# Morality as Natural History

## An adaptationist account of ethics

Oliver Scott Curry

o.s.curry@lse.ac.uk

Government Department

London School of Economics and Political Science

2004

## Abstract

What are moral values and where do they come from? David Hume argued that moral values were the product of a range of passions, inherent to human nature, that aim at the common good of society. Recent developments in game theory, evolutionary biology, animal behaviour, psychology and neuroscience suggest that Hume was right to suppose that humans have such passions. This dissertation reviews these developments, and considers their implications for moral philosophy. I first explain what Darwinian adaptations are, and how they generate behaviour. I then explain that, contrary to the Hobbesian caricature of life in the state of nature, evolutionary theory leads us to expect that organisms will be social, cooperative and even altruistic under certain circumstances. I introduce four main types of cooperation – kin altruism, coordination to mutual advantage, reciprocity and conflict resolution – and provide examples of 'adaptations for cooperation' from nonhuman species. I then review the evidence for equivalent adaptations for cooperation in humans. Next, I show how this Humean-Darwinian account of the moral sentiments can be used to make sense of traditional positions in meta-ethics; how it provides a rich deductive framework in which to locate and make sense of a wide variety of apparently contradictory positions in traditional normative ethics; and how it clearly demarcates the problems of applied ethics. I defend this version of ethical naturalism against the charge that it commits 'the naturalistic fallacy'. I conclude that evolutionary theory provides the best account yet of the origins and status of moral values, and that moral philosophy should be thought of as a branch of natural history.

## Acknowledgements

## Table of Contents

# Chapter 1 – Updating Hume

*[M]orals and politics would be very interesting if discussed like any branch of natural history.*
        – Charles Darwin[1]

## Introduction

What is morality and where does it come from? In the history of moral philosophy, there have been theological, cosmological, biological and sociological answers to this question. Some have argued that moral values are "thoughts in the mind of God". Some have argued that moral values are the products of rational reflection on objective truths about the universe, similar to mathematical truths. Some have argued that moral values are a product of human nature. And some have argued that moral values are merely social conventions or local cultural norms.

In his *Treatise of Human Nature*, David Hume set out to give an explanation of morality that was consistent with the rest of the natural sciences. As John Mackie observes, Hume saw the problem as follows:

> Here is this curious phenomenon, human morality, a cluster of attitudes, dispositions, practices, behavioural tendencies, and so on that we find almost universally among men, even in different societies and at different times; why is it there, and how did it develop? . . . [This question] may be answered in sociological and psychological terms, by constructing and defending a casual hypothesis; this is what Hume has done. . . . [The *Treatise*] is an attempt to study and explain moral phenomena (as well as human knowledge and emotions) in the same sort of way in which Newton and his followers studied and explained the physical world.[2]

---

[1] Quoted in: Cronin, 1992, p99.
[2] Mackie, 1980, p6.

Hume argued that morality is a product of human nature. According to Hume, the behaviour of humans (and other animals) is a product of passion and reason. Passion sets the goals of action, and reason works out how to achieve them. Some of these passions – for example, sympathy – are designed to promote 'the common good' of society. Hume called these 'moral passions'. The moral passions provide the springs of moral action and the criteria by which to judge the actions of others. The goals of these moral passions constitute the 'natural virtues'; and the ingenious ways that people found to extend the reach of the natural virtues – through invention and convention – constitute the 'artificial virtues'.[3]

By introducing 'the experimental method of reasoning into moral subjects' and placing the study of morality on a naturalistic and empirical footing, Hume left open the possibility that his account of morality could be updated in line with advances in science. The purpose of this thesis is to provide just such an update. I argue that Hume was basically correct about the origin and status of morality, and that we are now in a position to update Hume's account in line with what modern science tells us about the world and about our place in it.[4]

**Outline of the thesis**

The Darwinian update of Hume begins in Chapter 2 by placing the study of human nature and human behaviour in the context of modern evolutionary biology. According to the evolutionary account of psychology: 'passions' become a particular kind of 'adaptation'; 'values' become the goals that adaptations attempt to achieve; 'reason' becomes the information-processing that adaptations perform; and 'beliefs' become the states that adaptations can adopt. I show how this account of psychology incorporates the key insights of Hume's account whilst avoiding some of its problems.

---

[3] Hume, 1739/1985.

[4] "It is customary to study the works of these great political philosophers with the tools of textual analysis and intellectual history. One considers the writer's intellectual coherence or historical context rather than subjecting his theory to currently available scientific evidence. . . [but] the theories put forth by political philosophers [about human nature] *can* be evaluated more or less objectively." Masters, 1991, p144.

The next two chapters update Hume's account of the *moral* passions with what is now known about the evolution of 'adaptations for cooperation' in humans and other animals.

Chapter 3 shows that, contrary to the Hobbesian caricature of life in the state of nature, evolutionary theory leads us to expect that organisms will be motivated to cooperate and even be altruistic under certain circumstances. I introduce the adaptive problems represented by the four main types of cooperation – kin altruism, coordination to mutual advantage, reciprocity and conflict resolution – and provide examples of adaptive solutions from nonhuman species. In this way, evolutionary theory provides an explanation for the social behaviour of plants and animals. Biology can explain why and how organisms care for their offspring and their wider families, aggregate in herds, work in teams, practise a division of labour, communicate, share food, trade favours, build alliances, punish cheats, exact revenge, settle disputes peacefully, provide altruistic displays of status, and respect property.

Chapter 4 reviews the evidence for equivalent 'adaptations for cooperation' in humans. These include: adaptations for maternal care and for assessing paternity; 'theory of mind' and language; cheater-detection mechanisms and 'punitive sentiment'; costly and altruistic signals of fitness; and recognition and respect for private property. I also review the ways in which humans can be said to have 'extended their moral phenotypes' through the use of tools. I then show how this Darwinian account of human adaptations for cooperation corresponds to the Humean account of human moral passions.

Chapter 5 summarises the 'moral philosophy' that emerges from this naturalistic account of morality. I show that the view that moral values are the proximate goals of adaptations for cooperation makes sense of a variety of positions in traditional meta-ethics. I show that an account of the content of adaptations for cooperation provides a rich deductive framework in which to locate and make sense of a wide variety of

otherwise unruly, and apparently contradictory, positions in traditional normative moral philosophy. And I show how the account of decision-making implicit in evolutionary psychology provides a framework for the investigation of moral reasoning and moral discussion in the context of applied ethics.

Chapter 6 defends this Humean-Darwinian account of morality against the charge that it commits 'the naturalistic fallacy'. I show that 'the naturalistic fallacy' refers to a variety of different arguments, none of which pose any challenge to the Humean-Darwinian version of ethical naturalism.

Finally, Chapter 7 considers the future of ethics from this naturalistic perspective. I argue that if morality is an adaptation, then moral philosophy should be thought of as a branch of natural history, and moral philosophers should be biologists, psychologists and anthropologists. I then review some of the problems that such a naturalised moral philosophy should be tackling. These include: extending the list of adaptive problems of cooperation; filling the gaps in the existing empirical literature on moral psychology; and assessing whether Stone-Age intuitions continue to provide efficient solutions to Space-Age collective-action problems.

**Historical precedents**

The attempt to apply evolution to ethics is not new. Charles Darwin himself argued that "the moral sense" was an adaptation, designed by natural selection to facilitate cooperation between members of tribes.[5] But in the years immediately following Darwin, the opportunity to use the theory of evolution by natural selection to shed some light on the origins and status of moral values was squandered. For much of the 20th century, the idea that morality was an adaptation was overlooked in favour of theories that used evolution as a metaphor or fable of moral progress, or took various features of evolution or its products as the criteria of moral

---

[5] Darwin, 1871.

worth. These theories did much to bring 'evolutionary ethics' into disrepute.[6]

Evolutionary ethics didn't really return to the idea that morality was an adaptation until the flowering of evolutionary theory in the 1960s and 70s – as represented by the work of William Hamilton, George Williams, Robert Trivers, John Maynard Smith and Richard Dawkins.[7] EO Wilson can perhaps be credited with getting the ball rolling again. In *Sociobiology*, his massive synthesis of animal behaviour, he suggested that "Scientists and humanists should consider together the possibility that the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized".[8] In subsequent years, Peter Singer has argued that "human ethics has its origins in evolved patterns of behaviour among social animals".[9] Robert Trivers has argued that our "sense of fairness", and emotions such as amity, revenge, gratitude, sympathy, guilt, and a sense of justice, can be explained as adaptations for cooperation.[10] Robert Frank and a host of other 'behavioural economists' have argued that many human emotions – such as guilt, or revenge – can be seen as natural selection's solutions to various collective-action problems.[11] Geoffrey Miller has argued that human morality is a result of sexual selection: "a system of sexually selected handicaps – costly indicators that advertise our moral character".[12] Even John Rawls sees human moral sentiments as "the outcome of natural selection". "[T]he capacity for a sense of justice and the moral feelings," says Rawls, "is an adaptation of mankind to its place in nature."[13] And, closest to my own project, Michael Ruse sees himself as providing a Darwinian update of

---

[6] Farber, 1994.

[7] Dawkins, 1976; Hamilton, 1964; Maynard Smith, 1982; Trivers, 1971; Williams, 1966.

[8] Wilson, 1975, p562.

[9] Singer, 1981, p29.

[10] Trivers, 1983; Trivers, 1971.

[11] Elster, 1999; Frank, 1988; Nesse, 2001b; Schelling, 1978.

[12] "We have the capacity for moral behavior and moral judgments today," argues Miller, "because our ancestors favoured sexual partners who were kind, generous, helpful, and fair." Miller, 2000a, p294, p292.

[13] Rawls, 1971, pp503-4. Other notable attempts to investigate morality from an evolutionary or biological perspective include: Alexander, 1987; Arnhart, 1998; de Waal, 1996.

Hume, and has argued that "Darwinian meta-ethics . . . is almost exactly what one would expect from the pen of Hume, were he writing today."[14]

My account differs from these previous versions of evolutionary ethics in several important ways.

First, some theorists have been unclear as to the status of their argument, and have ended up moving from biological facts about human nature to normative statements about how we ought to live. (And have thus become unnecessarily tangled up in "the naturalistic fallacy".) The *most important* feature of Humean-Darwinian ethics is that it is primarily *a meta-ethical theory about what values are*. It is not a simple-minded attempt to derive normative statements from descriptive premises. In this respect, Humean-Darwinian ethics is no different from other meta-ethical theories – such as the view that values are thoughts in the mind of God, or arbitrary social conventions, or the conclusion of rational reflection – that begin by stating what, in fact, values are.

Second, many theorists have been unclear about the difference between gene-level and individual-level selection. Despite works with such unambiguous titles as *The Selfish Gene*, they have persisted in assuming that natural selection operates at the level of the individual and that, as a result, evolutionary theory expects individuals to be selfish.[15] Individual altruism is then seen as an anomaly to be explained by something other than evolutionary biology, or by postulating mechanisms that 'constrain' natural selfishness and rationality. In this thesis I take it for granted that genes are the units of selection, that under certain circumstances selfish genes will build selfless people, and that it is not necessary to go outside mainstream biology to explain human altruism.

---

[14] Ruse, 1986, p266. "I think of my position as being essentially that of David Hume brought up to date by Charles Darwin. . . . Hume is my mentor because he went before me in trying to provide a completely naturalist theory of ethics. He was no evolutionist, but he wanted to base his philosophy in tune with the best science of his day." Ruse, 1995, p256. Dennett discusses how close Hume came to foreshadowing the theory of evolution itself: Dennett, 1995, p28-33.

[15] "[T]he unit of selection in the Darwinian model is the individual. . . . If human nature, too, was shaped by the forces of natural selection, the apparently inescapable conclusion is that people's behaviour must be fundamentally selfish". Frank, 1988, p23.

Third, some evolutionary theorists have taken the Aristotelian view that the good life consists of fully developing and expressing all one's natural faculties; and that the significance of evolutionary theory is that it reveals what is natural for humans. These theorists then have to defend the premis that 'what is natural is good'. Humean-Darwinian naturalism makes no such assumption. Instead of taking all natural human goals as the standard of good, the Humean-Darwinian account takes 'the common good' as the standard, and uses this criterion to distinguish moral goals from the rest. Hence the account does not argue that all natural goals are good, but merely that good goals happen to be natural, which is a very different proposition.

Fourth, many theories have emphasised some adaptations for cooperation but not others. This results in an attempt to shoehorn all ethical phenomena into a few explanatory theories. When that fails, the conclusion is usually that there is 'more to ethics than evolution'. In this thesis I emphasise that there are many pathways to the evolution of human moral sentiments, and discuss four main categories of adaptations for cooperation – kin altruism, coordination to mutual advantage, reciprocity and conflict resolution. As far as I can tell, this has not been done before in evolutionary ethics. I show how starting from this broader base provides a more secure foundation for evolutionary ethics, and I argue that only after providing the most comprehensive treatment of adaptations for cooperation will we be in a position to assess whether or not there is more to ethics than evolution.

**The current terrain**

There exists in contemporary moral philosophy something of a standoff between scientists and philosophers. Scientists have in their possession a working model of moral psychology, but they don't realise that, according to Hume at least, this is also a theory of moral value. As a result, they are reluctant to draw any conclusions about morality *per se* for fear of trespassing on moral philosopher's turf and coming face to face with the fearsome 'naturalistic fallacy'. Moral philosophers, meanwhile,

express a passing interest in the biologist's endeavours but, chastened by previous dalliances with evolutionary theory, and convinced that theirs is a discipline set apart from natural science, argue that evolution can be of little relevance to ethics.[16] But moral philosophers cannot avoid making empirical assumptions about the way that the world works, about human nature, or about the nature of moral values. The result is that moral philosophers often fall back on intuitive, folk, ancient and medieval theories of psychology when investigating human morality.

This standoff cannot continue indefinitely. The scientists' model of moral psychology is set to improve in the wake of discoveries in genetics, animal behaviour, developmental psychology, cognitive neuroscience, artificial intelligence, robotics, and brain-imaging. There will come a time when it is no longer possible to ignore the chasm between what science is telling us about human nature and morality, and the outdated framework in which moral philosophy is conducted. At this point, scientists and philosophers will realise that they need to work together in order to identify the genuine problems of moral philosophy and to make progress towards their solutions. The over-arching purpose of this thesis is to bring about this new way of conducting moral philosophy sooner rather later.

---

[16] For a recent manifestation of anti-naturalism in contemporary ethics, see: Held, 2002.

## Chapter 2 – The origin of value

*[I]t is not profitable for us at present to do moral philosophy; that should be laid aside at any rate until we have an adequate philosophy of psychology, in which we are conspicuously lacking.*
　　　　– G. E. M. Anscombe[17]


**Introduction**

According to David Hume's account of psychology, human behaviour is the product of desires (or 'passions'), reason and belief. Desires set the goals of action; and reason operates over beliefs to work out how to achieve these goals. Hume thought that values should be understood as the product of desire. Desires determine what states of the world humans find agreeable, pleasurable, and *valuable*. Hume thought that desires project value on to the world in much the same way that the visual system projects colour onto the world; and he dismissed as 'vulgar philosophy' the idea that values, or colours, exist independently of human nature.

The purpose of this chapter is to ground Hume's psychology in modern evolutionary biology, and thereby present a Darwinian update of his subjectivist theory of value. To this end, I review: how evolution by natural selection gives rise to adaptations; how adaptations pursue goals; and how the goals of adaptations can provide a theory of value. According to the Darwinian update of Hume, a 'desire' is a particular kind of adaptation; a 'value' is the goal that an adaptation attempts to achieve; and objects or states of the world are 'valuable' to the extent that they are conducive to achieving that goal. I end the section by explaining why Humean ethical naturalism is to be preferred over Aristotelian ethical naturalism.

An additional advantage of the Darwinian update is that its account of information processing overcomes several problems with Hume's

---

psychology that have to do with how organisms acquire knowledge about their environments. Hence I argue that 'reason' should be seen as the information-processing that adaptations perform; and 'belief' should be seen as the different states that adaptations can adopt. And I show how this model of reasoning overcomes various versions of 'the frame problem' and 'the problem of induction'.

Hume's purpose in setting out a naturalistic, 'desire theory' of value was to lay the foundations for a naturalistic, 'desire theory' of *moral* value.[18] In a similar way, the discussion of the nature of adaptations in this chapter sets the scene for the discussion of adaptations for cooperation in the next chapter.

**Darwin's theory of evolution by natural selection**

Why do giraffes have long necks? Why are plants and animals so exquisitely well-suited to their environments? Why do species change, and where do new species come from? What is life, and how did life start in the first place? Charles Darwin's theory of evolution by natural selection provides an answer to all these questions.[19]

Darwin observed that, in a given population of plants or animals, individuals varied slightly in their ability to survive and reproduce. He also noted that individuals passed these abilities on to their offspring. Darwin then showed that if there were competition for the scarce resources needed to survive and reproduce, it would inevitably follow that the composition of the population would change. Individuals with traits that were better-suited to survival and reproduction would have more offspring, and hence these traits would become relatively more common in subsequent generations. So, for example, suppose that in a population of giraffes, a) some have longer necks than others, b) offspring inherit 'neck length' from their parents, and c) longer necks enable the giraffes to reach more leaves and subsequently have more offspring than

---

[18] See Ernest C. Mossner's introduction to Hume, 1739/1985.
[19] Darwin, 1859.

giraffes with shorter necks. If these three conditions hold, then it will inevitably follow that the average neck length of the giraffes in the population will increase over time. Given a constant supply of variation, this process could go on indefinitely, with members of the population accumulating features that enabled them to survive and reproduce better in their respective environments. Populations of similar organisms living in different environments may evolve in completely different directions, giving rise to new species. Conversely, if you were to wind the tape backwards, you would see complicated organisms evolving from simpler common ancestors, all the way back to the first glimmers of life: the chemicals in the primordial soup that had the unusual property of making copies of themselves.

Whereas Darwin focused on individual organisms as the locus of adaptation, modern biologists now focus on genes.[20] This is because individuals do not make replicas of themselves when they reproduce; but their genes do. As Richard Dawkins puts it: "The true unit of natural selection has to be a unit of which you can say it has frequency. It has a frequency which goes up when its type is successful, down when it fails. This is exactly what you can say of [buffalo] genes in gene pools. But you can't say it of individual buffaloes. Successful buffaloes don't become more frequent."[21] According to this 'gene's eye view' of evolution, genes replicate themselves by virtue of the effects that they have on the world. These effects include the construction of organisms. (Organisms can therefore be seen as the means by which genes replicate.) Genes that equip organisms with traits that are better-suited to survive and reproduce under prevailing conditions become more frequent in the

---

[20] Dawkins, 1976; Williams, 1966.

