region of this area. Patients with damage restricted to the ventro-medial region of the pre-frontal cortex (VMPFC) show no reduction in reasoning abilities. Such individuals retain full knowledge of moral rules and social conventions, display normal abilities to solve logic problems, financial problems, and even hypothetical moral dilemmas. However, when faced with real moral problems, those with VMPFC damage perform disastrously, showing poor judgement, indecisiveness, and what appears to be irrational behaviour.\textsuperscript{75}

It has been shown that the central deficiency resulting from the destruction of the VMPFC is that of emotional responsiveness to the world in general, and to one's behavioural choices in particular. Exposure to images that arouse strong skin conductance responses in normal people (e.g., nudity, mutilation, and death) produce no response in patients with VMPFC damage. Although these patients know that they should be affected by the images, they report no emotional reaction.\textsuperscript{76} Accordingly, damage to the VMPFC appears to produce results that mirror the lack of autonomic responsiveness observed in psychopaths.\textsuperscript{77}

Moral emotions such as sympathy, empathy, guilt, gratitude, and the moral sense that they comprise, serve to link us to the interests or welfare of others. It is the other-directed and pro-social nature of the moral emotions that instils in us a concern for interests that are not our own. Further, the moral emotions are clearly fundamental to the ordinary functioning of moral judgement and action—to both what we feel is morally permissible, and how we act as a result. As previously argued, there is good reason to

\textsuperscript{75} Ibid, p. 15.
\textsuperscript{76} Ibid, p. 15.
think that many of the moral emotions (or at least precursors to them) are also present in some non-human primates. As discussed in chapter one, evidence of other suffering emotions like sympathy, empathy, and community concern is given by a variety of behaviours. The presence of something approximating the other-condemning emotion of anger can also plausibly be interpreted as underlying the phenomenon of “moralistic aggression” or negative reactions to perceived violations of the group social code, suggested by studies of chimpanzees and macaques. Further, it is reasonable to speculate that a precursor to the other-praising emotion of gratitude may underlie the development of more cognitively demanding calculated reciprocal exchanges which appear to occur in some non-human primate species.

Evidence of behaviour that seems to require precursors to the self-conscious emotions of guilt and shame have also been documented by de Waal. It has been found, for example, that low-ranking male long-tailed macaques are overly submissive to dominants following illicit mating sessions. This occurs even when the indiscretion was not observed by dominants, and attempts are made to make the low-ranking male aware that the act was not observed. This suggests that inhibitions with regard to the reaction of dominants are rooted deeply enough to remain even when they are absent, and that some form of self-reproach is occurring. Even if the low-ranking male’s reaction were based

---

77 Ibid, p. 15.
78 Haidt notes that the presence of these emotions (or precursors to them) in non-human primates provides strong evidence that the morality is innate and is, like language, a major evolutionary adaptation. However, Haidt does not contend that morality is in no way enculturated. Rather, he argues that innate moral intuitions are shaped by culture during the development of an individual. Ibid, pp. 17-18.
79 Although Haidt does not directly discuss community concern, it seems reasonable to interpret it as a member of the other-suffering family directed not at individuals, but at the community as a whole.
81 Reciprocal exchanges are calculated when they are based on an ability to keep track of favours given and received. See chapter one.
on an a remaining fear of punishment resulting from an incomplete understanding of the
dominants' lack of knowledge, there is still reason to postulate the presence of a pre­
cursor to guilt. Although not the same as guilt, a fear of disciplinary actions by dominant
group members likely represents a stage in its evolution.83

Emotions such as sympathy, empathy, and community concern clearly provide
building blocks for the construction of rules for the regulation of natural and situational
advantages. The moral emotions motivate the maintenance of social order, and a regard
for the interests of others. As demonstrated by research into psychopaths and those with
damage to their VMPFC, without the moral emotions individuals do not react to the
suffering, needs, or interests of others or care about the effects of their behaviour. As
Haidt's research suggests, the moral emotions are elemental to the ordinary functioning
of moral judgement and action—to both what we feel is morally permissible, and how we
act as a result. Their other-directed and pro-social nature connects us to the needs and
interests of others in a way that encourages advantage regulation—and thus advantage
regulating rules—to arise.