[21] Dawkins, 1998b, p217. The view that natural selection consists of the differential replication of genes is often, erroneously, thought to contradict the view that biological systems are organised hierarchically into levels such as chromosomes, cells, individuals, and groups. However, these two views are not in conflict. It is simply the case that genes replicate themselves by means of their effects on the world, and that these effects include the formation of chromosomes, cells, individuals, groups and so on. The problem for biology is to specify the conditions under which genes for such 'higher-order' entities are favoured. (As a part of that project, this thesis aims to explain how gene-selection can give rise to adaptations for cooperation between individuals.) For this reason it is useful to distinguish between the "unit of selection" (which is always the gene) and the "level of adaptation" (which can be at any point along the resulting phenotype). For an overview of recent literature in this area, see: Keller, 1999.

population – the gene pool – than alternative genes (alleles). This process of differential replication is called *natural selection*; the resulting gradual change in the frequency of genes in the population is called *evolution*; and resulting well-designed replication-promoting features of organisms are called *adaptations*.

**Adaptations and their goals**

Biologists often distinguish between the *ultimate* and *proximate* goals of an adaptation. The *ultimate* goal of an adaptation is always to promote genetic replication; but, as we shall see, adaptations achieve this ultimate goal by pursuing a variety of more immediate or *proximate* goals, such as maintaining a certain body temperature, finding food, resisting parasites, or attracting mates.[22] In order to give a rounded picture of *how* adaptations achieve their proximate goals, it is necessary to draw attention to three aspects of the ways that adaptations can work: conditionality, movement and feedback.

*Conditionality and information-processing*

Adaptations – and the bundles of adaptations called organisms – can be seen as the genes' 'hypotheses' (theories, expectations, assumptions) about how to replicate in the kind of world into which they will be born.[23] Mutation generates novel hypotheses, and natural selection puts these hypotheses to the test, retaining the successful ones. So, a bird's wing is a hypothesis about how to get from A to B given the principles of aerodynamics; a snail's shell is a hypothesis about how to protect the snail's body given the earth's gravity and the typical strength of predators; a polar bear's fur is a hypothesis about how to remain hidden from prey given the colour of the world that it will inhabit; and so on.[24] In

---

[22] As George Williams puts it: "Each part of the animal is organized for some function tributary to the ultimate goal of the survival of its own genes." Williams, 1966, pp255-6. The ultimate-proximate distinction is often used to distinguish between *why* and *how* questions. So, sometimes biologists are interested in *why* natural selection designed an adaptive mechanism in a particular way; at other times, biologists are interested in *how* a given adaptation works.

[23] Dawkins, 1976, p55; Popper, 1972.

[24] Of course, none of these predictions need be explicit; the phenotype might not make any 'mention' of the level of gravity. It's just that we can see these adaptations as having been built on

this sense, thanks to millions of years of trial-and-error on the part of natural selection, organisms come into the world knowing what to expect.[25]

Some adaptations – such as the snail's shell – 'expect' a certain aspect of the world to be stable, and do their job of promoting replication (in this case, by protecting the snail's body from physical trauma) by remaining relatively fixed. Other adaptations 'expect' the world to vary, and do their job by adopting different states under different conditions.[26] These 'conditional' adaptations are 'uncertain' about which state to adopt; the uncertainty is resolved – a decision is taken – by attending to the conditions specified by the adaptation.[27] For example, a certain species of sea moss 'expects' that it will sometimes be preyed upon by sea slugs. This moss has the option of growing defensive spikes to protect against the predatory slugs. But the spikes are costly, and they are worth growing only if the slugs are present. Hence the sea moss is sensitive to the particular chemical cue indicating the presence of the slugs, and operates according to the rule: "If slugs, then spikes".[28]

In the moss's case there are only two possible states, and so the degree of prior uncertainty is halved when the slugs are detected. In information theory, the reduction of uncertainty by one half constitutes one 'bit' of information; hence the moss's adaptation is a one-bit information-processor.[29] Conditional adaptations can of course consist of an indefinitely large number of rules. Such adaptations are uncertain about more aspects of their world, process more information before making a

the assumption that the world works in a certain way. Consequently, a Martian biologist would be able to infer the gravitational pull of the earth by inspecting the snail's shell.

[25] "The tentative solutions which animals and plants incorporate into their anatomy and their behaviour are biological analogues of theories . . . Just like theories, organs and their functions are tentative adaptations to the world we live in." Popper, 1972, p145.

[26] Natural selection will favour conditional phenotypic effects when the benefits of changing state in response to changes in environmental conditions outweigh the costs of setting up the machinery required to do so. Godfrey-Smith, 1996.

[27] Other terms for conditional adaptations include: facultative, plastic, 'if/then' rules, and conditional strategies.

[28] As this example illustrates, a conditional adaptation embodies "a genetically based program (decision rule) that results [in]. . . alternative phenotypes (tactics)"; and this program operates "through a mechanism (physiological, neurological, or developmental) that detects appropriate cues and puts the strategy's decision rule into effect . . . ". Gross and Repka, 1998, p169-70.

[29] Dawkins, 1998a; Dawkins and Dawkins, 1973.

decision, and hence are capable of generating more sophisticated behaviour.[30]


*Movement and motivation*

Some conditional adaptations result in movement. These adaptations are *uncertain* about the location of various adaptive targets and hazards, and are equipped with systems that allow them to detect and move towards some things and away from others. Some single-cell organisms, for example, migrate up chemical gradients towards food; heliotropic plants move their leaves so as to maintain maximum exposure to sunlight.[31] But the "trick of rapid movement" has been developed most by the group of organisms known as animals.[32] Whereas plants tend to sit and wait for the good things in life to come to them, animals are go-getters.[33]

For example, noctuid moths are equipped with bat-avoider mechanisms. The moth's wing muscle is connected, via its nervous system, to an 'ear' on the opposite side. The moth's ear is sensitive only to the echolocation frequency of an approaching predatory bat. It operates like a simple circuit, which we might describe as embodying the strategy: "If 60khz, then dive". When the sonar hits the ear, the switch is triggered, the muscle is turned off, and the moth tumbles out of harm's way.[34]

---

[30] As Daniel Dennett has said: "Switches (either on/off or multiple choice) can be linked to each other in series, in parallel, and in arrays that combine both sorts of links. As arrays proliferate, forming larger switching networks, the degrees of freedom multiply dizzyingly…". Dennett, 2003, p162.

[31] Darwin, 1880/1994.

[32] Dawkins, 1976, p47.

[33] In plants, communication between effector and detector cells is largely chemical. In animals, it is electric. This allows animals to update their state in response to a changing world at speeds fast enough to avoid falling over or bumping into things. As Sir Fred Hoyle puts it: "Looking back [at evolution] I am overwhelmingly impressed by the way in which chemistry has gradually given way to electronics. . . . [P]rimitive electronics begins to assume importance as soon as we have a creature that moves around . . . The first electronic systems possessed by animals were essentially guidance systems . . . analogous to a guided missile" quoted in Dawkins, 1982. Dawkins comments that this is "what . . . any evolutionist must think about nervous systems".

[34] The actual mechanism is much more sophisticated than this, but it will serve as a useful example. Alcock, 1998, pp135-142.

It is relatively easy to make a simple animal on this basis.[35] All you need are two motors and two detectors. The motors provide the 'get up and go', and the detectors act as a kind of 'guidance system', steering the animal towards its target or away from hazards.[36] In order to build animals that pursue more than one target, all natural selection need do is to stack several of these systems on top of one another, and rank them in order of their respective contribution to reproductive success.[37] Which targets? Animal behaviourists joke that most of animal behaviour is captured by the 4 Fs: feeding, fighting, fleeing, and sexual intercourse.

*Feedback*

Some adaptive control systems, including some that generate movement, can be compared to thermostats, because they attempt to maintain a particular optimal state, detect departures from that state, and act to return a system to the optimal state. For example, the mammalian thermo-regulatory system detects departures from an optimal body temperature, and prompts particular behaviours – such as shivering, sweating, or moving to a different location – designed to return the body to the optimal state. Other systems make use of things in the world in order to return to the optimal state. For example, if the digestive system detects departures from optimal blood-sugar levels, it motivates the animal to go in search of food. As Bos, Houx and Spruijt put it: "Motivational states such as hunger, thirst, and libido arise because of a difference between actual and reference values in an animal's physiological systems, and subsequent behavior – appetitive and consummatory (Craig, 1918) – is directed at eliminating this difference."[38] Or, to quote Steven Pinker: "Wanting and trying are feedback loops, like the principle behind a thermostat: they receive information about the discrepancy between a goal and the current state of the world, and then

---

[35] Brooks, 1999; Clark, 1997. Brooks claims to have developed robots that exhibit "insect-level" intelligence.
[36] See Appendix 1, Figure 1.
[37] In robotics, a "subsumption architecture" ensures that System A always trumps System B, and System B always trumps System C… and so on, thereby forestalling internal conflict should more than one system be activated at once.
[38] Bos, et al., 2002, p99.

they execute operations that tend to reduce the difference."[39]

So, the picture of an animal that emerges from evolutionary biology is of a bundle of goal-seeking, information-processing adaptations, that are capable of generating sophisticated, environmentally-contingent behaviour (including, in some cases, movement).[40] And natural selection can design more sophisticated creatures by equipping them with more adaptations, and hence more goals, and by rendering each system more *uncertain* about the world, and thus requiring them to ask more discriminating questions before coming to a decision about how to act.

It follows that, in order to explain or predict an animal's behaviour, a biologist needs a description of its (motivational) adaptations (including the conditions upon which its adaptations are dependent), and an account of current conditions. For this reason, the study of animal behaviour consists, in large part, of the attempt to come up with an accurate 'circuit diagram' of an organism's motivational systems. (Appendix 1, Figure 2 illustrates the kind of sophisticated, environmentally-contingent, adaptive system needed to control the courtship behaviour of even a relatively simple creature, in this case a newt.) Hence evolutionary theory does not explain behaviour by positing an unconscious desire to spread one's genes; natural selection is not a theory of motivation. Rather, natural selection is a theory of design that is to be used to explain the particular specifications of adaptations, including motivational systems. Behaviour is the product of these adaptations at work – whether that involves protecting against physical trauma in the case of the snail shell, maintaining efficient photosynthesis in the case of the heliotropic plant, or avoiding predators in the case of the moth.[41] Evolutionists adopt the same approach when studying human

---

[39] Pinker, 2002, p32. See also: Dawkins, 1976, p51; Rosenblueth, et al., 1943.

[40] Note that, from the perspective of evolutionary biology, there is no hard and fast distinction between what an organism 'is' and what an organism 'does'. 'Behaviour' is just biology in motion. As the anthropologists John Tooby and Irvine DeVore put it: "there is no fundamental distinction between behavioural and morphological traits". Tooby and DeVore, 1987, p191. Or as Real puts it: "Behavior can be viewed as an exceedingly plastic aspect of the organism's phenotype." Real, 1994, p6.

[41] As Tooby and DeVore have said: "The psychology of an organism consists of the total set of proximate mechanisms that control behaviour. Natural selection, acting over evolutionary time,

behaviour.[42]


**A Darwinian theory of value**

We are now in a position to see how the biologist's account of adaptations and their proximate goals can serve as a replacement for Hume's 'desire theory' of value. Hume argued that humans (and other animals) possess a range of desires or passions, and that these desires determine what they value. The Darwinian, meanwhile, argues that organisms (including humans) consist of bundles of adaptations; and the proximate goals of adaptations (including sophisticated 'desire-like' motivational systems) determine what organisms value. As the theorist Larry Arnhart puts it: "In all animal behavior . . . there are natural goals, which are standards of achievement that we can identify as 'values' or 'goods'. If we define 'value' or 'good' in relational terms as whatever satisfies a desire, then all animals have values because they all have natural desires that they strive to satisfy as they gather information about their world."[43]


This Darwinian update of Hume's theory of value demystifies 'value' – making clear the place of value in a world of facts – and suggests that we

---

shapes these mechanisms so that the behaviour of the organism correlates to some degree with its fitness. However, in the lifetime of any particular animal, it is the proximate mechanisms that actually control behavior. If these can be understood, behavior can be predicted exactly...". Tooby and DeVore, 1987, pp197-8.

[42] As with other species, the picture that emerges of human motivational systems built on these principles is not one of brute urges and drives, but rather of vast computer programs of potentially unlimited sophistication. The psychologist Steven Pinker puts it as follows: "Most intellectuals think that the human mind must somehow have escaped the evolutionary process. Evolution, they think, can fabricate only stupid instincts and fixed action patterns: a sex drive, an aggression urge, a territorial imperative, hens sitting on eggs and ducklings following hulks. Human behaviour is too subtle and flexible to be a product of evolution, they think; it must come from somewhere else – from, say, 'culture'. But if evolution equipped us not with irresistible urges and rigid reflexes, but with a neural computer, everything changes. A program is an intricate recipe of logical and statistical operations directed by comparisons, tests, branches, loops, and subroutines embedded in subroutines. . . . Human thought and behaviour, no matter how subtle and flexible, could be the product of a very complicated program, and that program may have been our endowment from natural selection." Pinker, 1997, p27.

[43] Arnhart, 1998, p21. Or, to quote Karl Popper: "All organisms are problem finders and problem solvers. And all problem solving involves evaluations and, with it, values. Only with life do problems and values enter the world." Popper, 1990, p50. And, as Martin Daly puts it: "Natural selection doesn't have goals, but it's the reason organisms do". "Purposive ('teleological') concepts are *properly* applied to organisms because they have goal-seeking processes instantiated in their structures as a result of the evolutionary process . . .". Daly, 1991, p219. I intend to use 'goal' interchangeably with end, purpose, interest, preference, want, need, desire, value, and so on.

can investigate them in the same way we would investigate any other aspect of the natural world. Given that organisms are bundles of adaptations, each with its own proximate goal, for any given organism we can draw up a list of goals, or *desiderata*, that constitutes what the organism wants, needs or *values*. It follows that in order to arrive at a list of *human* values we must come up with a list of the adaptations of which humans are composed. Thus the investigation of human values is, in the first instance, a series of problems in evolutionary biology. This was certainly Hume's view. Hume expected there to be a continuity between the motivational systems ('passions') of humans and other animals – so much so that he suggested that cross-species comparative psychology might be used to shed light on human nature:

> 'Tis usual with anatomists to join their observations and experiments on human bodies to those on beasts, and from the agreement of these experiments to derive an additional argument for any particular hypothesis. 'Tis indeed certain, that where the structure of parts in brutes is the same as in men, and the operation of those parts also the same, the causes of that operation cannot be different, and that whatever we discover to be true of the one species, may be concluded without hesitation to be certain of the other. . . . Let us, therefore, apply this method of enquiry, which is found so just and useful in reasonings concerning the body, to our present anatomy of the mind, and see what discoveries we can make by it.[44]

Subsequent research conducted along these lines has tended to vindicate Hume's comparative approach. Darwin's *The Expression of the Emotions in Man and Animals* was devoted to demonstrating the continuity of the anatomical and physiological bases of motivation, as well as the behavioural – and even facial – expression of the emotions.[45] More recent work has revealed a surprising degree of conservation across

---

[44] Hume, 1739/1985, pp375-6.
[45] Darwin, 1872/1998.

evolutionary history of the basic motivational systems.[46] All of this allows us to conclude, as does Simon Blackburn, that: "the basic emotions of anger, fear, disgust, sadness, and happiness or joy can be thought of as the upshot of fairly specific 'affect programs' . . . homologous with similar systems found in other mammals, and especially primates".[47]

So, David Hume thought that human values were the products of such passions as: love, hatred, anger, malice, envy, and amorous passion, sympathy, meekness, beneficience, charity, generosity, clemency, moderation, equity, greatness of mind, industry, perseverance, patience, activity, vigilance, application, constancy, temperance, frugality, economy, resolution, prudence, courage, due pride, and humility.[48] A Darwinian, meanwhile, would argue that human values are the product of adaptations that include the usual mammalian motivational adaptations for thermoregulation, feeding, predator-avoidance, habitat-selection, mating, and – as we shall see in subsequent chapters – various forms of cooperation.[49]

**The varieties of ethical naturalism**

Given this updated desire-theory of value, the ethical naturalist has two options, as represented by the moral philosophies of Aristotle and Hume.

The Aristotelian observes that humans have a range of natural goals, and argues that 'the good life' consists in achieving these goals. For example, Arnhart argues that because "[t]he good is the desirable", the good life consists of fulfilling these natural desires. And we can "judge societies as better or worse depending upon how well they satisfy those natural desires".[50] The Aristotelian faces a problem when it comes to manifestly 'anti-social' adaptations. What if we were to find that some people, or all people some of the time, had adaptations for murder? for rape? for

---

[46] Lawrence and Calder, 2004 discusses work that shows that the 'emotional' systems underpinning fear, disgust and anger are homologous across mammals, reptiles and birds (and, in some cases, even fish and insects).
[47] Blackburn, 1998, p126.
[48] Hume, 1739/1985, p629.
[49] See, for example, Barkow, et al., 1992; Buss, 2000b.
[50] Arnhart, 1998, p17.

domestic violence? for sexual infidelity? for cheating?[51] In such cases the Aristotelian is forced to maintain that these acts are moral on the grounds that they are the product of natural desires. Arnhart attempts to escape this conclusion by arguing that, compared to the range of goods pursued by the rest of us, psychopaths lead impoverished self-defeating lives.[52] But either it is good to act on one's natural desires or it is not. Arnhart seems to be helping himself to some super-ordinate criterion of (moral) goodness not supplied by the theory.

The Humean faces no such problem. Hume intended his general desire theory of value to be only the backdrop to the more specific investigation of *moral* value, which he saw as the products of passions that aimed at the common good of society. So the Humean observes that humans have a range of natural desires or goals, but reserves the term 'moral' for goals that promote 'the common good' – the "publick interest"; the "public good"; a "common end"; "the general interests of society"; "the good of mankind".[53] The Humean therefore distinguishes passions in general from moral passions in particular. As we shall see, Hume's *moral* passions promote the common good by solving certain recurrent problems of social life, such as certainty of paternity, coordination problems, prisoner's dilemmas, the negotiation of hierarchies, and the defence of industry and property. So although Hume thought that all moral passions were natural, he did not think – and it does not follow – that all natural passions are moral. And so murder, rape, and theft are deemed immoral by our moral sentiments because such acts are contrary to the common good, irrespective of whether they are 'natural' or the product of adaptations or not. For the remainder of this thesis I will follow Hume in arguing that *moral* passions are to be distinguished from the entire range of passions on the basis of their contribution to 'the common good'.