As noted above, prescriptive rules arise when individuals learn to recognize the
contingencies between their own behaviour and the behaviour of others, and to act in
ways that minimize negative consequences. The moral emotions appear custom designed
to aid in the identification and navigation of precisely these social contingencies.
Morality, it seems, is a product of the moral sense, shaped into rules by evolved
dispositions, and discovered through the emergence of cognitive and reflective powers.
Accordingly, it is possible to see how the building blocks Flack and de Waal identify
provide raw materials from which moral systems can be constructed. As a result, it must

be concluded that we have satisfactory evidence that morality is in significant part biological in origin.
Although we have substantial reasons to believe that morality has biological origins, the question remains as to whether or not these origins are of any significance to substantive moral philosophy. Many logically and conceptually problematic arguments have been put forward contending that biological facts have implications for moral philosophy. Social Darwinists such as Herbert Spencer and William Graham Sumner argued that facts about human evolution directly entailed particular substantive moral principles. As Sumner warned, “[I]t is understood that we cannot go outside of this alternative: liberty, inequality, survival of the fittest; not-liberty, equality, survival of the unfittest. The former carries society forward and all its best members; the latter carries society downwards and favors all its worst members.”¹ According to the Social Darwinists, natural selection dictates that we adhere to an extreme \textit{laissez-faire} individualism. We have a positive moral obligation to foster an environment in which the struggle for existence can do its work. We are morally obligated to promote the forces of evolutionary change, or at least not stand in the way of their execution.²

Social Darwinist arguments like those advanced by Spencer and Sumner were rejected partly because they yielded some of the most “grotesque socio-politico-economic suggestions of our time.”³ They were also rejected because they committed a serious logical error—the naturalistic fallacy—by sliding from facts about what is the case to imperatives about what ought to be the case. Social Darwinists began with facts about

¹ Sumner, p. 79.
² Spencer, p. 346.
evolution and natural selection, and proceeded to derive moral imperatives from these facts without explaining the relationship between facts and values, or how the logical chasm between the two is to be bridged.

In the preceding chapters, I have argued that there is significant evidence that morality has biological origins. We have seen that the development of genuine altruism, often argued to be a distinguishing characteristic of human morality, can be explained biologically. As Flack and de Waal note, if human morality has biological origins, we ought to expect that elements of it will be present in other closely related social species. We have seen that this is indeed the case, and that elements of moral systems such as a sense of social regularity, and a capacity for moral emotions, are evidenced by conflict-management behaviours observed in some species of non-human primates. We have also seen how these elements provide building blocks from which moral systems can be constructed. In the chimpanzee sense of social regularity we see the beginning of prescriptive rules, and even of rules that specifically serve to limit the expression of dominance and regulate advantage. Further, we have seen that moral emotions such as sympathy and empathy inform the basis of moral judgement and action, and serve to connect us to the needs and interests of others in a way that aids in advantage regulation, and thus encourages advantage-regulating rules to arise.

Given the significant biological foundations of morality, it would be surprising if they had no impact, or placed no constraints on moral theorizing. As Michael Ruse has argued, “[b]eing a modified monkey counts, and nowhere more so than in the realm of social and moral behavior and thought.”

Yet, in light of the naturalistic fallacy, it may

---

3 Ruse (1990), p. 58.
appear that, no matter how interesting facts about the biological origins of morality may be, they can have no relevance to substantive moral philosophy without a solution to the significant and enduring logical problem of how an “ought” can be derived from an “is”. The naturalistic fallacy precludes the direct derivation of moral principles from biological facts (or facts of any other kind, for that matter). However, attempting to directly derive moral imperatives is only one way in which facts about the biological origins of morality could be brought to bear on moral philosophy.

Bernard Williams has objected in *Moral Luck*, that modern moral theories such as Kantianism and utilitarianism make such strong demands that “[o]ne is left, at any level of importance, only with purely moral motivations, and no limit to their application.”

Further, as a result, one has “no life of one’s own, except for some small area, hygienically allotted, of meaningless privacy.” Kantianism and utilitarianism are highly exigent moral theories. As Wilson notes, these theories are often seen as “too interfering and repressive where individual lives are concerned, and as provoking moral scepticism by setting standards of benevolence and self-sacrifice that normal human beings cannot ordinarily meet.”

Wilson notes that, although they are not fully hypermoral, utilitarianism and Kantianism are widely recognized as tending in the direction of hypermorality without reaching its extremes. They require profound renunciations of advantage, and do not leave any room for flexibility, or concessions for self-interest, in the application of moral

---

6 Ibid, p. 38.
7 Ibid, p. 58.
8 Ibid, p. 54. Wilson characterizes hypermoral positions as implying “a profound renunciation of advantage, a refusal to think flexibly and pragmatically, or to deploy all the resources that one has at hand, whether natural or cultural, native or acquired, to come out ahead, or even to equalize an unequal situation.” Ibid. p. 3.
rules. These theories are high-demand, and issue imperatives that are often out of touch with our moral intuitions and our everyday experience of right and wrong—with our intuitive sense of what morality requires of us. They impose moral rules that regularly “oppose our natural sentiments, frustrating their expression.” Utilitarianism and Kantianism often compel us to disregard the special relationships we have with our family, our friends, and with ourselves. These moral theories posit an overall obligation to consider the interests of others that is independent of the individual, and of any rewards that altruism may bring. As such, we feel burdened by them, resist them, and offer reasons for our non-compliance.

It is the purpose of this chapter to investigate whether or not the biological foundations of morality, and in particular its strong basis in the moral emotions, provides grounds for the rejection of particular kinds of highly exigent demands. As just noted, the demands of highly exigent moral theories such as utilitarianism and Kantianism often fail to accord with our sense of what we are morally required to do. I will argue that such exigent moral demands are disconfirmed, and are thus not morally binding, on the grounds that they are not supported by our experience of moral phenomena. In doing so, I will not argue that moral demands must be identical with factual evolutionary claims, or that moral demands must be directly derivable from factual evolutionary claims to be legitimate, as each of these violate the is/ought distinction. Rather, I will argue that moral imperatives must ultimately be explainable by factual evolutionary claims. Legitimate moral obligations must ultimately be explainable by facts about those deep preferences (or patterns of preference) built into our biological nature.

9 Ibid, p. 52.
4.1 Exigency in moral theory

Given what we know about the biological foundations of morality it is clear that, at least descriptively, as a phenomenon morality has something to do with the kinds of creatures we are as human beings. Human morality did not arise de novo, and it did not arise by the proclamation of some deity ex cathedra. Rather, it appears that morality is a product of the moral sense, shaped into prescriptive rules by evolved dispositions, and discovered through the emergence of cognitive and reflective powers. Morality is not a product of divine decree, or the reflection of some heavenly form of goodness, but of human nature. It is intimately bound to the kinds of creatures we are. In accordance with this, it is commonly accepted in moral philosophy that the "moral norms governing something must depend at least in part on what that something is." Moral philosophers since the ancient Greeks have assumed that there must be some connection between a moral theory and the beings that it is designed to govern. As John Rawls has succinctly put it "the correct regulative principle for anything depends on the nature of that thing." At bottom, formulas of moral obligation must have something to do with the kinds of creatures we are if they are to make moral demands of us at all.

The exigent demands of utilitarianism and Kantianism, however, cannot be dismissed out of hand on the grounds that they fail to connect the force of regulative principles with some account of the kinds of beings humans are. As Wilson notes, Jeremy Bentham, the father of classical utilitarianism, presented utilitarianism as being

---

11 ibid, p. 52.
12 Noggle, p. 533.
13 Noggle, p. 563.
14 Rawls, p. 29.
"based on a simple yet comprehensive account of human motivation."\textsuperscript{15} Quite simply, that we are motivated to pursue pleasure, and to avoid pain. As Bentham argued, "[n]ature has placed mankind under the governance of two sovereign masters, pain and pleasure. It is for them alone to point out what we ought to do, as well as what we shall do."\textsuperscript{16} Accordingly, standards of right and wrong are "fastened to their throne."\textsuperscript{17} Bentham also argued that, as it is pain and pleasure alone that are good and evil, it can be only the consequences of acts that matter. Acts, Bentham maintained, are to be morally evaluated only on the basis of the pain or pleasure that they create.

Utilitarians also contend that, in the calculation of pleasure, happiness, or "utility", no one person's interests are to be counted as more important than any others'. As Mill, a later utilitarian, argued in his work \textit{Utilitarianism},

\begin{quote}
the happiness which forms the utilitarian standard of what is right in conduct, is not the agent's own happiness, but that of all concerned. As between his own happiness and that of others, utilitarianism requires him to be strictly impartial as a disinterested and benevolent spectator.\textsuperscript{18}
\end{quote}

Mill developed on Bentham's formulation by distinguishing happiness from base physical sensation, and ranking pleasures with regard to both their quantity and quality (awarding greater value to those that engage the higher cognitive faculties). It is, Mill famously wrote, "better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied."\textsuperscript{19} Later utilitarians have attempted to further improve on Bentham's original account by continuing to broaden the range of states that are intrinsically valuable, and by arguing that it is rules and not individual acts that are to be evaluated for the utility they produce. Regardless of these modifications, at heart the

\textsuperscript{15} Wilson, p. 57.
\textsuperscript{16} Bentham, p. 11.
\textsuperscript{17} Ibid, p. 11.
utilitarian position is epitomized by its tenets that the only thing that is good in itself is some specific state (e.g., pleasure, or happiness), and that right acts (or rules of conduct) are determined purely as a function of their consequences—of how much of an intrinsically valuable state they produce. Those that result in the greatest good for the greatest number are morally right, and those which diminish overall utility are morally wrong. Thus, according to the utilitarian we are morally obligated to perform those acts which will result in the greatest good for the greatest number.