---

[51] Buss, 2000a; Daly and Wilson, 1988; Dugatkin, 1997b; Mealey, 1997; Thornhill and Palmer, 2000
[52] Arnhart, 1998, Ch, 8.
[53] Hume, 1739/1985, p532, p580, p590, p620, p628. In general, argued Hume, "men receive a pleasure from the view of such actions as tend to the peace of society, and an uneasiness from such as are contrary to it." (p585).

The question, then, is whether human nature comprises passions aimed at the common good, as Hume supposed. That question is the subject of the next two chapters. The remainder of this chapter reviews some of the ways that the Darwinian view of psychology differs from the Humean view, and how the Darwinian view may provide possible solutions to some of the problems inherent to Hume's psychology.

**The acquisition of knowledge**

Hume had a standard 'desire-belief' – or 'passion-reason' – model of the mind. He adopted the traditional assumption that the passions were locked in combat with reason, but went on to argue, contrary to tradition, that the passions won. Indeed, Hume famously concluded that: "Reason is, and ought to be the slave of the passions, and can never pretend to any other office than to serve and obey them."[54] The Darwinian, however, has a different ontology of mental entities. Instead of expecting there to be two different types of mental state, the Darwinian sees multiple adaptive motivational systems, each of which, as we have seen, processes information that leads it to adopt different states. If we begin with adaptation, and if we identify value with the proximate goal of the adaptation, then we might identify 'reasoning' with the information processing that adaptations perform, and 'belief' with the states that adaptations adopt. (To give a very simple example, we might say that the noctuid moth *wants* to avoid bats, that it *reasons* that if it hears a certain sound then it ought to dive, and that if it hears and dives then it *believes* that a bat is present.) Thus, desire, reason and belief are not alternative or competing mental entities; they are terms that refer to different aspects of the operation of (psychological) adaptations. For the Darwinian, "Information-processing is, and ought to be, the slave of the adaptations".[55]

---

[54] Hume, 1739/1985, p460.

[55] This alternative way of dividing up the mind leads evolutionary psychologists to characterise human decision-making as the product of "reasoning instincts", "strategic emotions", "fast and frugal heuristics" and so on. Cosmides and Tooby, 1992; Frank, 1988; Gigerenzer, et al., 1999.

This Darwinian account solves or avoids a number of problems inherent to Hume's account of how organisms acquire knowledge of the world, including how they learn from experience.[56]

Hume had a standard empiricist view of how humans and other animals acquired information and formed beliefs about their surroundings. In Book I of *The Treatise* – entitled "Of the Understanding" – Hume argues that our knowledge of the world is arrived at through the senses. He claims that sensation creates impressions, which on 'reflexion' become 'ideas'. The faculty of reason – 'the understanding' – combines 'simple' ideas together to form 'complex' ideas according to seven principles or 'relations': "resemblance, identity, relations of time and place [contiguity], proportion in quantity or number, degrees in any quality, contrariety, and causation".[57] For example, if you discover that you like apples, then you might infer that you like oranges also because apples 'resemble' oranges in some ways. In this way, the relation of 'resemblance' allows you to generalise from one situation to another. To take another example, if one has the idea of 'putting one's hand in a flame' followed closely by the sensation of 'pain', then 'the understanding' might come to 'associate' putting one's hand in a flame with the painful consequences by means of the relation of 'contiguity'. These associations are, according to Hume, strengthened or reinforced by repetition, in much the same way that a path becomes worn with use. If putting one's hand in a flame is repeatedly followed by pain, then the association between these two events becomes increasingly strong.

There are several well-known problems with this standard empiricist view of the mind. For example, there is the problem of which aspects of the world an organism should attend to. Given that there is an infinite number of things that an organism could pay attention to – an infinite number of potential sources of information, stimuli, cues – how does an organism know which 'sense data' to let in, and which to ignore as irrelevant? This problem, sometimes referred to as "the frame problem",

---

[56] The argument of this section closely follows the arguments set out by Karl Popper: Popper, 1999; Popper, 1972; Popper, 1990.
[57] Hume, 1739/1985, p117.

has led some to propose that animals "filter out" irrelevant information, or that the solution lies in equipping organisms (or robots) with sufficient "ignoring power".[58]

There is also the problem of what to do with the sense data once it is 'let in' – in particular, how to direct it to the 'right' bit of circuitry in the brain (sometimes referred to as 'the input problem'). How does information about food arrive at the food module, and not at the mate selection module? It seems as if there must be some super-ordinate mechanism responsible for recognising and directing different packets of information to their respective domains.

Once the information gets where it's going, there is also a problem with how the information is processed. Take 'resemblance', one of Hume's 'relations'. Any two objects in the universe are similar in an infinite number of ways and dissimilar in an infinite number of ways. Apples resemble oranges in being small, round, fruits, full of vitamin C, composed of atoms, to be found on Earth, and so on. But they are dissimilar in that they have different colours, different chemical compositions, are members of different species, you can't make orange pie, and so on. What counts as 'resemblance' differs depending on the purposes for which the objects are being used or the comparison is being made. There is no objective standard of 'similarity' for the content-free 'resemblance' faculty to latch on to; similarity is in the eye of the beholder. And so merely positing a 'resemblance' faculty cannot explain how inferences are made from one object to another, or from one event to another. What one needs instead is a specification of the particular mechanisms that enable an organism to make particular inferences from certain classes of objects to others.

It is also a problem for Hume's empiricist psychology to explain how

---

[58] Alcock, 1998, pp135-142; Dennett, 1998, Ch. 11. Note that this approach sets up a problem of which one of the premises is: 'organisms are sensitive to an infinite number of things in the world'. And it sees the solution of this problem as mechanism that can ignore 'infinity minus $n$' things, where $n$ is the number of things that the organism in question is sensitive to. Of course, it is not possible – for natural selection or anything else – to design a mechanism that can ignore an infinite number of things. Fortunately, natural selection never had to (see below).

organisms learn from experience. Take 'contiguity'. The common-sense view is that one learns by associating events that occur close together in time (or space). This is the basic associationist view of learning, and it remains the dominant view of learning in biology and psychology. For example, in the context of discussing the predictions that genes make about their world, Richard Dawkins comments:

> One way for genes to solve the problem of making predictions in rather unpredictable environments is to build in a capacity for learning. Hence the program may take the form of the following instructions to the survival machine: 'Here is a list of things defined as rewarding: sweet taste in the mouth, orgasm, mild temperature, smiling child. And here is a list of nasty things: various sorts of pain, nausea, empty stomach, screaming child. If you should happen to do something followed by one of the nasty things, don't do it again, but on the other hand repeat anything that is followed by one of the nice things.'[59]

This neat summary of the commonsense view of learning illustrates why such a psychology is impossible. The problem is that any one event is preceded and followed by an infinite number of other events. How is an organism supposed to know which of these infinite events are to be associated with one another? How is the survival machine supposed to know which of the infinite number of things that it has just done – scratched its nose, looked at the sky, eaten a mushroom, walked east, and so on – has resulted in the good (or bad) consequence? How is an organism supposed to know what constitutes 'an event' in the first place? In the absence of a particular hypothesis, theory or expectation of which kinds of things are likely to cause which other events, an organism has no hope of learning anything in this way. There is no such thing as a 'general purpose hypothesis', hence there can be no such thing as 'general purpose learning'. A "capacity for learning" or "association" is not an explanation; it is merely a relabelling of the phenomenon that we wish to

---

[59] Dawkins, 1976. p57.

explain.

Hume and others have often supposed that this kind of association can be saved by repetition or 'induction'. An animal may not be able to work out that eating mouldy food caused nausea the first time; but if it happens again, several times, and if the animal observes the connection repeatedly, then it has a better chance of detecting the connection, and of being more sure of that connection. But this is no help. 'Repetition' presupposes a repetition *of something*, and it is the absence of the something – that is, of a hypothesis – that causes the problem in the first place. Unless the animal has picked up on the connection in the first place, there is no possibility of it observing it again. No amount of further 'statistical data' can help if the creature has no idea what to look for.

Incidentally, Hume himself provided the logical refutation of induction. Hume noted that although inductive inferences claimed to rely on nothing but experience and observation, they tacitly assumed that the future would be like the past, which itself cannot be confirmed by experience. For example, in order to infer from past experience that placing one's hand in a flame will be painful in the future, one must also assume that 'the future will resemble the past' in the relevant ways. (You might, after all, contract a neurological disorder that numbs your hands.) But then the question is: From where did you derive the additional premise that the future would resemble the past? Either this premise is *a priori*, or else you learnt it. Admitting the existence of *a priori* knowledge invalidates the empiricist assumption that all knowledge is acquired through the senses. But arguing that the assumption was learnt (by induction) simply moves the problem one stage back, and opens an infinite regress. Hume recognised this problem, but fudged its solution. He conceded that inductive inferences are logically invalid, and cannot be relied upon, but argued that we nevertheless come to make inductive inferences through habit and custom. Hume did not take the extra step and see that what is logically impossible must also be psychologically impossible – after all, the problem is not that the premises of the inductive inference are unreliable, it is that they are unavailable. In Popper's words:

"Having cast out the logical theory of induction by repetition he [Hume] struck a bargain with common sense, meekly allowing the re-entry of induction by repetition, in the guise of a psychological theory."[60]

The Darwinian update neatly avoids all of these problems.

First, organisms do not gather information in the same way that they, for example, gather food; there are no little packets of information sitting out in the world waiting to be consumed. Rather, the acquisition of information by an organism consists in the reduction of the prior uncertainty of its adaptations. As we saw, organisms confront the world with a range of conditional adaptations that are 'uncertain' as to which state to adopt. In other words, a creature approaches the world with certain questions, and it acquires knowledge by having its questions answered. (To quote Popper: "our senses can serve us . . . only with yes-and-no answers to our own questions".[61]) The reduction of this uncertainty – the processing of information[62] – constitutes the acquisition of knowledge. And, as we have seen, natural selection designs an adaptation to be sensitive only to those aspects of the world that reduce uncertainty about which state to adopt (and ultimately, how to survive and reproduce). So, what constitutes 'sense data' in the first place is a function of the conditions upon which the adaptation is dependent. And as a result, what constitutes 'sense data' for one creature is not what constitutes sense data for another. Bees are sensitive to ultraviolet light in ways that bats are not; bats are sensitive to sonar in ways that lobsters are not; lobsters are sensitive to the earth's magnetic field in ways that bees are not; and so on. Each species, including humans, inhabits its own particular sensory world that is different from the worlds of all other species.[63]

Hence organisms do not face the problem of having to 'ignore' infinite amounts of information. They are simply sensitive to a particular subset

---

[60] Popper, 1963, pp45-6.
[61] Popper, 1990, pp46-47.
[62] Dawkins, 1998a; Dawkins and Dawkins, 1973.
[63] Uexküll, 1934/1957.

of things in the world and not to any other of the infinite alternative subsets. As Popper puts it: "Classic epistemology which takes our sense perceptions as 'given', as the 'data' from which our theories have to be constructed by some process of induction, can only be described as pre-Darwinian. It fails to take account of the fact that the alleged data are in fact adaptive reactions."[64]

Second, it follows from this view of knowledge acquisition that there are no 'sensations' or 'ideas' floating around in the foyer of the mind, waiting for 'the understanding' to organise, redirect or combine them. The information that 'sense data' refers to is the particular state adopted by a particular adaptation. The structure of this information, and the uses to which it will be put, is already determined by the possible states of the adaptation. For example, the noctuid moth's nervous system does not first encode a representation of a bat and then face the problem of shunting this representation to the appropriate bat-avoiding mechanism. This portion of the moth's nervous system just *is* a bat-avoiding mechanism, and the presence of the bat is 'represented' by the mechanism adopting a particular state (in this case, a dive). This 'representation' of the bat – which takes the form not of a 'picture', but of an activated circuit – is then available for other circuits to latch on to. Of course, it is no easy task to explain how this model of the mind 'scales up' to accommodate human thought. But we can at least we can be confident that we are starting in the right place with the right problems; and given that natural selection managed to solve these problems, we have good reason to hope that where evolution has led science shall surely follow.[65]

Third, the Darwinian account of psychology avoids the problems of induction by not relying on induction at all. Hence, learning is explained with reference to the prior expectations or theories that the organism is equipped with. In the simplest case we might suppose that 'learning' involves the execution of second-, third-, and nth-order decision-rules. A *first-order* conditional can be described as "If X, then Y". For example, a rat

---

[64] Popper, 1972, p146, p145.
[65] Cosmides and Tooby, 2000.

might operate according to the rule: "If you smell cheese, then move towards it and eat it". A *second-order* conditional would be "If A, then (If X, then Z), else (If X, then Y)". For example, a rat might operate according to the rule: "If a particular cheese has made you ill, then (If you smell the cheese, then ignore it), else (If you smell the cheese, then move towards it and eat it)".[66] Thus, the rat 'learns' not through the accumulation of (potentially fatal) experiences, but through the application or a prior 'theory' of what kinds of things are likely to induce nausea.[67] To quote the psychologists Garcia and Koelling: "The hypothesis of the sick rat, as for many of us under similar circumstances, would be 'it must have been something I ate'."[68] So, organisms do not learn through the gradual accumulation of 'data'; instead they "jump to a conclusion" and then put that conclusion to the test.[69] And more sophisticated creatures can generate more sophisticated hypotheses, and thereby discover more about their worlds. As Popper put it: "all knowledge is a priori, genetically a priori, in its content. For all knowledge is hypothetical or conjectural: it is our hypothesis. Only the elimination of hypotheses is a posteriori, the clash between hypotheses and reality. In this alone consists the empirical content of our knowledge. And it is enough to enable us to learn from experience; enough for us to be empiricists"[70] Or, to quote Noam Chomsky: "the general form of a system of knowledge is fixed in advance as a disposition of the mind, and . . . the function of experience is to cause this general schematic structure to be realized and more fully differentiated."[71]

---

[66] This example is taken from Alcock, 1998, pp102-3. It is representative of a large ethological literature that developed in reaction to behaviourist and Skinnerian views of learning as conditioning and reinforcement. See, for example: Breland and Breland, 1961; Lorenz, 1966a.

[67] We may say that "an organism 'learns from experience' only if its dispositions to react change in the course of time, and if we have reason to assume that these changes do not depend merely on innate [developmental] changes in the state of the organism but also on the changing state of its external environment." Popper, 1972, p343.

[68] Garcia and Koelling, 1966, p124, quoted in Kamil, 1994, p29.

[69] 'Jumping to a conclusion' also explains how 'one-shot learning' is possible; that is, how animals can learn, for example, to avoid a foodstuff after only one unpleasant experience, instead of requiring many such experiences.

[70] Popper, 1999, p47.

[71] Chomsky, 1965, pp51-2. Chomsky continues: "It is a matter of no concern and of only historical interest that such a hypothesis will evidently not satisfy the preconceptions about learning that derive from centuries of empiricist doctrine. These preconceptions are not only quite implausible, to begin with, but are without factual support and are hardly consistent with what little is known about how animals or humans construct a 'theory of the external world'." (p58).

Thus we escape Hume's problem of induction by accepting that expectations about the future – that 'the future will resemble the past' in the relevant ways – are *a priori*. Hume was reluctant to take this step because at the time of writing, God was the principal theory as to the origin of *a priori* knowledge, and this was the view that Hume was trying to get away from. It was left to Kant to develop the notion of *a priori* knowledge; and to Darwin to wrest *a priori* knowledge from God and from transcendental realms, and instead provide an entirely naturalistic explanation of the form and content of innate knowledge. Under this view, "the regularities we try to impose are *psychologically a priori*, but there is not the slightest reason to assume that they are *a priori valid*, as Kant thought".[72] It comes as no surprise, therefore, to find that prominent evolutionary psychologists see themselves as providing "an evolutionary Kantian position" with regard to innate knowledge.[73]

It follows that there will have to be many different mechanisms for many different kinds of knowledge acquisition. For example, in the domain of food and foraging, an organism may 'discover' that there is fruit in a particular valley; it may 'learn from experience' that green fruit makes it ill; it may 'develop' a particular metabolism in response to shortages of food during a critical period of development; it may 'acquire' certain food preferences by attending to the smell of its mother; it may learn how to wash sweet potatoes by 'imitating' others; and so on. And there will be equivalent mechanisms for other domains, such fighting, fleeing, mating, and cooperating. There is no single mechanism that could accomplish these diverse feats; and these different processes are not sufficiently distinguished or suitably explained with reference to 'a capacity for learning'.[74] Hence we arrive at the conclusion – which seems counter-intuitive on the standard view – that rich and diverse *a priori* knowledge is necessary for learning to occur. By making explicit the content of this evolved *a priori* knowledge, evolutionary psychologists are attempting to explain how learning is possible; they not presenting prior knowledge as

---

[72] Popper, 1972, p24.
[73] Tooby and Cosmides, 1992, p70.
[74] See: Gallistel, 1999.

an alternative to 'learning' or other forms of knowledge acquisition.

Incidentally, in *Conjectures and Refutations* Popper used the notion of building "an induction machine" to illustrate why it is impossible in theory for natural selection, or anything else, to build an organism capable of induction:

> To sum up this logical criticism of Hume's psychology of induction we may consider the idea of building an induction machine. Placed in a simplified 'world' (for example, one of sequences of coloured counters) such a machine may through repetition, 'learn', or even 'formulate', laws of succession which hold in its 'world'. If such a machine can be constructed (and I have no doubt that it can) then, it might be argued, my theory must be wrong; for if a machine is capable of performing inductions on the basis of repetition, there can be no logical reasons preventing us from doing the same. The argument sounds convincing, but it is mistaken. In constructing an induction machine we, the architects of the machine, must decide *a priori* what constitutes its 'world'; what things are to be taken as similar or equal; and what *kinds* of 'laws' we wish the machine to be able to 'discover' in its 'world'. In other words we must build into the machine a framework determining what is relevant or interesting in its world: the machine will have its 'inborn' selection principles. The problems of similarity will have been solved for it by its makers who thus have interpreted the 'world' for the machine.[75]

Unfortunately, not everyone heeded Popper's warning. Indeed, in some ways, the early history of artificial intelligence can be seen as the attempt to build just such "an induction machine". The guiding assumption has been that humans learn by induction (or 'association'), and that in order to create a machine that recreates the scope and power of human-like

---

[75] Popper, 1963, p48.

intelligence, one must build a machine capable of learning by induction (or 'recognising patterns', or having its 'connections reinforced'). The suggestion that one must equip an artificially-intelligent machine with 'prior' knowledge was met with suspicion, hostility, and the vague sense that this would be 'cheating'. Not surprisingly, artificial-intelligence engineers encountered numerous problems in the course of attempting to build machines that lacked any "framework determining what is relevant or interesting in its world", and these problems came to be known as 'frame problems'.[76] Fortunately, in the 1980s, some AI-engineers and roboteers began to turn their backs on what they dubbed 'good old fashioned AI', and began to develop robots inspired by evolutionary biology. In this way, they managed to avoid 'frame' and related problems altogether.[77]

**Conclusion**

This chapter has provided a Darwinian update of Hume's account of psychology, and Hume's naturalistic account of value. Evolutionary theory sees organisms as bundles of adaptations designed by natural selection to solve the problems of survival and reproduction that faced their ancestors. In mobile organisms such as animals these adaptations include 'motivational systems' that move the organism towards adaptive targets (such as food and mates) and away from adaptive hazards (such as predators). The information-processing performed in pursuit of these goals constitute 'reason'; and the different states that adaptations can adopt constitute 'belief'. The proximate goals of these adaptations constitute *desiderata*, or what that organism values.