In contrast, Kant maintained that we are morally obligated to perform those acts dictated by the absolute moral law, regardless of their consequences. At the heart of Kant’s ethical theorizing is his supreme principle of morality—the Categorical Imperative. This imperative is binding on all rational beings unconditionally, and has various formulations. The most prominent of these is the universalizability version that “I ought never to act except in such a way that I can also will that my maxim should become a universal law.”20 Alternatively, there is also the practical version: “Act in such a way that you always treat humanity, whether in your own persons or in the person of any other, never simply as a means, but always at the same time as an end.”21 It is from the Categorical Imperative that all specific substantive rules are to be derived, as it summarizes a procedure for determining which acts are morally permissible.

Kant argued that the source of moral obligation is to be found “not in the nature of man nor in the circumstances of the world in which he is placed, but solely a priori in the

19 Ibid, p. 140.
20 Kant, p. 261.
21 Ibid, p. 274. Kant also gave a third version: “[A]ct only so that the will through its maxims could regard itself at the same time as universally lawgiving.”
concepts of pure reason. Morality is absolute and categorical, and thus cannot have arisen from the contingencies of human nature or circumstance, Kant held that "a law has to carry with it absolute necessity if it is to be valid morally—valid, that is, as a ground of obligation." Kant maintained that morality is something (synthetic \textit{a priori}) necessary which emerges from the interaction of rational beings. Our capacity for rationality allows us to know and revere the moral law. Because humans are "endowed with a faculty of reason that enables them to conceive themselves as partially exempt from the laws of nature, they can be roused to the admiration of pure moral agency and respect for the moral law." It is when we act from the motive of pure reverence for the moral law alone, or from the "good will," that we act morally on Kant's account. Intention is everything, and consequences, either to ourselves or others, play no part in determining the rightness of any act.

According to Kant, the source of our moral obligation is our rationality. Morality is a set of laws that rational creatures spontaneously prescribe to themselves in accordance with the Categorical Imperative. It is in virtue of the fact that humans are rational creatures that we are subject to the moral law at all. In this respect, Kant's moral theory stays within the constraint that formulas of moral obligation must have something to do with the kinds of creatures we are if they are to make legitimate moral demands.

Both utilitarianism and Kantianism are highly exigent moral theories. Even though these theories connect the force of regulative principles with some account of the kinds of beings humans are, they issue imperatives that are often seen as interfering and

\begin{footnotes}
\footnote{22 Ibid, p. 256.}
\footnote{23 Ibid, p. 256.}
\footnote{24 Ruse (1990), p. 67.}
\footnote{25 Wilson, p. 54.}
\end{footnotes}
repressive, and set standards of benevolence and self-sacrifice that normal human beings cannot ordinarily meet. Further, they often fail to connect to our intuitive sense of what is morally required of us. To see this clearly, consider the following examples. The utilitarian asserts that we are morally obligated to perform those acts (or follow those rules of conduct) that will result in the greatest good for the greatest number. Accordingly, if it were the case that neglecting one’s own family in order to campaign heavily for the poor of an impoverished nation resulted in higher overall utility, then one would be morally required to do so. Indeed, as much of the world is very poor, one is likely required to donate all of one’s income, above that required to maintain a minimal level of subsistence, to charity. Dining out, watching movies, engaging in personal hobbies and projects, and even buying birthday gifts are all morally forbidden where that money or time could be used to contribute to the greater good.

According to Kant’s moral system, we are required to follow the moral law regardless of any consequences. Moral rules are absolute and hold for all persons under all circumstances. A well known example of a rule that passes the Categorical Imperative is that of not lying. Should a known child molester, or murderer come to your door and inquire as to whether or not your children are at home, Kant would maintain that you must answer the villain honestly—even if it means grievous injury to your offspring. With regard to the time and money spent on those things we enjoy above a minimal level of subsistence, the moral demands of Kantianism and utilitarianism are the same. The Kantian too would will it to be universal law that people in wealthy

---

26 One could not will it to be universal law that everyone lie, as this would be self-defeating. People would quickly learn that they could not believe what others say, and thus lies would no longer be believed.
nations forgo their dinners out, movies, and the expenses of their personal projects and hobbies, if it would save strangers in developing nations from starvation.\(^{27}\)

As Bernard Williams has noted, "[t]here is no limit to what a given person might be doing to improve the world, except the limits of time and strength."\(^{28}\) Yet, the majority of us who could be sacrificing all our luxuries, or donating all our spare time to those in need do not do so. Although we feel sympathetic towards those who are less fortunate, and may often donate our time and money to charitable causes in the developing world and at home, we do not feel that we are under obligations that are as strong and binding as those commanded by utilitarianism and Kantianism. I doubt it too implausible a conjecture that there are few people who would think it permissible, or even required, to neglect the legitimate needs and interests of their own children in order to campaign for someone else’s.\(^{29}\) It is difficult to imagine a person of sound mind who would permit a child molester or murderer access to their children because they are morally required not to lie. Both utilitarianism and Kantianism connect the force of regulative principles with some account of the kinds of beings humans are. However, given the demands they make of us, it appears that they have got something terribly wrong.

### 4.2 Moral sentiment and feelings of obligation

---

\(^{27}\) Wilson, p. 61.

\(^{28}\) Williams (1985), p. 77.

\(^{29}\) I count among legitimate needs and interests those whose satisfaction contributes to a generally agreeable (rather than minimal) level of health, wealth, education, enjoyment, and the like. I will not argue for an account of legitimate needs and interests here. Clearly however, I do not want to count desires for maximal pleasure, wealth, power, or other extremes among legitimate needs or interests.
Although I have been considering utilitarianism and Kantianism in their more simplified forms, it has been noted that “even in their updated versions, the famous modern moral theories appear to a sector of their readers to suffer from the same underlying problems as their predecessors.” They still often fail to connect with our sense of what we are morally required to do. Wilson observes that “[t]he relationship between natural sentiment and moral obligation is unsatisfactorily developed in Kantianism, and incoherent... in utilitarianism. High-demand moral theories are not a by-product of deep instincts and dispositions. Rather they are a new imposition that arose as a response to industrialization, the emergence of contract relations between persons who did not know each other well, popular government, and the spread of objective knowledge about others.” As Wilson notes, not everyone is threatened with the demands of impersonal moral theory. It is only a small literate fraction of humanity that concerns itself with such theories. The majority of people in the world interest themselves very little with the problem of benevolence to strangers, and instead live in blissful ignorance and indifference.

Certainly, the fact that high-demand moral theories are a new imposition, or that most of the world does not concern themselves with them, in itself gives us no grounds to reject their demands. As stated at the outset, though, I maintain that the fact that these theories issue imperatives that strongly conflict with our intuitive sense of what morality requires of us does speak against them. It cannot be denied that we feel we have

30 Wilson, p. 57.
31 Ibid, p. 57. Wilson maintains that the relationship between natural sentiment and moral obligation is incoherent in utilitarianism because, although the utilitarian argues that we are entirely governed by pain and pleasure, she also maintains that one is obligated to suffer or even perpetrate direct harm if it brings about greater utility overall.
32 Ibid, p. 74.
33 Ibid, p. 74.
important moral obligations to help and co-operate with others. However, few if any among us live our lives as if we have equal responsibilities to all. As a general rule, individuals value their own needs and interests, and those of their family and friends, over those of strangers. Not only do our own children, for example, likely come first in our affections, we feel that they really ought to. These feelings of partiality are ours by nature, instilled through natural selection. As Ruse observes:

Given the mechanisms of biology, it seems highly improbable that there would simply be an even sentiment, applied indifferently to all and every human. One would expect stronger feelings towards close kin, weaker towards non-relatives in one’s group, and finally a falling away as one deals with complete strangers.

We care about ourselves and our loved ones so strongly that our partiality of concern is unshakeable. Wilson notes that “[p]artiality towards our own causes . . . appears to be resistant not only to extinction, but even to modification.” Although it is true that, in a limited range of cases, we can sometimes extend our concerns in the direction of impartial benevolence, it is “always for a specific reason, and not because the action performed accords with a formula of obligation derivable in a moral theory.” Although I may well forego a pair of stylish new shoes and send money to a charity because I am moved by a television appeal, I will never do so because I believe that utilitarianism or Kantianism are true or irrefutable. At best, I might offer a utilitarian or Kantian principle post-hoc as a way of explaining or exalting the motives behind my charitable act.