Having made explicit the Darwinian theory of value, the next task is to provide a Darwinian update of Hume's theory of *moral* value. This will require demonstrating that Darwinian theory can explain the existence of 'passions aimed at the common good'. In Chapter 3 I show that: contrary to the Hobbesian caricature of 'life in the state of nature' evolutionary

---

[76] Lormand, 1999.

[77] Brooks, 1999; Clark, 1997. For a discussion of how natural selection avoids 'the frame problem' see: Sifferd, 2002.

theory leads us to expect that organisms will be designed to be social, cooperative and even altruistic under certain circumstances. And in Chapter 4 I show that such 'adaptations for cooperation' can be found in humans. In Chapter 5 I shall show that this Humean-Darwinian account of the moral sentiments is consistent with a wide range of traditional views of the nature and content of moral values.

## Chapter 3 – Life in the state of nature

> *The position I have always adopted is that much of animal nature is indeed altruistic, cooperative, and even attended by benevolent subjective emotions, but that this follows from, rather than contradicts, selfishness at the genetic level. Animals are sometimes nice and sometimes nasty, since either can suit the self-interest of genes at different times. That is precisely the reason for speaking of 'the selfish gene' rather than, say, 'the selfish chimpanzee'.*
>
> – Richard Dawkins[78]

**Introduction**

According to Thomas Hobbes human nature is entirely selfish; life in the state of nature is 'nasty brutish and short'; and morality and political society are artificial inventions that must be imposed on humans by an external authority.[79] David Hume took issue with Hobbes' account and argued that human nature was not entirely selfish, but comprised a range of passions that are aimed at the common good. Hume called these 'the moral passions', and hence argued that morality and to some extent political society are not artificial inventions, but that they are the products of natural human sentiments.

Hume relied on introspection, everyday observations, and anthropological evidence that seemed to suggest that sociality and morality were human universals, to make the case for natural human moral passions. But, as we saw in the previous chapter, Hume also suggested that cross-species comparisons – "the correspondence of passions in men and animals"[80] – might be used to establish the naturalness of certain human passions. This chapter adopts just such a comparative approach, and reviews the literature presenting the theory of -- and evidence for -- the evolution of adaptations for cooperation in

---

[78] Dawkins, 1998b, p212.
[79] Hobbes, 1651/1958.
[80] Hume, 1739/1985, p376.

nonhuman species. Recent developments in evolutionary theory lead us to expect that organisms will be social, cooperative and even altruistic under certain conditions. For example, evolutionary theory can explain why animals are motivated to care for family members, to coordinate to mutual advantage, to punish free-riders, to settle disputes peacefully, and to respect property. The finding that such adaptations are widespread vindicates Hume's comparative approach, and bolsters his optimistic assessment of human nature. And having familiarised ourselves with the logic and structure of 'adaptations for cooperation' in general in this chapter, we will be in a better position to identify examples of such adaptation in humans in the next chapter, and thereby give Hume's account of human moral passions a Darwinian update.


**The 'state of nature'**

Some standard interpretations of Darwin's theory of evolution by natural selection seem to support the Hobbesian view that life in the state of nature is 'nasty, brutish and short'. Thomas Henry Huxley wrote that: "From the point of view of the moralist, the animal world is on about the same level as the gladiator's show. . . . the weakest and the stupidest went to the wall, while the toughest and the shrewdest, those who were best fitted to cope with their circumstances, but not the best in any other way, survived. Life was a continuous free fight, and . . . a war of each against all was the normal state of existence." The conclusion that Huxley drew was that: "the ethical progress of society depends, not on imitating the cosmic process [evolution], still less in running away from it, but in combating it". But against this view of the natural world has always stood the well-documented existence of extensive cooperation among animals.[81] As Petr Kropotkin put it: "The ants and the termites have renounced the 'Hobbesian War' and they are better for it". Again, the moral was clear: "Don't compete! . . . [C]ombine – practice mutual aid! That is what Nature teaches us."[82] The tension between what the theory of natural selection seemed to suggest and what the evidence clearly showed – between

---

[81] Allee, 1931.
[82] Kropotkin, 1902.

irresistible force and immovable object – created 'the problem of altruism', which EO Wilson called "the central theoretical problem of sociobiology".[83]

The advent of 'selfish gene' theory in the 1960s and 70s seemed to some to support Hobbes's and Huxley's side of the argument, placing selfishness at the very core of animal and human nature.[84] But, as the quote at the beginning of the chapter points out, that is not the conclusion to draw from selfish gene theory; in fact, the theory provides a way of reconciling the two apparently incompatible positions. This is principally because 'selfish gene theory' is not a theory of motivation; it is a theory of design. And it is simply the case that the differential selection of genes can produce phenotypic effects that we characterise as cooperative and altruistic. The next section provides a brief summary of the logic of cooperation. This is followed by sections that review the evolution of adaptations for four distinct types of cooperation: kin altruism, coordination to mutual advantage, reciprocity and conflict resolution.

**The evolution of cooperation**

The very first replicators to emerge on Earth were lone agents. But, as the population size increased, their selective environment soon came to be filled with other replicators and their effects. We might say that social life began when replicators constituted a significant part of the selection pressures on other replicators.

The caricature is that relations between these replicators are necessarily competitive, but this is not the case. Interactions in which one replicator advances at the expense of another are only one of four logically possible interactions. A selfish replicator can promote its replication at the expense of another (+/-); a cooperative replicator can promote its replication whilst also promoting the replication of another (+/+); an altruistic replicator can promote the replication of another at a cost to itself (-/+);

---

[83] Wilson, 1975.
[84] Dawkins, 1976.

and a spiteful replicator can harm its replication whilst also harming the replication of another (-/-). But because replicators that are, on balance, altruistic or spiteful remove themselves from the gene pool, we are left with replicators that have 'selfish' and 'cooperative' phenotypic effects.[85] Hence evolutionary theory expects genes to be 'self-interested', in the sense of promoting their own replication; but it does not expect all genes to be 'selfish' in the sense of always promoting their replication at the expense of others'.[86]

What opportunities does nature provide for the selection of cooperative replicators? Evolutionary theorists, making use of game theory, have identified at least four categories of cooperative interactions: kin altruism, coordination to mutual advantage (coordination), reciprocity (prisoner's dilemma) and conflict resolution (chicken).[87] Kin altruism refers to situations in which a gene helps copies of itself that happen to reside in other individuals. Coordination to mutual advantage (mutualism) refers to situations in which genes benefit from working together, and have solved various problems of spatial or temporal coordination in order to do so. Reciprocity refers to a particular solution to the problems of delayed or uncertain mutualism. And conflict resolution refers to the means by which genes avoid some of the costs of conflict and dispute. Each of these opportunities presents different pathways to cooperation, and each of these paths contains different obstacles. The theory behind these four types of cooperation, and examples of adaptations in action, are given below.

---

[85] Replicators that are decreasing in frequency are not targets of the cumulative selection that is necessary for adaptation. So, while an adaptation might suddenly find itself in an environment in which it reliably delivers benefits to others, there can be no adaptations that have been specifically designed for this kind of altruism.

[86] This distinction between selfish and cooperative interactions maps onto the distinction in game theory between zero-sum and nonzero-sum games. In game theory, interactions in which one player's gain is another's loss are called zero-sum or constant-sum games. Interactions in which one player's gain can be another's gain are called nonzero-sum or variable-sum games. 'Cooperative' is usually reserved to describe games in which players can communicate; but I will use the term in the more colloquial sense here. Note also that I am following the convention in evolutionary biology of using game theory to model the effects of natural selection on a population of genes, and *not* to model the decision-processes of individual organisms. According to the biological use of game theory, natural selection 'chooses' the best available phenotypic 'decision-rules', and individual decision-making is seen as the execution of these rules: Dawkins, 1980.

[87] See Appendix 1. For example, see Nunn and Lewis, 2001.

Note that I am using game theory to illustrate 'ideal types' of cooperation, and that I have chosen examples from the animal behaviour literature that tend to exemplify these types of cooperation. In the real world, however, adaptive problems are not always as clearly defined as this, and there is no guarantee that all adaptations for cooperation will fall neatly into one or other of these four categories. For example, in order to successfully hunt in packs, a creature may need to coordinate on an equilibrium in a repeated prisoner's dilemma with members of its family – hence the adaptation (or adaptations) responsible for such a behavioural trait may be designed to solve multiple, overlapping problems. Note also that I am using 'toy games' merely as analytic categories to distinguish between different types of cooperation; I am not proposing any quantitative accounts of actual selection pressures – filling in the relevant variables with ecologically-valid measures of costs and benefits – in order to arrive at detailed predictions about, for example, the precise amount of time members of a particular species will devote to one form of cooperation or another.

Finally, note that a complete account of the origin and evolution of cooperation would begin with 'naked replicators', proceed through the emergence of chromosomes and single cells, and the emergence of multi-cellular organisms, and it would include a full account of the social lives of microorganisms and plants.[88] However, my interest here is not so much with adaptations for intra-individual cooperation as with adaptations for inter-individual cooperation – that is, cooperation between replicators housed in different individuals. Hence this chapter will focus on adaptations for cooperation between multicellular organisms, usually animals.

---

[88] For intra-individual cooperation between genes, see: Haig, 2003; Maynard Smith and Szathmáry, 1995; Ridley, 2000; Skyrms, 1996. Microorganisms, meanwhile, "demonstrate all the hallmarks of a complex and coordinated social life" including "cooperation, division of labour, eusociality, cheating, complex communication networks, high genetic relatedness and recognition of kin". Crespi, 2001, p178. Trivers discusses the evolution of chemical 'warning calls' in plants: Trivers, 1985, pp60-62. For more on the social behaviour of plants see: Charnov, 1984.

**Kin altruism**

One of the principal sources of common interest between organisms is that they share genes. Under such circumstances, a gene that bestows a benefit on another organism will spread if the benefit is greater than the cost, modified by the probability that they share the relevant gene.[89]

As we saw in the previous chapter, natural selection can be characterised as "the process whereby replicators out-propagate each other" by virtue of the effects that they have on the world.[90] A curious but inevitable feature of this process is that replicators can have effects on copies of themselves. Hence the frequency of a particular *type* of replicator in a gene pool will be a function not only of the replicator's effects on its own replication, but also of its effects on the replication of other copies of itself. And this applies even if the replicas happen to reside in other individuals. As Dawkins has said:

> What is the selfish gene? It is not just one single physical bit of DNA. . . it is *all replicas* of a particular bit of DNA, distributed throughout the world. . . . [W]hat is a single selfish gene trying to do? It is trying to get more numerous in the gene pool. Basically it does this by helping to program the bodies in which it finds itself to survive and reproduce. But now we are emphasizing that 'it' is a distributed agency, existing in many different individuals at once. . . . [A] gene might be able to assist *replicas* of itself that are sitting in other bodies.[91]

So, genes that benefit other copies of themselves will be selected if they benefit other replicas more than they cost themselves (where both benefit and cost are measured in terms of increase or decrease in frequency). The

[89] Hamilton, 1964.
[90] Dawkins, 1982, p133.
[91] Dawkins, 1976, p88. David Haig provides an alternative way of making the point that 'selfish genes' are distributed. Haig distinguishes between the *material* gene and the *informational* gene. The *material* gene is the stretch of DNA, the *informational* gene is the information that this stretch of DNA carries. It follows that there can be multiple *material* copies of the same *informational* gene; indeed, Haig suggests that the material gene can be seen as a *vehicle* for the informational gene. Haig, 1997.

result is that genes will be selected to bring about an efficient allocation of reproductive effort across replicas of itself. Hence, we should expect to find adaptive phenotypic effects in one organism that have been designed to benefit replicas in another, under certain circumstances. Which others? And under what circumstances?

Individuals with whom one is likely to share genes are called one's genetic relatives or family. Evolutionary theory expects a gene in one individual to benefit a family member up to the point where the benefits, modified by the probability of sharing that gene, are greater than the costs (Br>C). So, a gene in one individual has, on average, a 50% chance of finding a replica in its parents, its offspring or its full siblings; a 25% chance of finding a replica in grandparents, grandchildren, aunts and uncles, nieces and nephews; a 12.5% chance of finding a replica in first cousins; and so on.

Under what circumstances might it pay for a gene to stop promoting its own replication and instead promote the replication of replicas; to stop using up its own resources, but pass them on to others? Clearly, if genes are similarly situated, then there is no gain to transferring resources between them. But we might expect kin altruism to be selected when they are differently situated, for example, when one organism is experiencing diminishing marginal returns on consumption of a resource.

One particular source of asymmetry occurs between parents and their offspring. When designing the 'life history strategy' of an organism, natural selection "discounts the future" in line with the risk of death present in the organism's ecology. Thus, organisms are not built to last indefinitely, but exhibit 'planned obsolescence'. The result is that, as Trivers observes, "Since the reproductive value of a sexually mature organism declines with age, the benefit to him or her of a typical altruistic act also decreases, as does the cost of a typical act he or she performs."[92] Hence a transfer of resources from the older organism to the younger can represent a net gain to the genes involved.

---

[92] Trivers, 1981, p12.

Hence parental investment in offspring is perhaps the most familiar product of kin selection – from water bugs carrying eggs on their back, to birds feeding chicks in the nest, to suckling in mammals.[93] In sexually-reproducing species, equal parental investment by both sexes is unstable, and one sex usually specialises in 'parenting effort', while the other specialises in 'mating effort'.[94] Early on in the evolution of sexual reproduction, these differences were manifest only in different sized gametes: one sex (females) produced a few large immobile resource-laden gametes, while the other sex (males) produced numerous, small, mobile gametes. But over evolutionary time, these initial differences have snowballed to produce the differences in parental investment characteristic of all sexually-reproducing species.

Mammals are characterised by an especially high degree of parental, and almost always maternal, investment.[95] In female mammals, well-documented differences in genes, and subsequently hormones, give rise to adaptations for gestation and lactation, orchestrated by the hormones prolactin and oxytocin.[96] Commenting on the neural substrate of the psychology of maternal investment, the neurobiologist Jaak Panksepp notes: "It was a momentous passage in biological evolution when neural circuits emerged in the brain that encouraged animals to take care of each other. The fact that these urges evolved from pre-existing sexual circuits should come as no surprise to those who appreciate the tinkering ways of evolution. Thus, one of the key neuromodulators that helps sustain female sexuality – namely, oxytocin – is also a key player in the initiation of maternal urges in first-time mothers."[97]

Altruism between siblings is rarer than between parents and offspring, largely because there is less asymmetry between potential donors and recipients. Nevertheless, there is still plenty of it. 'Helpers at the nest' –

---

[93] Clutton-Brock, 1991; Trivers, 1972. On 'life history strategy', see: Charnov, 1993.
[94] Dawkins, 1976, Ch. 9; Trivers, 1972.
[95] But see: Buchan, et al., 2003.
[96] Hrdy, 1999.
[97] Panksepp, 2000, p148.

where mature offspring stick around to help their parents raise their young – have been reported in over 100 species of bird, as well as numerous examples in mammals, including our closest relatives.[98]

Cooperation among siblings reaches its most extreme form in the case of the social or eusocial insects, such as ants, bees and termites, which practise a division of labour that includes a reproductive caste and sterile workers. William Hamilton showed that this extreme division of labour was facilitated by a quirk of these insects' genetic system. In a multicellular organism, the probability that a neighbouring cell shares a copy of a gene is close to 1; in mammals, the chance that a full sibling will share a gene is, on average, 0.5. In certain species of social insect the probability that a worker shares a gene with the Queen is around 0.75.[99] In such species, genes in worker bees can produce sterility and self-sacrifice because by doing so they promote the replication of copies of themselves, in the Queen bee, more efficiently than if they tried to go it alone.

In short, kin selection is an inevitable result of how natural selection works on genes. Adaptations for kin altruism are widespread and, as a result, family groups are a ubiquitous feature of the social lives of animals. But kinship is only one source of common interest. Other sources of common interest flow from the benefits that animals derive from 'working together' in various ways. Cooperation between unrelated individuals raises two main problems that animals have to overcome. First, they have to identify and coordinate their behaviour with other members of the 'team'; second, they have to defend cooperative schemes against free-riders. Let's look at each of these problems, and how they are solved, in turn.

---

[98] Goodall, 1994; Trivers, 1985, pp184-198.

[99] This difference between the relatedness of cells in a body and bees in a hive means that there is greater scope for conflicts of interest among bees. Under certain conditions female worker bees lay their own eggs, an activity that is countered by the 'policing' activities of other bees that destroy worker-laid eggs. Maynard Smith and Szathmáry, 1995, pp264-265.

**Coordination to mutual advantage**

An organism might benefit another if by doing so it receives a direct benefit in return.[100] Typical examples of this kind of mutually-beneficial exchange, or 'mutualism', include herding and cooperative hunting. Individuals may benefit others as a mere by-product of their ordinary activity – as in the case of penguins huddling together for warmth, or when birds eat parasites from the backs of rhinos – or as a product of adaptations specially designed for mutualism. This section will focus on the second kind of benefit – those bestowed by specially-designed adaptations for mutualism – and on the problems that animals face in trying to work together. In particular, it will focus on cases where individuals are *uncertain* about how to coordinate to mutual advantage – about what to do, when to do it, or where to do it – and where they reduce their uncertainty by attending to the behaviour of others. For example, an individual fleeing from a predator wants to remain in the herd in order to reduce his exposure to danger; he will reduce his uncertainty about which way to run by attending to the behaviour of other members of the herd. A prairie dog wants to flee from a predator at the same time as other prairie dogs, but might be uncertain as to when that will be; it will reduce this uncertainty by attending to the warning calls made by others. A hunting lioness wants to adopt the most effective position in the hunt; she will reduce her uncertainty about whether to adopt the position of 'wing' or 'centre' by attending to the positions adopted by other members of the hunt. Situations of this kind are modelled in game theory as 'coordination problems'.[101]

Adaptations for coordination to mutual advantage have received relatively little attention in the evolution of cooperation literature (as compared to the attention given to kin altruism, reciprocity and costly signalling). This is largely because coordination to mutual advantage does not involve any apparent altruism, and its explanation requires no major theoretical advance. As the animal behaviourist John Alcock has

---

[100] As Dawkins puts it: "If animals live together in groups their genes must get more out of the association than they put in." Dawkins, 1976, p166.
[101] See Appendix 2, Table 2.

said: "When both helper and recipient enjoy reproductive gains from their interaction, they have engaged in mutualism or cooperation, which requires no special evolutionary explanation."[102] The result of this relative neglect of coordination has been that the evolutionary theory of coordination has not been richly developed; the scope of evolutionary explanations of social and cooperative behaviour has been unnecessarily restricted; and adaptations for signalling and for 'social intelligence' (amongst others) have not been recognised as adaptations for cooperation. This section attempts to correct these oversights.

In the simplest case we might suppose that an adaptation for coordination to mutual advantage consists of a conditional decision rule in which the 'condition' is some aspects of another's behaviour. Such strategies could be elaborated in numerous ways, in terms of the number of decision points, and the types of cues that it relies upon. In order to be more specific, and to impose some order on the nascent literature on the evolution of coordination I have adopted the theoretical framework proposed by Thomas Schelling and, especially, David Lewis. Schelling and Lewis mention several different ways in which uncertainty about others' behaviour might be reduced and hence coordination problems might be solved: salient focal points, prediction, communication, shared expectations, agreements, precedent, and leadership.[103] *Salience* refers to a feature of an equilibrium that makes it more 'noticeable' to the players (also referred to as a "focal point").[104] Players coordinate their actions by choosing the equilibrium that they expect to be salient to others. Individuals can make *predictions* about the behaviour of others based on a familiarity with their goals and the information that they have available to them. *Communication* can help to facilitate or confirm those predictions, leading to the generation of *mutual* or *shared expectations*.[105] (When communication is used to generate shared expectations about the future, it becomes an *agreement*.) *Precedent* is a source of salience that "achieves coordination by means of shared acquaintance with the achievement of

---

[102] Alcock, 1998, p562.
[103] Lewis, 1969; Schelling, 1960.
[104] Schelling, 1960, pp68-70.
[105] Lewis, 1969, p27.

coordination in a single past case exactly like our present coordination problem".[106] *Leadership* refers to situations in which players converge on the equilibrium chosen by a salient individual. In the next few sections I provide examples of each of these types of coordination from the animal behaviour literature.

*Salience*

To illustrate the notion of coordinating by means of a salient focal point, Thomas Schelling gives the example of a game in which two parachutists are dropped somewhere in a territory and are required to meet. Neither knows the location of the other, and they cannot communicate in any way. The territory includes one bridge and several houses. Most players choose to go to the bridge, because they expect that that is where the other player will choose to go also. There is nothing about the bridge *per se* that makes it a focal point; in versions of the game where there are several bridges and only one house, people chose to meet at the house. It is merely that the uniqueness of the bridge serves to "precipitate" or "crystallize" the players' expectations: "to fill the vacuum of indeterminacy that otherwise exists". Schelling observes that in order for something to act as a focal point, it must enjoy "prominence, uniqueness, simplicity, precedent, or some other rationale that makes it qualitatively different from the continuum of possible alternatives."[107]

Midges face a similar coordination problem, and solve it in a similar way.[108] Like many sexually-reproducing species, male and female midges face a potential coordination problem when it comes to finding a mate. They would both benefit from arriving at the same place at the same time; but the world is big and midges are small. Males, it seems, solve this problem by hovering above a "conspicuous object" such as a post, or a human head. So, males operate according to the rule: "If you want a mate, then hover above the nearest conspicuous object" (where 'conspicuous' would have some more precise specification in terms of the midge's

---

[106] Lewis, 1969, p41.
[107] Schelling, 1960, pp68-70.
[108] This discussion is based on an observation by: Hamilton, 1971, p251.

visual system). Females, it would seem, have been designed by natural selection to 'take it for granted' that males will behave in this way, and so operate according to a similar rule. In this case, natural selection has seen to it that flying to the highest conspicuous object has become the only stable strategy in the population, although it could have been otherwise (for example, 'hover above shiny objects').

Midges resolve their uncertainty about others' behaviour by attending to certain features of the world. Coordination problems are more difficult to solve if you are attempting to coordinate with a moving target.

*Prediction, expectation and anticipation*

Members of coordinating groups are often uncertain of the position or activities of other members of their team, and must resolve this uncertainty, by attending to the behaviour of others, in order to successfully coordinate. Theory requires that in order to coordinate to mutual advantage, individuals must have adaptations that motivate and enable them to do so; that is, they must have adaptations that embody theories or expectations about how to behave in order to successfully coordinate, and hence react in specific ways to specific aspects of their teammates' behaviour. However, the literature on coordination to mutual advantage has yet to fully develop this aspect of the theory. Researchers often talk as if animals can merely 'learn' how to behave by 'observing' others, without ever specifying the mechanisms that make this possible.[109] There is, however, no doubt about the fact *that* animals coordinate their behaviour, in ways that cry out for explanation by suitably-developed evolutionary theories of coordination.

For example, individuals benefit from being part of a herd, a flock or a school because other animals can act as 'cover' against predators. (Members of such aggregations also benefit from there being multiple sets of sensory systems on the lookout for predators.)[110] The tendency for

---

[109] Boinski and Garber, 2000.
[110] 'Sensory integration' has been observed in urchins, mysid shrimp, birds and fish, Thompson's gazelle, and dolphins. Norris and Schilt, 1988, p157.

animals in such aggregations to attempt to adopt central positions or to otherwise reduce their exposure to predators gives rise to what are called 'selfish herds' and 'polarized schools'.[111] Polarized schooling has evolved in a wide variety of taxa, including "echinoderms, schooling mollusks, arthropods, fish, amphibians, diving and flocking birds, herding ungulates, and dolphins".[112] In other species, reactions to predators may involve a coordinated division of labour, as in the case of the defensive formations formed by musk ox under attack by wolves, or the different sentry positions adopted by prairie dogs under attack by different types of predator (see below).[113]

Some predators have also taken advantage of the benefits of working together. Species that practise cooperative hunts include: bacteria, spiders, tuna, gulls, hawks, wolves and other wild dogs, lions and other big cats, dolphins, whales, and some primates – some 22 vertebrate species in total.[114] Some species operate a *division of labour* when hunting. For example, yellowtail tuna form teams to hunt mackerel. About six Yellowtails swim in a line along the seaward side of a shoal of mackerel. They then adopt a crescent formation and herd a small group of the mackerel away from the main shoal, and towards the shore. There, in shallower water, and with one exit blocked by the shore, they form a semi-circle around the prey. One of the Yellowtails then darts into the shoal, forcing the mackerel to flee into the mouths of the other members of the hunting party. To take another example, Harris hawks have at least three different strategies for cooperative hunting; one involves a surprise pounce from several different angles, another involves one hawk flushing the prey out from cover while the other two wait to ambush it, and a third involves a kind of relay attack. Group hunts among coyotes, wolves, spotted hyenas, lions, bottlenose dolphins, killer whales and chimpanzees

---

[111] Hamilton, 1971.
[112] Norris and Schilt, 1988, p151.
[113] Mech, 1970; Slobodchikoff, 2002.
[114] Examples of cooperative hunts can be found in: Dugatkin, 1997a. See also: Boesch, 1994; Lenski and Velicer, 2000; Smolker, 2000. The coordinated action of several predators can be enough to overcome the defences (even the collective defences) of their prey. For example, by approaching a school of fish from different angles, predators can send 'contradictory messages' through a school, creating a 'zone of confusion' at the point where the messages meet – long enough to individuate those fish and make them vulnerable.

have also been observed to involve "intelligent coordination and division of labour".[115] In lions, chimpanzees and some wild dogs, participants "regularly assume particular hunting positions relative to the prey and to their companions".[116]

In other situations, it is not enough merely to *react* to what others are doing; animals must make predictions about future behaviour in order to anticipate it. Making such predictions merely involves natural selection taking existing expectations about the behaviour of others, and extrapolating them into the future. So, during the course of the arms race between predator and prey, or during the more benign co-evolution between cooperative partners, we might expect selection to favour more sophisticated hypotheses about the behaviour of others. We should expect A's hypotheses about B to reflect the fact that B itself is a collection of hypotheses – of conditional adaptations that adopt different states under different conditions. Hence, we might expect A's theory of B will come to reflect B's various adaptive goals (desires), and be sensitive to aspects of the world that indicate what state B might be in (beliefs). (For example, "If there's a loud noise, I can assume that he's heard it" or "If he *looks* at something, then he knows that it is there".) Eventually, this system of theories and expectations might develop to the point where one would want to call it a 'model' or 'simulation' of the other organism.[117] And perhaps further innovations might see one animal's model of another including the second animal's model of the first, and so on. (Of course, what one does with this theory depends on the interests of the parties involved. A predator may use a predictive theory of its prey to intercept and devour it; one organism may use its theory to frustrate the behaviour of a rival or combatant; a teammate may use it to generate mutual benefits.)

---

[115] Holekamp and Engh, 2002, p372.

[116] Holekamp, et al., 2000, p618. The researchers go on to say that differences in position may stem from "morphological differences among individual hunters that result in differential hunting success". Alternatively, the organisms may be using arbitrary morphological differences to break the symmetry that might otherwise hamper attempts to coordinate upon an efficient division of labour.

[117] Whiten, 1996. Of course, one organism's model of another need not be an exact or complete match of the other's motivational systems; it need only model what is adaptively relevant in as efficient a manner as necessary.

This would be the ideal jumping off point to discuss the evolution of 'theory of mind'.[118] However, many researchers have been reluctant to attribute rudimentary forms of theory of mind to nonhuman species. For example, some researchers distinguish between behaviour that is the result of 'decision rules' and behaviour that is the product of a real 'theory of mind' without ever specifying what the difference is, or why the latter is not merely a particularly sophisticated instance of the former.[119] Thus, the emergence of early theories of mind represents another future growth area for evolutionary theories of cooperation.

*Communication, mutual expectation, and agreement*

Whereas prey might attempt to frustrate the predictions of a predator, to the extent that an organism has an interest in being predicted, we should expect it to act in ways that make the prediction more accurate. To the extent that it is in the interests of the participants, and to the extent that it is cost-effective to do so, we should expect one organism to facilitate the prediction of another by exaggerating those aspects of its phenotype that the other uses to resolve uncertainty – in other words, by signalling.[120]

As Darwin noted, signals evolve by exaggerating, amplifying or stylising previously arbitrary aspects of the phenotype, including size, colour, odour, sound and behaviour patterns.[121] The results, as the primatologist Marc Hauser puts it, are that: "Fireflies flash, honeybees dance, ants lay perfumed trails, midshipmen hum, electric fish zing high voltages, lizards flash dewlaps, bullfrogs belch, chickens crow, kangaroo rats drum, horses whinny, wolves howl, lions roar, dolphins click, whales sing, baboons grunt, gibbons duet, human infants babble, and human adults talk."[122]

---

[118] Baron-Cohen, 1995; Byrne and Whiten, 1988; Humphrey, 1976; Povinelli and Godfrey, 1993; Whiten and Byrne, 1997.
[119] For example, see: Holekamp, et al., 2000.
[120] Johnstone, 1998.
[121] For more on signalling and communication, see: Darwin, 1872/1998; Johnstone, 1998.
[122] Hauser, 2001, p221.

Some species use signalling to maintain group cohesion whilst on the move. "Coatis, social mongooses, giant river otters, and some canids emit quiet contact vocalizations more or less continuously during group travel."[123] In many social species – including ground squirrels, chickens and several species of monkey – the individual that first detects a predator will emit a warning call to the others, thereby coordinating the behaviour of the rest. And in some species, individual foragers alert others with whom they associate to the presence and location of food so as to better defend it against rival groups of foragers. For example, ravens will alert their nestmates to the location of fresh carcasses.[124]

These signalling systems can be fairly sophisticated. Vervet monkeys have specific alarm calls for leopards, lions and hyenas, hawks, snakes, baboons and unfamiliar humans. Each different call elicits a different response from the other members of the group. Vervets don't give warning signals if they are on their own, and are more likely to give the signal the more kin members there are around.[125] North American prairie dogs use a variety of calls to communicate to the other dogs in its colony not only the identity of the approaching predator (coyote, human, domestic dog, or red-tailed hawk), but also the colour, size and shape of the predator, and information about its speed and movement. The other prairie dogs react differently, with different tactics, depending on the nature of the threat. Slobodchikoff observes that the "sources of information in alarm calls appear to function as a primitive grammar, composed of nounlike, adjectivelike, and verblike elements".[126]

Communication can be used to confirm relatively elaborate expectations about future behaviour. Troops of baboons often sleep together, but fragment whilst foraging during the day, only to meet up later at a distant watering hole. Before dispersing, different groups will initiate movement towards their preferred watering hole, and the troop will disperse only

---

[123] Holekamp, et al., 2000, p607.
[124] Dall, 2002.
[125] Cheney and Seyfarth, 1990.
[126] The combinatorial nature of the system allows them to describe entirely novel objects, such as the experimenter's 'black oval. And calls vary between colonies, exhibiting dialects Slobodchikoff, 2002, p258.

once a consensus has been established as to which hole to meet up at. As the primatologist Richard Byrne observes:

> In some populations, baboons communicate about out of sight locations that are not immediate goals by mass initiations of movements in the direction of these locations, usually shady water holes in desert habitat. The direction of movement eventually chosen predicts which water hole will be visited some hours later, but not the paths of baboon groups in between. This seems to be the strongest evidence in wild nonhuman primates for an ability to conceive and communicate about events displaced in time and space . . . [127]

*Precedent*

In some cases, the behaviour of others is not available to inspection, and the animal will have to remember what others did in the past, and act on the assumption that they will act the same way in the future. For example, many species of birds congregate annually for the purposes of selecting mates. These congregations – known as 'leks' after the bird in which they were first studied – could take place in any number of possible sites. Each bird wants to arrive at the same site as every other bird (as presumably more individuals means more choice and more chance of being chosen). It seems that the birds (and some species of bat) solve this problem by returning to "traditional" display sites every year.[128] Juveniles may learn the location of these sites from their parents.[129]

*Leadership*

Leadership provides another solution to the problem of coordination.[130] For example, animals that travel in groups often face coordination problems as to when and where to move. As the primatologist Sue

---

[127] Byrne, 2000, p518.
[128] Alcock, 1998, p510.
[129] For examples, see: Avital and Jablonka, 2000.
[130] See Foss, 2000 for a review of leadership theory.

Boinski puts it: "*Who* leads group travel, decides *where* the group will travel, using *what* information, and *how* do they accomplish these tasks?"[131] One solution is for dominant individuals to make the decision. For example, in African baboons, alpha males "have a disproportionate share in deciding travel direction", and signal their decision with a variety of "[s]pecialized visual and vocal displays". The process is not autocratic: "Other individuals regularly initiate moves that may be accepted or reject by the decision maker . . . There is also evidence that 'weight of numbers' favoring a particular direction of movement has some influence on the final decision."[132] Dominant individuals also take the lead in group hunts. Alpha wolves and cape hunting dogs "usually [lead] . . . the pack and makes the first lunge".[133]

*From selfish herds to sympathy*

In addition to adaptations for coordination, mutualism seems to have fostered the extension of 'sympathy' from kin to unrelated individuals. To the extent that an animal comes to rely on the participation of others in cooperative schemes – especially if the animal interacts repeatedly with the same individual(s) – then that individual may come to have a direct interest in the continued existence and well-being (measured in terms of ability to contribute to the cooperative scheme) of those other individuals.[134] Members of regularly interacting teams might come to have a stake in one another's continued welfare. In such cases we might expect the evolution of adaptations that benefit others not only as a direct result of participation in cooperative teams, but also indirectly, as a result of adaptations designed to ensure that the individual will be fit to cooperate in the future. As Frans de Waal observes, kin altruism seems to provide ample 'pre-adaptations' for more generalised sympathy:

> [W]ith the evolution of parental care in birds and mammals came feeding, warming, cleaning, alleviation of distress,

---

[131] Boinski and Garber, 2000, p4.
[132] Byrne, 2000, p517.
[133] Holekamp, et al., 2000, p614.
[134] Tooby and Cosmides, 1996.

and grooming of the young, which in turn led to the development of infantile appeals to trigger these activities. Once tender exchanges between parent and offspring had evolved – with the one asking for and the other providing care – they could be extended to all sorts of other relationships, including those among unrelated adults.[135]

Hence, dolphins carry struggling associates to shallow water, elephants stand guard around injured herdmates, chimps lick one another's wounds. As de Waal observes: "We are not surprised to find that dolphins, elephants, canids, and most primates respond to each other's pain and distress, because the members of these species survive through cooperation in hunting and defense against enemies and predators."[136] When unrelated individuals form stable, mutually-advantageous associations of this kind, animal behaviourists tentatively use the term "friendships".[137] And when one animal has a stake in the welfare of another, and hence acts to alleviate their distress, animal behaviourists tentatively use the term "sympathy".[138]

In summary, some adaptations for mutualism must overcome coordination problems in order to cooperate successfully. These adaptations include: the ability to generate and share expectations about the behaviour of others; to confirm expectations by signalling; to copy the behaviour of one's group; to adopt local solutions to coordination problems, such as precedents; to lead and be led; and, in some cases, to exhibit 'sympathy' for regular members of one's group. These adaptations give rise to herds, teams, partnerships, and friendships. We will now look at what is in effect a special case of mutualism: delayed mutualism, or reciprocity.

---

[135] de Waal, 1996, p43.
[136] de Waal, 1996, p80.
[137] Dunbar, 1996; Silk, 2002; Smuts, 1985.
[138] de Waal, 1996, Ch. 2; Flack and de Waal, 2000.

**Reciprocity**

The second problem facing cooperating animals is they must defend themselves against free-riders, who take the benefit of cooperation without paying the cost. Perhaps one member of a hunting team will fall back and let others do the work. Or perhaps an individual will accept your help when she's in trouble, but refuse to help you in return when the situation is reversed.

Situations in which there is uncertainty about whether a benefit is being returned, or whether it will be returned in the future, are usually modelled as a prisoner's dilemma.[139] Defect (cheat) is the only stable strategy in such games. In evolutionary game theory terms, delayed exchange could not evolve in an ecology that was characterised by one-shot prisoner's dilemmas. But in a repeated game, under certain conditions the situation can come to resemble an *assurance* game. In such a game, the problem becomes one of how to coordinate on and maintain a superior equilibrium.