34 Ibid, p. 74.
35 Ruse (1986), p. 239.
36 Ruse (1990), p. 69.
37 Wilson, p. 62.
38 Ibid, p. 62.
I will also never sacrifice the legitimate needs or interests of myself or my loved ones for those of strangers because such an action accords with a formula of obligation derivable in a moral theory. I may do so because I feel sympathetic to those less fortunate, or guilty about good fortune at the expense of others, but not because it is my duty to promote utility for the greatest number, or to uphold the Categorical Imperative. As Peter Strawson showed clearly in “Freedom and Resentment”, there simply are limits to the changes in our basic commitments and practices that we can make for purely theoretical reasons. We could, Strawson argued, never abandon our common-sense view of persons as being (for the most part) responsible agents, no matter how irrefutable arguments for determinism appear.\textsuperscript{40} Just as the apparent irrefutability of determinism cannot stop us from holding others responsible, it appears that the fact that the modern moral theories demand impartiality cannot not stop us from holding our own preferences and desires, and those of people we care about, more dear than those of strangers.

As argued in the preceding chapter, we make judgements about our moral obligations primarily on the basis of quick gut feelings and affectively laden intuitions. We do not stop and try to assess what judgement and corresponding action would accord with a particular moral theory. Evidence suggests that moral reasoning is triggered only under special circumstances, and primarily as an ex-post facto social product.\textsuperscript{41} As we have seen, affective processing is triggered more quickly in real-time judgements, and yields judgements that are stronger and less likely to be revoked than those made by rational processing.\textsuperscript{42} Subject responses to the incest case (and cases like it), demonstrate that even when moral reasoning is engaged, it may be disregarded if it yields results that

\textsuperscript{40} Strawson, p. 68.
\textsuperscript{41} Haidt (2003), p. 18.
conflict with strong affective judgements. When it comes to our own interests and those of our families, our powerful feelings of importance, love, and responsibility yield judgements that are exceedingly unlikely to be revised just because they conflict with obligations derivable in a moral system.

4.3. Exigency and disconfirmation

The intentions of the preceding section were descriptive—to demonstrate the degree of dissonance between the exigent impersonal demands of the modern moral theories and our sense of what we are morally required to do. Although these facts are suggestive of reasons for rejecting such exigent impersonal moral demands, the objection that moral psychology has no bearing on the question of our real obligations must be acknowledged. Even if we are biologically programmed to think and feel in a certain way, we can still always ask whether it is morally right or even defensible to think and feel those ways. We have arrived once again at the difficult problem of the difference between “ought” and “is.”

In this final section, I will argue that facts about moral psychology do indeed have a bearing on the kinds of demands that can be morally binding. As stated at the outset, I will not argue that moral demands must be identical with factual evolutionary claims, or that moral demands must be directly derivable from factual evolutionary claims to present real obligations. As noted, each of these strategies violate the is/ought distinction. Rather, I maintain that legitimate moral obligations must ultimately be explainable by facts about those deep preferences (or patterns of preference) built into

---

43 Ibid, p. 2.
our biological nature. Demands that fail to be explainable in terms of such preferences are disconfirmed, and cannot be said to present us with legitimate moral obligations.

On Wilson’s view, a system of formulas of obligation, judgements, and ideals can be said to be confirmed when it “commands agreement by appropriately informed morally intentioned persons.” A system that has been confirmed can be regarded as true. Wilson proposes that “[t]heorists ought to try to confirm a system of obligations, judgements, and ideals, and should accept any system that has been confirmed.” Although theorists may not succeed in completely confirming a system, and may only be able to provide good reasons in favour of a theory, as a matter of professional obligation, theorists ought to aim at confirmation. Further, insofar as we agree that a moral theory is a theory about what we ought to do, we must agree that if a moral theory has been confirmed we ought to do what it says. We must agree that it presents us with binding moral obligations. In contrast, we can say that a system of formulas of obligation, judgements, and ideals is disconfirmed when it is rejected by appropriately informed and morally intentioned persons. A system that has been disconfirmed can be regarded as false, and as failing to require us to do what it says. Such a system fails to present us with obligations that are binding because they are derivable in that system.

Clearly, the exigent impersonal demands of the modern moral theories make them prime candidates for disconfirmation. Although it is difficult to precisely define the sort of person who is properly called “appropriately informed and morally intentioned,” it is reasonable to assume that such a definition should not rule out ordinary agents who are

---

44 Wilson, p. 87.
45 Ibid, p. 87. Wilson adds the qualifier here that we ought to accept such systems provided there is reason to believe that moral systems are intrinsically benign.
aware of the pertinent facts, and care about doing the right thing as they see it. We would not want to make such a definition so stringent that only the agreement of moral theorists could matter. As we have seen, although some moral theorists might, few ordinary agents would agree to exigent impersonal demands of the sort that command us to make great sacrifices for strangers at the expense of the legitimate needs and interests of our kith and kin. We are partial in our interests, and if we are honest with ourselves we realize that we would look with disdain at those who are not. A parent with means who kept his or her children on a bare subsistence level diet, in sufficiently warm but threadbare clothes, and did not provide them with the best education available—all in order to save funds for the poor of a developing nation—would be a poor parent indeed.