Exchanges cease to be prisoner's dilemmas, and become more tractable assurance games, if the players value future cooperation (R) more than they value immediate defection (T) – modified by the rate at which the player discounts the future (*d*) – and if the players are sufficiently likely to meet again (*w*). In other words, a delayed exchange will resemble an assurance game if: $wd$>T-R/R-P.[140]

In some species, background features – such as the fact that members are long-lived, are physically joined to one another, live in groups, or share neighbouring territories – ensure that the possibility and expected value of future interaction is high enough to sustain cooperation.[141] In other species, *w* and *d* may be more variable, and so reciprocity requires that

---

[139] See Appendix 2, Table 3. For an extended discussion of assurance games, see Skyrms, 2004.

[140] As Axelrod puts it, "Mutual cooperation can be stable if the future is sufficiently important relative to the present". Axelrod, 1984, p126.

[141] As Trivers reports, sparrows, vervet monkeys and Belding's ground squirrels respond more aggressively to strangers than they do to neighbours. They seem to behave according to the rule: "This is my neighbour. As long as he stays in his territory he is fine with me. I will not waste energy in foolish strife". As a result, "[r]elations between neighbours to become more peaceful the longer the neighbours associate". Trivers, 1985, pp366-7.

potential cooperators 'reassure' one another by displaying (costly) signals of *immobility* or *low discount rate* that demonstrate that it will be in their interests to reciprocate when the time comes, and thereby prompt the other player to begin with a cooperative move.[142]

Having established that the payoffs constitute an assurance game, the next thing to do is to coordinate on a superior equilibrium. One simple and successful way of coordinating on the superior equilibrium is to open with a cooperative move in the expectation that it will be in the interest of the other player to follow suit. In an assurance game with only two players, two available moves, and two equilibria, it is relatively easy to see which equilibrium is superior, and hence which move is cooperative. In more complicated games, with more players, more moves, and many more equilibria, a more sophisticated strategy may be needed to successfully coordinate on an equilibrium. In such situations it might be a good rule of thumb to open with the move that you would like others to play, so that your move is available as a focal point for others to coordinate upon.

Players that successfully coordinate on a superior equilibrium can be expected to continue to cooperate until such time as circumstances change and the situation reverts to a prisoner's dilemma. Should this happen, and one player defects, then we should expect the other player to follow suit. The policy of meeting defection with defection ensures that cooperators are not exploited by defectors, and that it is never in the interests of a long-term partner to defect.

Some strategies respond to defection with defection or by breaking off relations, thereby depriving the defector of future opportunities for exploitation and cooperation. Some strategies go further and impose an

---

[142] Eric Posner explains the logic as follows: "Because a good type [i.e., a reciprocator] is a person who values future returns more than a bad type does, one signal is to incur large, observable costs prior to entering a relationship. For example, if a good type values a future payoff of 10 at a 10 percent discount and a bad type values the same payoff at a 30 percent discount, the good type can distinguish himself by incurring an otherwise uncompensated cost of 8, which is less than the good type's discounted payoff (9) and greater than the bad type's discounted payoff (7). Because the recipient of the signal realizes that only the good type could afford 8, the recipient is willing to enter the relationship." Posner, 2000, p19.

*additional* cost on defectors – they punish.[143] Punishment can help to make cooperation stable over shorter time periods, or with individuals with higher discount rates.

(Punishment can be used to overcome some obstacles to cooperation in a repeated prisoner's dilemma. If a penalty (F) imposed on a cheat is greater than the benefit attained from cheating (T-R), modified by the chances of being caught (w), and the rate at which the cheater discounts the future (d) – in other words, if F>T-R/wd – then the payoffs in the game no longer resemble a prisoner's dilemma.[144] If the cost of inflicting punishment is in part a function of its severity, then punishment will be an efficient strategy only if the costs (cF) are less than the net gain from cooperation; if cF<R-P. In turn, the costs of imposing a penalty might be dependent on the condition of the punisher – his relative strength, the size of his kin group, the number of allies he has, and so on. All in all, punishment will be more successful in maintaining cooperation the more likely the cheat will get caught, the sooner he is caught, and the cheaper it is to administer the penalty.)

What if the circumstances that lead a player to cheat were to change, and once again come to resemble an assurance game? Or what if the original defection was a mistake? Can mutually-beneficial cooperation be re-established under such conditions? In order to remove any incentive for cheating, we might expect the cheated player to cooperate only if the cheater divested himself of the benefit of the original defection. This cost could be imposed by punishment, or perhaps the cheat could provide a costly benefit to the cheated player, for example, by playing a cooperative move against the cheated player's defection. Such a strategy ensures that the relationship is robust against error or changing circumstances, and can return to a superior equilibrium after a period of mutual defection.

---

[143] The availability of other partners, transaction costs, costs of imposing penalty, and so on are among the factors that determine which will be the most efficient response to defection.

[144] For example, if the temptation (T-R) is 10, and the probability that the cheat is caught is 1 (*w*), then the cost imposed will have to be greater than or equal to 10. But if the probability that the cheat is caught is 0.5, then the cost imposed will have to be greater than or equal to 20. Likewise, for a benefit of 10, if the cheat's discount rate is 0.5 per time period, and the cheat is caught on average after two periods, then the penalty will have to be 40 (that is, 10/0.5 x 0.5).

In a now famous computer tournament, Robert Axelrod showed that a relatively simple strategy – Tit-for-Tat – that combined many of these features was stable and successful against a range of equivalent strategies in a repeated prisoner's dilemma. (We might describe such a strategy as: "IF wd>T-R/R-P, THEN Cooperate. IF the other player Cooperates, THEN Cooperate in the next round. IF the other player Defects, then Defect in the next round. Repeat.") So, TFT opens with a cooperative move, and then does whatever the other player did in the previous round. Defecting against defectors defends against exploitation; and cooperating with cooperators maintains a mutually-beneficial equilibrium and makes it possible to restart cooperation after a bout of mutual defection.

So much for the theory.[145] Has reciprocity ever evolved in practice?

There have been numerous reports of reciprocal cooperation in the animal behaviour literature since Robert Triver's landmark paper in 1971.[146] Examples include: reciprocal grooming (cleaner wrasse, impala, numerous primates); reciprocal parental-care (pied kingfishers, dwarf mongoose, green woodhopoe, hyena females, silver backed-jackals); reciprocal food sharing (brown vampire bats); reciprocal predator-inspection (sticklebacks, guppies); and reciprocal alliances (vervet monkeys, Rhesus macaques, stump-tail macaques, bonnet macaques, baboons, chimpanzees, lions and dolphins).[147] The problem is that,

---

[145] The 'folk theorem' in game theory states that there are an infinite number of stable strategies in an indefinitely repeated prisoner's dilemma. CDCCD… is the best response to CDCCD… DCCCD… is the best response to DCCCD… and so on. The theory of games, and even Axelrod's tournaments, provides no *guarantee* that a tit-for-tat-like strategy will inevitably evolve. We may nevertheless proceed to investigate whether, in fact, such a strategy has actually evolved. (This is generally the case with evolutionary theory; an adaptationist prediction about the form that an 'optimal' strategy might take is no guarantee that any such adaptation exists. It merely tells us what to look for.)

[146] Trivers, 1971.

[147] Dugatkin, 1991; Harcourt and de Waal, 1992; Leimar and Axén, 1993; Ligon, 1991; Milinski, et al., 1990; Seyfarth and Cheney, 1984; Wilkinson, 1988; Wilkinson, 1984. The payoff structure between genes on a chromosome, and between parasite and host, also seem to resemble repeated prisoner's dilemmas; with the result that cooperation breaks down under predictable circumstances. See: Axelrod, 1984, Chapter 5; Cronin, 1992, p259; Nesse and Williams, 1994, pp57-61. But, as before, in this section I will be focusing on the explicit reciprocal exchange of goods and services between mobile multicellular organisms.

because mutualism shades almost imperceptibly into reciprocity, it has not been easy for field-workers to distinguish between the two. And not all these examples of cooperation are universally accepted as examples of reciprocity. What we need is a clear, unilateral provision of benefit at a cost to the actor, followed either by the return of the benefit, and then perhaps repeated interactions with trusted cooperators; or failure to return the benefit, leading to refusal to aid in future or punishment, and perhaps some kind of reparation.

However, two celebrated examples of reciprocity – cleaner fish and their clients, and blood-sharing between vampire bats – do seem to display the requisite features.

Many fish make use of smaller cleaner fish, usually wrasse, to remove debris and ecto-parasites from their mouth and gills.[148] The client fish are groomed, and the cleaners get an easy meal. The interaction between client and cleaner can be modelled as a prisoner's dilemma. Clients can be cleaned and then let the wrasse go (cooperate); or they can be cleaned and then eat the cleaner (defect). The cleaners, meanwhile, can clean the grouper (cooperate); or they can clean the client and then take a bite out of its exposed and vulnerable mouthparts (defect). What makes cooperation possible in this situation? First, clients need to be groomed regularly, and cleaners need to feed regularly, and so both value future opportunities for cooperation. Second, cleaners maintain fixed territories, which makes them easy to find; and because it costs time and energy to find alternative cleaners, it pays the client to return to the same cleaner. Hence both cleaner and client can expect to interact repeatedly in the future. Cooperation occasionally breaks down because cleaners take a bite from less sensitive areas of the client's mouth which they hope will go undetected. If they are detected the response of the client depends on whether they have access to other cleaners. Clients with only one cleaner on their territory will respond by chasing the cleaner, thus imposing a cost on it. Clients with access to more than one cleaner will respond by

---

[148] Trivers, 1971.

taking its business elsewhere.[149] In order to re-establish cooperation after such a breach, cleaners will stop feeding in order to massage their client's dorsal area with their fins.[150]

Vampire bats regurgitate blood to unrelated roost mates who have failed to find a meal that night.[151] The exchange constitutes a significant benefit to those involved, because hunting is a risky way of life, and blood-sharing acts as a kind of insurance policy that can save individuals from starvation. The exchange between bats can be modelled as prisoner's dilemma. A bat can beg for blood, and give blood when asked (cooperate), or it can beg blood, but refuse to give blood when asked (defect). What makes cooperation possible in this situation? Hunting, and unsuccessful hunts, are an everyday occurrence for vampire bats, and so the benefits of immediate defection are outweighed by benefits of future opportunities for cooperation. Second, bats live in communal roosts for several years and so have a high probability of meeting again. Individuals recognise one another by their calls, and tend to preferentially regurgitate to those with whom they have associated in the past. An individual who fails to regurgitate when asked (thereby cheating) will be punished by having her subsequent requests for blood refused.

However, the picture is much less clear when we turn to examples of nonhuman primate reciprocity. Coalition formation between male baboons provided, for a long time, one of the most celebrated examples of primate reciprocity.[152] Two male baboons will often team up to defeat a third, usually in a dispute over a female. Because individuals tend to solicit and receive assistance from those whom they have assisted in the past, it was assumed that this constituted reciprocal altruism between the two males. However, males may have been cooperating because both had an immediate interest in doing down the third. The solicitation could be translated as "I want to attack that individual, you want to attack him too. Shall we do it together?" If this was the case, then it is no surprise that

---

[149] Bshary and Grutter, 2002; Bshary and Schäffer, 2002.
[150] Bshary, 2001.
[151] Wilkinson, 1988; Wilkinson, 1984.
[152] Packer, 1977.

individuals with similar interests – and similar rivals – regularly assist one another. Moreover, as Ronald Noë has argued, in order for the cooperation to count as reciprocity, the baboons would have to take turns mating with the female. If baboon A helps baboon B win the female this time, then baboon B should help baboon A win the female next time. But this is not what happens. After the coalition has successfully defeated the third-party, the allies enter a scramble competition for the female, which the dominant member usually wins. It seems that the members of the coalition are cooperating to defeat the rival, and taking their chances when it comes to the prize. As Noë points out, there need be nothing reciprocal about this. A subordinate may have only a 10% chance of 'getting the girl', but will cooperate anyway because 10% is better than nothing. Furthermore, argues Noë, individuals may be choosing partners on the basis of ability (rather than reciprocity), and hence ending up choosing the best available fighters each time.[153]

At the very least, primates seem to operate a kind of *modus vivendi*, a live-and-let-live form of 'negative reciprocity'. That is, they will retaliate against those who aggress against them. Individual chimpanzees will exact revenge against those that aggressed against them, including "allies that fail to support them in competitive interactions with third parties."[154] Retaliation can even be aimed at relatives of the aggressor.[155] These feuds are sometimes brought to an end by reparations and attempts to make amends through affiliative behaviour such as grooming.[156] At other times it takes the intervention of dominant individuals to bring the cycle of violence to an end and restore order.[157]

---

[153] Noë, 1992.

[154] Clutton-Brock and Parker, 1995, p214. See also: de Waal, 1994.

[155] "In vervet monkeys, adult females who have been displaced from food sources may seek out and attack their displacer's relatives. In macaques, members of different matrilineal groups ally with eachother and individuals that have been displaced or attacked by members of another matriline commonly respond by attacking a vulnerable member of their aggressor's matriline. Attacks on a member of a matriline are commonly followed by retaliation against members of the aggressor's matriline by relatives of the victim." Clutton-Brock and Parker, 1995, p211.

[156] Aureli and de Waal, 2000; Silk, 1998.

[157] Aureli and de Waal, 2000; de Waal, 1989. "In many social vertebrates, dominant individuals will also intervene to end aggressive exchanges . . . " Clutton-Brock and Parker, 1995, p214.

This behaviour could be interpreted as a kind of implicit social contract in which everyone cooperates by not using violence to get what they want, but where retaliatory violence – perhaps as an extension of self-defence – is acceptable as a deterrent. If this were to count as reciprocity, then it would be hard to tell who was reciprocating with whom, because you have to conclude that, in the absence of open hostilities, everybody was cooperating with everyone else all the time. Alternatively, retaliatory aggression could be seen as a coercive strategy used by some animals to threaten others into doing what they want or punish them for not doing so (as opposed to aggression used to punish defectors and pave the way for cooperation), and the ensuing conflicts could be seen as a breakdown of the usual mechanisms used to resolve conflicts (as discussed in the next section).

It may be that reciprocity is rare among primates because they have little to exchange, and when they do, as in the case of a successful hunt, reciprocal relations are swamped by strong kinship bonds and hierarchical social relations. However, the absence of abundant evidence for reciprocity is not necessarily evidence of absence. As de Waal comments: "Monkey groups may be veritable marketplaces in which sex, support, grooming, food tolerance, warnings of danger, and all sorts of other services are being traded. To us scientists falls the task of figuring out the worth of each service and of following the relationships over a long enough period to understand that deals are being struck."[158]

In summary, while few species seem to have mastered full-blown reciprocity, many species exhibit aspects of it. And whether or not nonhuman primates turn out to practise reciprocity, it is clear that they have ample preadaptations for it – in terms of long-term cooperative associations, individual recognition, performance assessment, punishment, appeasement and so on – should the conditions for its evolution arise.

---

[158] de Waal, 1996, p156.

**Conflict resolution**

The view that life in the state of nature is 'a war of all against all' is further contradicted by ways in which animals manage and resolve conflicts of interest. The widespread tendency of animals to 'restrain' themselves during combat, or replace actual fighting with ritual contests was once attributed to group selection or to 'the good of the species'.[159] It took the injection of game theory to formalise how such strategies could benefit the genes involved. In the paper that introduced evolutionary game theory – *The Logic of Animal Conflict* – John Maynard Smith and George Price showed that, contrary to what was once widely assumed, natural selection does not favour all-out aggression in competition between organisms, for the simple reason that fighting has costs that are sometimes best avoided.[160]

Suppose that two evenly matched organisms are competing for a resource (V), and that they can choose to fight or to withdraw. If they both fight, then they have an equal chance of winning the resource, but suffer the costs, in terms of time, energy and injury, of the fight (c). If one fights and the other withdraws, then the fighter gets the resource and the other one gets nothing. If they both withdraw, then we might suppose that they each have an equal chance of walking away with the prize. The evolutionary-stable strategy in this game is to fight with a probability of V/c. If V<c then each player will withdraw a certain proportion of the time.[161]

In this simple version of the game, both players prefer the outcome in which one player threatens to fight and the other withdraws over the worst outcome, in which both players fight. Despite appearances, the players have a common interest in minimising the costs of a dispute. However, with mixed strategies, the players are leaving the outcome to

---

[159] The ethologist Konrad Lorenz noted that: "[N]obody with a real appreciation of the phenomena under discussion [use of ritual in animals] can fail to have an ever recurring sense of admiration for those physiological mechanisms which enforce, in animals, selfless behaviour aimed towards the good of the community, and which work in the same way as the moral law in human beings." Lorenz, 1966b, p94.

[160] Maynard Smith and Price, 1973.

[161] See Appendix 2, Table 4. If V>c then the situation resembles a Prisoner's Dilemma, and the players will always fight.

chance – they will still come to blows $(V/c)^2$ of the time. Maynard Smith has shown that a range of conditional strategies that make use of some asymmetry between the players are better at coordinating on one of the equilibria outcomes and are evolutionarily stable against a range of other strategies.[162] These conditional strategies can be divided up according to whether they make use of *correlated* or *uncorrelated* asymmetries to settle the dispute.

*Correlated asymmetries*

*Correlated* asymmetries refer to differences between players that correlate with the probability that they would win if it came to a fight – often referred to as the player's 'resource holding power' (RHP). In the simple game, the players are equal in every respect. But in practice, animals are likely to differ in their fighting ability (hence their chance of winning the fight will not be 0.5), in how much they value the resource (V), or in their ability to endure the costs of the fight (c). These differences mean that the players differ in their willingness to fight. The more likely a player is to win, the more he values a resource, or the more he can endure the costs of fighting, the more likely he is to opt for a fight (and hence the more sense it makes for the other to give way). Given that these asymmetries will tend to settle the conflict, it pays both players to discover them *before* the conflict rather than *afterwards*. The outcome is the same, but both winner and loser benefit from avoiding the costs of conflict. So we might expect, prior to – or instead of – coming to blows, participants to display traits relevant to their chances of success, to assess the other's traits, and to moderate their willingness to fight accordingly. They might, for example, come to adopt a strategy such as: "Display. If opponent is bigger/stronger/faster/heavier/hungrier… then retreat."

The signals used to settle disputes – referred to as costly signals, honest signals, fitness indicators – can take a variety of forms depending on the nature of the contest and the qualities needed to win it.[163] In competition

---

[162] Maynard Smith, 1982.
[163] The signals used to settle disputes are a particular instance of the general rule that information transfers between parties with conflicts of interest requires some mechanism to ensure their

over resources or territory, the signals will reflect the physical prowess of the participants. The classic example comes from a study of stag red deer competing over the control of harems.[164] The contest begins with a roaring match lasting several minutes. Roaring is an honest signal of size and strength. If the stags are too closely matched for this to settle the dispute, the contest moves to a 'parallel walk' stage where the combatants have the chance to size one another up. If this doesn't settle the dispute, then the stags lock antlers and begin a pushing contest, and the loser retreats. In other competitions, age, size, weight, strength, skill, experience, or number of allies may carry the day.[165]

Settling disputes through costly signalling also occurs in the context of female choice of mates. In such cases, males compete with one another for access to females by displaying the qualities – usually 'good genes', or some form of paternal investment – that constitute the criteria of female mate choice.[166] This 'sexual selection' of fitness indicators has produced some of the most elaborate displays in the animal world. The most celebrated example is the peacock's tail. This enormous and cumbersome tail imposes a cost on its bearer that only the highest-quality males could endure. Moreover, the tail acts as a billboard advertising the peacock's genetic quality both directly, and through the individual's parasite load. Other examples of sexually-selected fitness indicators include: birdsong, bower bird's nests, and dolphin "air art".[167]

It is worth drawing attention to two further features of costly signalling. The first is that costly signals can often produce 'positive externalities' for

'honesty'. 'Honesty' or 'credibility' is usually achieved by making the signal 'hard to fake' or 'costly' such that low-RHP individuals cannot, or cannot afford to, display them. (Obviously, if the signal led the signaller to win more fights was 'cheap', then it pays use the signal irrespective of one's actual RHP. Consequently, receivers would not benefit from responding to the signal, and should no longer pay attention to the signal, so both players would be back where they started, incurring costly fights.) For example, 'large size' cannot be displayed by low-RHP individuals because they do not possess it. Alternatively, 'charging' is a behaviour that both low- and high-RHP individuals can display, but that might prove prohibitively expensive for low-RHP individuals, in the sense that the cost that they incur outweighs any benefits they may acquire through bluffing. Furthermore, what start out as a fairly 'crude' signals of resource-holding power can become more exaggerated, elaborate and 'symbolic' over evolutionary time.

[164] Clutton-Brock and Albon, 1979.
[165] For a review, see: Riechert, 1998.
[166] Miller, 2000a; Ridley, 2000.
[167] Miller, 2000a; Zahavi and Zahavi, 1997.

their audience. For example, in many species of insects, birds, and mammals, males compete by providing females with packets of food – "nuptial gifts" – which provide benefits to the recipient female and her offspring in addition to honestly signalling the male's foraging prowess.[168] Ravens feed on the carcasses of dead animals; but it is risky business to check whether an animal is in fact dead, and not asleep or injured, or whether the carcass is being protected by some other scavenger. The best strategy would seem to be to hang back and let others take the risk of establishing the status of the body, and only come forward when they sound the all clear.[169] But, as experiments with stuffed models of predators have shown, males actually *compete* to take risks. It appears that "the occasional boldness of corvids serves to enhance status and impress potential mates by demonstrating that they have the courage, experience, and quickness of reaction to deal with life's dangers". This interpretation is supported by the observation that males often "alternate between acts of bravery and courtship of female onlookers".[170]

The second feature is that, in stable social groups, the results of ritual contests can form the basis of 'dominance hierarchies'. In stable social groups in which members interact repeatedly, there may be no need for repeated displays; individuals can merely remember what happened last time, or else watch how others fare. Situations in which the results of previous contests are used to settle current disputes – in other words, when players trade on their reputation – are referred to as "dominance hierarchies". Dominance hierarchies represent a further de-escalation of hostilities, reducing the costs of even the ritual contests for players, and ensuring a greater degree of stability (until, of course, a player decides to challenge the 'status quo'). To quote Preuschoft and van Schaik: "dominance in groups seems to function as a conflict management device, preventing escalated competition by conventionalizing means and

---

[168] Alcock, 1998, pp457-61.
[169] As we have seen, ravens often recruit others to feeding sites so as to better defend them against other groups of ravens and other scavengers.
[170] de Waal, 1996, p134.

priority of access [to scarce resources], thus allowing for peaceful coexistence of group members."[171]

Some species combine these two features of costly signalling: they compete for status in social groups by performing acts that benefit onlookers. For example, male Arabian Babblers (a social songbird, about the size of a sparrow) compete for status by acting as lookouts or "sentinels", alerting other members of their group to the approach of a predator.[172] Acting as sentinel is costly, as it prevents the bird foraging, and leaves it exposed to predators. But instead of trying to shirk their responsibilities – as one might expect from a Hobbesian perspective – males actually compete to act as the sentinel.

Competition through 'public service' also seems to be a feature of primate dominance hierarchies, especially among chimpanzees. De Waal reports that chimpanzee males take risks in order to provide the group with food, are generous with their own kills, and confiscate the kills of others and redistribute them. "Possibly, then, generosity serves political ends: food distribution enhances an individual's popularity and status. . . . [I]nstead of dominants standing out because of what they take, they now affirm their position by what they give."[173]

Some primates have gone even further. They compete for status by shouldering the costs of solving collective-action problems, thereby turbo-charging cooperation. We have seen how dominant individuals provide leadership when it comes to coordinating the activities of groups of roaming foragers, and in the context of cooperative hunts. Individuals are apparently competing for dominance through their experience, their

---

[171] Preuschoft and van Schaik, 2000, p90.

[172] Zahavi and Zahavi, 1997.

[173] de Waal, 1996, p144. The notion of 'tolerated theft' provides a framework in which to understand the evolution of altruism as a costly signal of fitness. Successful hunters bring meat back to camp; itself a costly signal of fitness. There is more meat than he and his family can eat. It does not pay for him to defend the surplus, and so he tolerates/lets occur scrounging. Tolerating scrounging itself becomes a signal of high-status; amplified by being 'public', that is the fact that it is witnessed by an audience of scroungers at least. "Tolerating theft" becomes "showing off". So the hunter begins to positively encourage 'scrounging', by providing more meat and more surplus. And so on. Thus there is selection for generosity and other forms of provision of public goods. Blurton Jones, 1984.

knowledge, and the wisdom of their decisions.[174] Dominant chimpanzees, as well as stump-tailed monkeys and gorillas, also seem to compete by intervening to end disputes among subordinates.[175] Performing the so-called 'control role' may not be entirely altruistic. By breaking up fights, dominant individuals may be protecting their own position or the position of their allies; they may be shoring up the hierarchy from which they benefit; or they may be securing the support of those that they help. But neither is such behaviour entirely self-serving. Unlike the interventions of most group members – who support family and allies – control males seem to intervene in a way that is most likely to restore order. "Dominant chimpanzees generally break up fights either by supporting the underdog or through impartial intervention. . . . They seem to interfere on the basis of how best to restore peace rather than how best to help friends."[176] And ultimately, to the extent that individuals benefit from group living, taking action to secure the cohesion and functioning of groups is to the benefit of all. It would appear, then, that male chimpanzees compete for status in part by being an effective and impartial mediator, a keeper of the peace. As de Waal comments, "the group looks for the most effective arbitrator in its midst, then throws its weight behind this individual to give him a broad base of support for guaranteeing peace and order. . . . A fair leader is hard to come by, hence it is in the community's interest to keep him in power as long as possible. . . . A higher authority who takes care of these problems [disputes] with impartiality and minimum force must be a relief for all."[177]

Finally, we should spare a thought for the subordinates, the 'good losers', the doves, without whom the resolution of conflict would not be possible. Just as costly signals of fitness involve the exaggeration and elaboration

---

[174] "Circumstantial evidence indicates that competent leadership of troop movement requires familiarity with the ranging area. . . . [O]lder, and presumably more experienced, troop members predominating in leadership roles . . . Ungulates provide stronger evidence than primates that leaders are the oldest group members and also the most complete repositories of knowledge of the ranging area. Upon the death of their leader, family units of African elephants become disoriented, seemingly rudderless, and lose their normal ranging patterns. Thin-horn mountain sheep (*Ovis dalli*) acquire their adult ranging patterns as a direct consequence of which adult leader they elect to follow as juveniles." Boinski, 2000, p457.
[175] Das, 2000; de Waal, 1989.
[176] de Waal, 1996, p129.
[177] de Waal, 1996, pp130-1.

of attributes predictive of success, cues of submission involve the exaggerated concealment of these same attributes. "While threat displays accentuate size and weapons and elicit yielding on the part of the recipient, displays of submission reduce apparent size, conceal weapons, and correlate with yielding on the part of the sender."[178] The same exchange of signals – albeit in a swifter, and more symbolic form – takes place when members of dominance hierarchies meet. These 'greeting rituals' "function to reassure partners that are uncertain or ambivalent about their relationship. Reassurance is accomplished by demonstrations of trust and trust worthiness: partners make themselves vulnerable to each other." There are also more formal, symbolic signals of subordination. Macaques give a 'silent bared-teeth display', and chimpanzees "use a vocal-gestural signal of subordination consisting of repetitive pant-grunting and bowing towards the dominant."[179]


*Uncorrelated asymmetries*

*Uncorrelated* asymmetries refer to differences between the players that do not correlate with the probability that a player would win a fight. In theory, these differences can be entirely arbitrary.[180] In practice, however, uncorrelated asymmetries usually have to do with prior possession of a resource, and the strategy takes the form of: "If owner fight, if intruder submit". The point, in either case, is that the strategy provides a fast and efficient way of arriving at one of the equilibria outcomes and hence avoiding fights; and that, once established, the strategy is stable against invasion by alternative strategies that don't obey the convention. In a population of owner-occupiers, a mutant strategy such as "If intruder fight, if owner submit" would incur the costs of fighting on every occasion, and be removed from the population.

---

[178] Preuschoft and van Schaik, 2000, p85.
[179] Preuschoft and van Schaik, 2000, p90, p93, p96.
[180] Cronin gives the example of the strategy "Fight if you approach from the North, submit if you approach from the South". Cronin, 1992, p315.

The recognition of prior ownership as a means of settling disputes over scarce resources – territory, mates, food – is widespread in nature.[181] For example, male speckled wood butterflies take possession of small patches of sunlight on the forest floor. Should another butterfly arrive at the same patch there is a brief tussle, but the interloper will defer to the previous owner. By temporarily removing the butterflies, experimenters can fool both into thinking that they 'own' the patch. Now deprived of their conventional means of settling disputes, when the butterflies next meet, the fight that usually last for a few seconds continues for up to 40 seconds.[182] In another elegant experiment, two sticklebacks were allowed to establish and defend territories at the far ends of an aquarium. The fish were then placed in large glass test-tubes, and, predictably, when the two tubes were brought together they tried to attack one another. Moreover, the experimenter could cause one to attack and the other to retreat by moving the two tubes from one end of the tank to the other. The point at which each fish was equally belligerent constituted the boundary between their respective territories.[183]

Given that insects and fish can recognise and respect the property of others, it comes as no surprise to find that, as Marc Hauser notes, "in almost all territorial species, intruders respect territory ownership". Hauser goes on to say that "The space that a territory owner defends is functionally equivalent to his property, and an intruder's respect reveals his acknowledgment of ownership and property rights."[184]

Animals usually have little property to defend other than their territories. But the same recognition of ownership seems to apply on those occasions when they do have valuable (and mobile) resources, such as food and mates. Experiments by Hans Kummer and others have shown that,

---

[181] Robert Ardrey's *The Territorial Imperative* provides the classic summary – although, like Lorenz's *On Aggression*, it suffers from being somewhat group-selectionist. Ardrey, 1966.
[182] Reported in: Cronin, 1992, pp315-6.
[183] Dawkins, 1976, pp79-80.
[184] Hauser, 2001, p303. The notion of 'property' or 'ownership' presumably evolved as an extension of the 'zone' surrounding an individual. So, many animals might be expected to keep away from other animals – respect their 'personal space' – lest they make themselves vulnerable to attack or retaliation. It would be relatively easy for natural selection to extend this no-go zone to include whatever area or resource is likely to be defended by the individual, and hence subject to reprisal.

among macaques and baboons, 'carrying an object close' acts as a cue of ownership.[185] As Kummer notes, among the species of monkey studied so far, "the respect for prior ownership of a valuable object is well established even when the rival is dominant".[186] de Waal reports that among chimpanzees, once "food is firmly in a subordinate's hands, his or her ownership is generally respected".[187] And many animals take a similarly proprietary attitude to their sexual partners. For example, hamadryas baboons "do not interfere with each other's bonds with females: even large, totally dominant males are inhibited from taking over the female of another male after having seen the two together for a couple of minutes".[188]

In summary then, a variety of species use displays of prowess, genetic quality, generosity and leadership to settle disputes, and these displays give rise to stable social hierarchies, lubricated by ritualised greetings. Many species also rely on the recognition of prior ownership to settle or forestall disputes over scarce resources such as territory, food and mates.

**Conclusion**

This brings to an end our brief summary of the social lives of animals. We have seen how, contrary to the caricature of life in the state of nature, evolutionary theory leads us to expect organisms to be motivated to cooperate and even be altruistic under certain circumstances. Under the headings of kinship, coordination, reciprocity and conflict resolution we have seen that a variety of animals: care for their offspring and their wider families; form teams and partnerships, coordinate to mutual advantage, and value the welfare of other team-mates; practise a rudimentary form of reciprocity, including punishment and restitution; resolve disputes by means of ritualised contests and dominance hierarchies, which may in turn provide benefits for onlookers and other members of the group; and respect the property of others.

---

[185] Kummer and Cords, 1991; Sigg and Falett, 1985.
[186] Kummer, 1991, p70.
[187] de Waal, 1996, p152.
[188] de Waal, 1996, p245

It would appear that cooperation is far more widespread, sophisticated, and deeply entrenched in the animal world than Hume would have imagined or dared hope. Equipped with this catalogue of adaptations for cooperation, we may now proceed to investigate whether or to what extent humans have retained – or evolved equivalent versions of – these adaptations.

## Chapter 4 – Human adaptations for cooperation

*[I]t can hardly be disputed that the social feelings are instinctive or innate in the lower animals; and why should they not be so in man?*

– Charles Darwin[189]


**Introduction**

In the previous chapter we saw that, contrary to the Hobbesian caricature of life in the state of nature, evolutionary theory leads us to expect that organisms will be social, cooperative and even altruistic under certain circumstances. Under the categories of kin altruism, coordination to mutual advantage, reciprocity and conflict resolution, evolutionary theory can explain why and how some organisms care for their offspring and their wider families, aggregate in herds, work in teams, practise a division of labour, communicate, share food, trade favours, build alliances, punish cheats, exact revenge, settle disputes peacefully, provide altruistic displays of status, and respect property.

Given that adaptations for cooperation are widespread in other species, and seem especially well-developed in non-human primates, the default assumption of research on human psychology should be that humans have retained versions of these primate adaptations. And indeed, research conducted on this basis has begun to uncover many of the predicted features of such a suite of adaptations. The purpose of this chapter is, first, to review this literature, and second, to ask whether human adaptations for cooperation correspond to what Hume called 'the moral passions'. If humans do have adaptations for cooperation, and if they correspond to Hume's moral passions, then we may conclude that evolutionary biology provides a firm foundation for Hume's account of moral psychology and the subjective nature of moral value.

---

[189] Darwin, 1871, p71.

**The burden of proof**

Before getting into the details of human adaptations for cooperation, it is important to be clear about the form that the argument takes. It runs as follows. First, adaptations for cooperation are widespread in other species and seem extremely well-developed in primates. Second, humans are descended from social primates. Therefore, in the absence of any explanation of how or why natural selection would have removed adaptations for cooperation from the human line, the default assumption should be that humans continue to possess such adaptations.

As such, the question facing psychologists is not *whether* humans have adaptations for cooperation, but *which* adaptations for cooperation do they have? In particular, how have human adaptations for cooperation been modified by the specific set of selection pressures at work since we split from our last common ancestor with chimpanzees six million years ago? According to the anthropologists John Tooby and Irvine DeVore, the principal difference between pongid and hominid evolution was the reliance of the latter on hunting as a means of subsistence.[190] Hunting, they argue, created the right selective environment for the evolution of increased male parental investment, extended coalition-formation, and full-blown reciprocity. Hunting also seems to constitute an additional arena for costly signalling. So, we should expect humans to possess adaptations for: maternal care, paternal investment and kin affiliation; forming interest groups, coordinating behaviour and extending sympathy to regular collaborators; detecting and punishing cheats; settling disputes and forming dominance hierarchies by means of costly – sometimes altruistic – signalling; and for respecting prior ownership.

Against this background, it would be reasonable to assume that humans are equipped with such adaptations *even in the absence of any further positive evidence*. This is because in order to argue that humans *do not* have such adaptations for cooperation, one would have explain how and why natural selection *erased* or *removed* such adaptations from the human line. It is difficult to imagine what such an argument would look like, and, as

---

[190] Tooby and DeVore, 1987.

far as I am aware, no-one even attempts to make it. Fortunately, however, there is considerable empirical evidence for the continued presence of human adaptations for cooperation. As Appendix 3 suggests, evidence for human adaptations for cooperation could come from genetics and heritability studies, from research into hormonal, neurological and other proximate mechanisms, from behavioural and experimental studies, and from anthropological, sociological and historical surveys of large-scale patterns of human behaviour. In an ideal world, evidence would come from all these sources, and this chapter would present a full circuit-diagram – a Gray's Anatomy – of human adaptations for cooperation. But adaptationist research into human psychology is at a very early stage, and such a complete description is not available.[191] Instead there is a handful of preliminary – but suggestive – evidence from experiments on human psychology, and plenty of circumstantial evidence of these adaptations at work. Below, I provide a brief sample of this evidence, from each of the four categories of cooperation: kin altruism, coordination to mutual advantage, reciprocity and conflict resolution.[192]

**Kin altruism**

As we saw in the previous chapter, genes can spread by helping copies of themselves that reside in other individuals. One common result of this 'kin selection' is parental investment, but it has also led to cooperation among siblings and forms the backdrop to numerous other forms of social and altruistic behaviour.

When it comes to parental investment, we saw how, in sexually-reproducing species, equal parental investment by both sexes is unstable, and that one sex usually specialises in 'parenting effort', while the other specialises in 'mating effort'. This division of labour is particularly pronounced in mammals, in which the females have adopted an intensive paternal investment strategy that includes gestation and lactation as well

---

[191] I return to what needs to be done to complete such an account in the final chapter.
[192] As mentioned above, the collective-action problems faced by any particular species do not necessarily fall neatly into one or other category, but are combination of them. As a result, the adaptations for cooperation of a given species may bear the stamp of various combinations of selection pressures.

as costly periods of care of the infant. Humans are no exception to this pattern.