When we will not agree to demands derivable from a moral theory’s formulas of obligation—even when we are appropriately informed and morally intentioned—such a theory is disconfirmed. Yet, we must again return to the objection that our feelings, or what we would agree to, are not what is at issue when we are trying to determine what we are morally obligated to do. We must ask why such things matter at all. Historically, moral intuitionism has been the most widely recognized and thoroughly discussed method of justification for moral theories. It rests on the claim that a moral theory should not deviate too much from our pre-theoretical moral convictions (or at least those that we are prepared to hold on reflection). Moral intuitions are spontaneous moral judgements about a particular agent or act, type of act, or a more general moral rule or principle.

---

47 As noted in chapter three, evidence suggests that, although a person could in principle simply reason her way to a judgement that contradicts her initial intuitions, such an ability may be common only among philosophers, who have been extensively trained and socialized to follow reasoning even to very disturbing conclusions. Haidt (2000), p. 22.
48 Noggle, p. 531.
49 McMahan, pp. 93-94.
McMahan notes in his survey of recent literature on intuitionism, “our intuitions often compel belief in a way that, for most of us, no moral theory does.”50 Further, there is a surprising uniformity in our intuitions across a range of cases.51 As a result, many moral theorists have held that the correct way to determine our moral obligations is to discover what moral theory can best be brought into equilibrium with our pre-theoretical moral convictions.

Not all moral theorists, however, assign such an authority to moral intuition. Peter Singer has argued that “all the particular moral judgements we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past.”52 As a result, “it would be best to forget all about our particular moral judgements.”53 It seems the critic who charges that our feelings or intuitions about our moral obligations say nothing about what our true moral obligations are has still not been successfully answered. Our intuitions may simply be misguided relics, or unreliable guides to our true obligations, and thus lack authority.

In his work *Taking Darwin Seriously*, Ruse considers the question as to why morality evolved, and argues that morality is a biological adaptation that has arisen to provide us with the “extra push” required to get us to help and to co-operate with others. Our normal non-moral feelings are by their very nature generally limited and restricted, and for the most part selfish and immediate.54 As Haidt has argued, what is special, and distinguishes the moral emotions, is their ability to connect us to the needs and interests

50 Ibid, p. 97.  
51 Ibid, p. 106.  
52 Singer, p. 516.  
of others. They serve to take us beyond the direct interests of the self. Biologically, we need such feelings, and the rules they give rise to, to get us to interact in ways that will benefit us in the long term. If we help and co-operate with others they are more likely to return the favour. However, without morality it is less likely that we would be driven to sacrifice immediate natural or situational advantage for long-term gain.

As such, morality evolved to take us beyond our normal feelings, and to instill in us a sense of obligation to others. From a biological perspective, morality simply would not work if we did not believe that it is objective. Morality is an adaptation that takes us beyond our regular wishes, desires, and fears, and encourages social interaction, and it does this by filling us full of thoughts about our obligations and duties to others. Morality moves us to action precisely because we think that it is an external constraint—that although we have a choice about whether to do right or wrong, we have no choice about what right and wrong are in themselves. If people did not feel that morality is objective, they would be more likely to cheat in their co-operative relationships, and be disinclined to come to the aid of others (especially non-kin). As Ruse notes, “moral altruism would go, and with it would go biological altruism. Hence, our biology leads us

56 In his book Passions Within Reason, Robert Frank makes a similar argument. Frank argues that moral emotions evolved to help us solve an important and recurring problem of social interaction—the commitment problem. The commitment problem arises when it is in a person’s interests to behave in a way that will later seem contrary to self-interest. Moral emotions such as sympathy and anger help individuals to make, and then keep, their cooperative agreements even though material incentives at a given moment may unambiguously dictate defection. This allows individuals to reap the rewards of cooperative interaction that are in their long-term interests. pp. 46-51.
to 'objectify' morality."58 Morality requires an air of objectivity to accomplish what it has evolved to do.59

In a sense, then, substantive morality is a "collective illusion foisted upon us by our genes."60 However, Ruse is quick to note that "the illusion lies not in morality itself, but in its sense of objectivity."61 Morality itself is very real, it is the apparent objective referent of morality that is not. Once one accepts that morality is a biological adaptation, she is directed to a meta-ethical conclusion about its status—that morality is a function of human nature, that reduces to feelings and sentiments. It is ultimately subjective and has neither meaning nor justification outside of the human context.

There is no need to postulate morality "out there" beyond the feelings and sentiments of human beings. Indeed, to do so is redundant as there is nothing about morality left over to be explained. Further, even if there were an external morality we have no assurance that it would be cognitively accessible to us, within our capacity to understand, or benign.62 To suppose that natural selection would align our moral sentiments with the "true" morality is to "impute a direction to evolution altogether alien to modern biological thought."63 What evolves is what works, and had evolution taken our development down a different path we might have thought moral what we now find repugnant.64

This returns us once again to the distinction between "is" and "ought." We have seen that, on the basis of a particular factual story about the nature and process of

62 Wilson, p. 85.
64 Ibid, p. 254.

77
evolution, a complete explanation of morality can be provided. Although it may initially appear that this is a gross violation of the is/ought barrier, it is not, because no attempt to equate or logically derive moral claims from factual claims is involved. Instead, on Ruse’s account, evolutionary claims are used to explain away the apparent objective referent of morality. That which gives morality its binding prescriptive nature, and takes it beyond the factual, is explained by facts about why we have the feelings and sentiments that we do.  

As a result, the reasoned justification of morality can only be taken so far. There simply comes a point at which there are no further reasons to be had. As Jeffrie Murphy observes,

> Value judgements are properly defended in terms of other value judgements until we reach some that are fundamental... However, suppose that we seriously raise the question of why these fundamental judgements are regarded as fundamental... We reject simplistic utilitarianism because it entails consequences that are morally counterintuitive, or we embrace a Rawlsian theory of justice because it systematizes (places in ‘reflective equilibrium’) our pretheoretical convictions. But what is the status of those intuitions of convictions? Perhaps there is nothing more to be said for then than that they involve deep preferences (or patterns of preference) built into our biological nature.

Let us return then to the critic who charges that our feelings and sentiments about our moral obligations say nothing about what our “true” moral obligations are. Clearly, this cannot be the case. Indeed it seems that, in the end, our feelings and sentiments say everything about what our moral obligations are. If there is no objective referent of morality, and morality is ultimately a function of human nature that reduces to feelings and sentiments, then it follows that it can have neither meaning nor justification outside of the human context. As noted above, it has often been charged that whether humans are biologically programmed to think and feel in a certain way or not, the question remains as to whether it is morally right or defensible for us to think and feel in these ways.

65 Ibid, p. 256.
However, once it becomes clear that there is no external standard of morality, this question no longer makes sense. The way we think and feel is the only determinant of moral rightness there is.

As a result, facts about our moral psychology must indeed have a bearing on substantive moral philosophy. Recall that a system of formulas of obligation, judgements, and ideals is disconfirmed when it fails to command agreement by appropriately informed and morally intentioned persons. A system that has been disconfirmed can be regarded as false, and as failing to present us with obligations that are binding because they are derivable in that system. Clearly, Kantianism and utilitarianism are disconfirmed. As we have seen, many of the obligations derivable in these systems strongly conflict with our sense of what morality requires of us. In particular, we are highly resistant to those exigent impersonal demands that obligate us to disregard the special concern we have for the legitimate needs and interests of our family, our friends, and for ourselves. As such demands fail to be explainable in terms of those deep preferences built into our biological nature, they cannot be said to present us with legitimate moral obligations.

Although there appear to be limitations to the kinds of sacrifices we can be morally obligated to make, we have not merely ended up at the same grotesque conclusions as the Social Darwinist by a different route. We have powerful feelings of partiality towards our kith and kin, and ourselves. Partiality is a deep preference built into our biological nature. However, we also have powerful preferences for altruism, and for co-operation with others. Where funds can be spared, we feel that we ought to be charitable, where it is possible to save or improve the life of another at a manageable

---

66 Murphy, p. 112. n. 21

79
expense to ourselves, we feel we ought to do so. We are driven to help one another, and we take pleasure in improving one another's lives as a result. Moral emotions motivate us to come to the aid to others, to change relationships with those who treat others unfairly, to respect the interests of others, and to praise pro-social behaviour. They instill in us a sense of impartial obligation to others.

Inevitably, our preferences for partiality and concern for strangers come into conflict, and any system of formulas of obligation, judgements, and ideals will have to accommodate this conflict if it is to stand some chance of being confirmed. However, although any successful account of our moral obligations must take our preference for partiality into account, this by no means implies that we are free of obligations to others. It is part of our very nature that we can be moved to come to their aid. As Ruse observes, "[t]he whole point about ethics is that we are all in it together."68

BIBLIOGRAPHY


Darwin, Charles. (1979 [1871]) The Decent of Man, and Selection in Relation to Sex. place not given: Arden Library.