Humans continue to manifest typical patterns of mammalian sex differences: men and women have different genes and chromosomes, leading to different hormones, which lead to differences in the design of their bodies and brains, and ultimately in their reproductive strategies and mate preferences.[193] Later in this chapter, we shall look at the role of androgens in preparing males to be more aggressive, competitive and risk-taking. At this point, however, we merely need to note that, in the absence of such hormones, the default female developmental track produces individuals that are, relative to males, less aggressive, competitive, and risk-taking, and are more nurturing, 'people oriented', and interested in infants.[194] (These traits are noticeably reduced or lacking in girls who have been exposed to abnormally high levels of male hormones.)[195]

The end product is a predictable sex difference in willingness and ability to provide primary childcare – a sexual division of labour that continues into the modern world: "Females in all known societies exhibit more nurturing behavior than males both inside and outside the family . . . Sex differences in parental care are universal across cultures. While the level of paternal involvement varies considerably among societies, there is no society in which the level of direct paternal care approaches that of mothers."[196] Again, in order to argue that women's primary role in

---

[193] Buss, 1994; Geary, 1998; Symons, 1979.

[194] Baron-Cohen, 2003 p33-4. Alexander and Hines report that the 'unsocialised' toy-preferences of female and male vervet monkeys manifest a sex-difference similar to that observed in human girls and boys. Given the choice, female vervets preferred a doll and a pot, whereas male vervets preferred a car and a ball Alexander and Hines, 2002. For a review of the sophisticated software underlying female reproductive and mothering strategies, see: Hrdy, 1999.

[195] Congenital adrenal hyperplasia (CAH) refers to a condition in which the fetal adrenal gland produces an excess of androgens during critical periods of embryonic development. Females with CAH are born with male external genitalia, but female internal reproductive organs. With the help of corrective surgery, and treatment for the original condition, girls with CAH usually grow up to become normally fertile females. Compared to controls, CAH girls exhibit more typical masculine male behaviour and less feminine behaviour – in short, they tend to be tomboys. In a review of the literature, Kingsley Browne reports that girls with CAH: "exhibit substantially less interest in infants and express a lesser desire to marry and have children when they grow up, and they have more male-like occupational preferences". Browne, 2002, p110.

[196] Browne, 2002, p21.

childcare is *not* the result of the asymmetry between male and female reproductive strategies as manifest in mammalian biology, but is instead the product of arbitrary socialisation or 'culture', one must first explain how and why – and provide evidence that – natural selection removed the psychology of maternal investment from the human line. I am not aware of any such arguments.

Turning to fathers, the increased dependency of human infants on their parents, and the increased reliance on hunting over the past six millions years, seems to have increased the demand for, and the supply of, male parental investment. The result is that, unlike the 95% of mammalian fathers who invest nothing at all in their offspring, and unlike the majority of primate fathers, who do little more than occasionally protect their offspring from lethal violence, *human* fathers seem to invest heavily in their offspring under the right conditions.[197]

Which conditions? In species with internal fertilisation, fathers can never be as sure as mothers that 'their' child is really their own, and so evolutionists predicted that male parental investment would vary with 'certainty of paternity'.[198] Although it is not known exactly how humans identify kin, one suggestion is that men use facial resemblance to regulate their uncertainty about paternity, and that the degree of resemblance would effect their investment decisions (whereas there would be no such effect for women). To test this prediction, the psychologist Steven Platek presented subjects with an array of five children's faces, one of which had been morphed with their own, and asked them about their feelings towards the children. As predicted, Platek found that:

---

[197] For comprehensive treatment of human paternal investment, see: Geary, 2000.

[198] Robin Baker reports that in the modern, Western world: "On average, about 10 percent of children are not sired by their supposed fathers. Some men, however, have a higher chance of being deceived in this way than others – and it is those of low wealth and status who fare worst. Actual figures range from 1 percent in high-status areas of the United States and Switzerland, to 5 to 6 percent for moderate-status males in the United States and Great Britain, to 10 to 30 percent for lower-status males in the United States, Great Britain and France. Moreover, the men most likely to sexually hoodwink the lower-status males are men of higher status." Baker, 1996, pp124-5. See also: Betzig, 1993.

> [M]ales were more likely to choose the face that resembled
> their own as the one that they would be most likely adopt,
> the most attractive, the child they would spend the most
> time with, the child they would spend money on, and the
> child would least resent having to pay child support for.
> Unlike males, females were relatively indifferent to whether
> the children's faces resembled their own.[199]

Subsequent imaging research has shown that "the male, but not the female, brain responds differentially to children's faces that are morphed with their own face", leading Platek to suggest that "the male brain possesses specific neurocognitive substrates responsible for responding to resemblance expressed in children's faces."[200] These discoveries dovetail neatly with an earlier finding that other interested parties seem to be sensitive to uncertainty about paternity: in maternity wards, the mother's family, more than the father's, can be expected to comment on how much the baby looks like its father.[201]

Indirect evidence of the importance of 'certainty of paternity' comes from cross-cultural studies of male parental investment. One study of 186 pre-industrial societies found that where confidence of paternity is high, men invest far more in terms of wealth, position, and personal involvement than where it is low.[202] Uncertainty of paternity also has effects beyond the nuclear family. As kin selection theory accurately predicts, when confidence of paternity falls below 33%, a man will invest not in his wife's children but in his *sister's* children (his nieces and nephews), for they are likely to be his closest genetic relatives in the next generation.[203] In the

[199] Platek, et al., 2002, p164.
[200] Platek, 2003; Platek, et al., under review.
[201] Daly and Wilson, 1982. Evolutionary theory suggests that different members of the family may recognise kin in different ways. Genetic fathers share 50% of their genes with their offspring, hence phenotypic cues such as facial resemblance provide reliable cues of relatedness. Full siblings, however, share 50% of their genes only on average – some siblings share more some share less. Hence other cues, such as residing in the same home, or being cared for by the same parents, may be more reliable cues of relatedness (Cronin, personal communication). For evidence that siblings use 'co-residence' to establish relatedness, see: Lieberman, et al., 2003.
[202] Gaulin and Schlegel, 1980.
[203] Alexander, 1979, p72. A man's degree of relatedness (r) to his son is 0.5 modified by his probability of paternity (p): 0.5p. A man's degree of relatedness to his sister is also modified by

anthropological literature this is referred to as the 'avunculate' or 'mother's brother' system. More recently, it has been discovered that uncertainty of paternity leads to a sex difference in the altruism of grandparents. The father's father has two degrees of uncertainty between him and the child, whereas the mother's mother has none. As expected, most grandparental investment comes from the mother's mother and least from the father's father, with mother's father and father's mother intermediate.[204]

Generally speaking, the commonality of interests among family members is reflected in the fact that: humans continue to live in family groups, and wider networks of kin relations;[205] human priorities in rescuing people from a burning house, leaving money in our wills or choosing who to turn to for advice neatly track genetic closeness;[206] and social relations between genetic relatives tend to be more harmonious than between non-genetic relatives and strangers. Daly and Wilson report that, despite the proximity of family members, and the amount of time spent with them, only 6% of homicide victims are genetic relatives of the perpetrator.[207]

Common interests among family members has also been reflected in legal systems. Until relatively recently: crimes against an individual were also considered crimes against that individual's kin group; kin groups were often the agents of revenge; retribution was visited upon the perpetrator and on the perpetrator's kin; and kin groups were the recipients of reparations – peone, wergild, blood money – supplied by the kin of the offender. The deterrent function of one's kin group, and the self-policing that they could enforce, constituted the majority of what passed for law in pre-political societies. The role of kin groups in the maintenance of social

---

*his* father's probability of paternity (0.25+0.25p). If brother and sister had the same father, then r=0.5; if they had different fathers, then r=0.25. And the sister's degree of relatedness to her children is 0.5. So, a man is more closely related to his sister's children than his wife's children when: 0.5p < 0.5(0.25+0.25p) or when p<0.333.

[204] Euler and Weitzel, 1999.
[205] Alexander, 1979, pp144-7; Brown, 1991. And all attempts to erase the family, as in Israeli kibbutzim, Soviet Russia, or religious communities, have routinely failed. See, for example, Tiger and Shepher, 1975.
[206] Buss, 1999, pp230-40.
[207] Daly and Wilson, 1988.

order seems to have declined with the relative decline of kin groups themselves, as well as the rise of a centralised authority.[208]

Kin altruism is conspicuous also by its absence. Because parental investment is a costly resource, psychologists predicted that the evolved psychology of parents would be designed to preferentially invest in one's own genetic children, and to be reluctant to invest in children that are not one's own. This is indeed the case with step-parents. Compared to their genetic counterparts, step-parents and children view the relationship as less loving and less dependable emotionally and materially; step-parents withhold investment and look forward to the children leaving home; and step-children do indeed leave earlier. One consequence of this predictable difference in feelings is an enormous differential in the risk of violence. Children are at much higher risk of abuse and murder from step-parents than from genetic parents. The presence of a step-parent turns out to be the single most powerful risk factor for severe child maltreatment yet discovered.[209]

In summary, there is no evidence to suggest that natural selection has removed the standard mammalian-primate adaptations for kin altruism from the human line. On the contrary, human mothers continue to have adaptations for caring for their offspring, fathers have adaptations for investing in their offspring under the right conditions, and everyone it seems has adaptations for caring for their family members. We may now turn to cooperation between unrelated individuals.


**Coordination to mutual advantage**

As we saw in the previous chapter, individuals with common interests can often benefit from working together as a team. Typical examples include collective defence against predators, collective hunting, and the formation of alliances and coalitions in dominance hierarchies. In addition to identifying individuals with common interests, successful

---

[208] Daly and Wilson, 1988, Chapter 10; Posner, 1981, Chapters 5-8.
[209] Daly and Wilson, 1998. Several common objections to the theory are dealt with in Daly and Wilson, 2001.

mutualism requires that individuals overcome the 'coordination problems' that arise as a result of uncertainty about *where* and *when* to do *what*. In such circumstances, selection will favour enhanced predictive and signalling abilities. We also saw how iterated mutualism between individuals might foster an independent interest in one-another's continuing welfare, and could lead to the evolution of 'sympathy' and 'friendship'.

Once again, there is no reason to suppose that any of these abilities of other primates have been removed in the course of human evolution. On the contrary, given that the major difference between pongid and hominid evolution seems to have been the reliance of hominids on collective hunting – together with a possible increase in direct inter-group aggression and warfare – the expectation should be that humans will be in possession of highly-sophisticated adaptations for coordination to mutual advantage, including prediction, signalling, coalition-formation and sympathy. And this seems to be the case: natural selection has taken the relatively rudimentary predictive abilities possessed by other primates and given them a massive upgrade in humans.

Like some other creatures, humans have a suite of adaptive heuristics that attribute goals and beliefs to others in order to predict and interpret their behaviour – a 'theory of mind'.[210] While other primates, such as vervet monkeys, are capable only of zero- or perhaps first-order predictions (he thinks that…), humans seem capable of anything up to fifth- or sixth-order (he thinks that I think that he thinks that she thinks that they are going to…).[211]

'Theory of mind' emerges in an increasingly well-charted developmental sequence. As Griffin and Baron-Cohen report: from one to three months, infants display expectations about the movement of physical objects, and

---

[210] Pitchford, 2001. 'Theory of mind' is also referred to as: "mind reading" Baron-Cohen, 1995, "social intelligence" Humphrey, 1976, "Machiavellian intelligence" Byrne and Whiten, 1988; Whiten and Byrne, 1997, "political intelligence" Orbell, et al., 2000, and "the intentional stance" Dennett, 1987.
[211] Seyfarth and Cheney, 2002.

are 'surprised' when their expectations are confounded; from three to six months, infants distinguish the 'biological' motion of animate objects from the motion of inanimate objects; from twelve months, infants are capable of following others' eye-gaze, using contingent reaction to infer animacy, and begin to use mental state terms when speaking; from eighteen months, infants begin to attribute goals and intentions to others' actions (and use them to correct for others' mistakes when imitating their behaviour), use others' intention to infer the meaning of words, and come to understand that others may have desires different from their own; from age two, infants are capable of engaging in pretence; from age four and a half, children are capable of recognising false beliefs (that others may have beliefs different from their own); and by age six, children are capable of full belief-desire description and prediction of the behaviour of themselves and others.[212]

Neuroscience is also beginning to home in on the neural location of the various components of 'theory of mind'. Neurons in and around the "superior temporal sulcus" – a subset of the visual system responsible for the detection of motion – are specialised for detecting "biological motion" (such as "walking dancing and throwing"), "goal-directed" movement (such as "reaching, grasping, picking, [and] tearing"), as well as the detection of head- and eye-direction which are required for recognising shared attention, gaze direction, and cues of submission. The "orbitofrontal-amygdala-temporal circuit" has been implicated in several models of social cognition in monkeys, apes, and humans, including attributing intentions from eye movement, the recognition of facial expression, especially fear. And neuro-imaging studies have led some researchers to conclude that "the areas bordering the anterior cingulate and the medial frontal cortex" is "the locus of the mechanism that represents the mental states of self and other".[213] And some or all of these mechanisms can be damaged or absent, as is the case with people with the heritable condition of autism.[214]

---

[212] Griffin and Baron-Cohen, 2002, pp86-90. See also: Baron-Cohen, 1995.
[213] Griffin and Baron-Cohen, 2002, pp86-90, pp96-104.
[214] Baron-Cohen, 1995.

Systems for generating expectations about others' behaviour go hand-in-hand with systems for reducing uncertainty about behaviour through signalling. In addition to the usual primate 'body language' and facial expression of emotion,[215] humans have of course evolved the recursive, combinatorial, vocal signalling system known as language. Noam Chomsky demonstrated that language relies on a basic innate competence, and Steven Pinker has successfully argued that this innate competence couldn't be anything other than a Darwinian adaptation.[216]

The increased reliance on teamwork during the course of human evolution seems to have increased the stake that individuals have in the well-being of other members of the team. This has led to a predictable growth in the human capacity for 'sympathy' – the tendency to detect and attempt to alleviate the distress of others.[217] The investigation of the brain mechanisms responsible for sympathy is well under way.[218] Against this background, friendship can be seen as individualised, iterated mutualism, in which individuals have a significant stake in one another's continuing welfare.[219] The importance of this type of cooperation may explain why the absence of friendships is a major source of anxiety, stress, and ill-health.[220]

Advanced 'theory of mind' and language have opened up a wide range of novel opportunities for cooperation among humans (and are good candidates for explaining the difference between the societies of humans and other animals).

---

[215] Darwin was the first to place the study of human communication on a naturalistic footing by emphasising the continuity in form and function of the mechanisms underlying "body language", gesture and facial expression of emotion from animals to man: Darwin, 1872/1998. His general contention, that humans share a basic suite of emotions and in certain contexts signal them to others with a common repertoire of facial expressions has been generally confirmed: Brown, 1991; Ekman, 1973; Tooby and Cosmides, 1990.

[216] Chomsky, 1965; Pinker, 1994. If, like other forms of communication, language is used to reduce uncertainty, then the meaning of a sentence is a product of the prior expectations of the audience. Baron-Cohen, 1988; Grice, 1989; Lewis, 1969; Sperber and Wilson, 1986.

[217] Hoffman, 2001; Preston and de Waal, 2002.

[218] Decety and Chaminade, 2002, p127, reports on a "mass of converging evidence from developmental psychology and cognitive neuroscience" concerning the neural basis of sympathy.

[219] Tooby and Cosmides, 1996.

[220] Wilkinson, 2001.

For example, as Leda Cosmides and John Tooby note, the ability to generate "hypotheses about what other individuals are likely to value" together with the ability to signal which particular objects one values (by pointing, looking, saying), vastly increases the scope of social exchange. Instead of being restricted to fairly limited and stereotypical exchanges – such as those that take place between cleaner-fish and their clients – humans can identify, and therefore cooperate with, individuals that share any of a wide-range of interests.[221] This ability is presumably a necessary pre-requisite for the widespread exchange of goods and services in modern (that is, post-Pleistocene) societies.

Theory of mind and language make it possible for expectations about behaviour to be communicated to, and shared by, a large number of individuals – a process referred to by the sociologists Peter Berger and Thomas Luckmann as "the social construction of reality".[222] Berger and Luckmann's account of 'social construction' runs as follows: Humans face the problems of how to coordinate their actions over space and time. Solutions include facial expression, language and other signalling systems, adoption of predictable roles and routines, and use of artefacts such as calendars and watches. These solutions to coordination problems constitute shared expectations about how others will behave; thus social constructions can be understood as the arbitrary conventions used to solve coordination problems. Established conventions are called "institutions". Once in place, others can come to adopt them; when children do so it is called "socialization".[223] These social constructions constitute "social facts"; and it is the study of these social facts that marks sociology out as a distinct discipline.[224] Berger and Luckmann conclude

---

[221] Cosmides and Tooby, 1989.

[222] Berger and Luckmann, 1966.

[223] "*A* watches *B* perform. He attributes motives to *B*'s actions"; as a result, "each will be able to predict the other's actions. . . . This relieves both individuals of a considerable amount of tension . . . [and] makes possible a division of labour between them". Berger and Luckmann, 1966, pp74-5.

[224] Berger and Luckmann, 1966, p204. Berger and Luckmann continue as follows: Sometimes the original reason for the institution is forgotten, and so new reasons are invented to explain and justify it (this is called "legitimation"). These reasons form the myths and legends of a society. (Myths in the service of the partial interests of a particular sector of society are called "ideologies".) A society must defend its myths against those of sub-cultures, and of other societies. For the purposes of analysing the behaviour of a society and its members, these conventions, or "social facts" can be treated as things, and as such compose the "reality" in which

that "Man is biologically predestined to construct and to inhabit a world with others"; but they refrain "from any causal or genetic hypotheses" about – or "assertions about the ontological status" of – social constructions.[225] The evolved psychology of 'theory of mind' (and language) provides the missing link between biology and sociology by explaining how it is that humans come to generate and share expectations about behaviour. In this way, 'evolutionary psychology' explains and thereby incorporates 'social constructionism'.[226]

Note that, because not all aspects of social life are coordination problems, not all aspects of social life are coordination solutions (that is, social constructions) contrary to what some proponents of social constructionism might argue. To paraphrase George Williams, we might say that: "Social construction is a special and onerous concept that should not be used unnecessarily, and a social phenomenon should not be called a construction unless it is the arbitrary solution to a coordination problem maintained by shared expectations."

At higher-orders, 'theory of mind' also makes 'common knowledge' possible. As Chwe explains: "We say that an event or fact is common knowledge among a group of people if everyone knows it, everyone knows that everyone knows it, everyone knows that everyone knows that everyone knows it, and so on."[227] Common knowledge is necessary for certain forms of coordinated social activity. For example, in the story of the Emperor's new clothes, everyone knew that the emperor was naked, but they didn't know that everyone else knew, and no-one wanted to be the first and only person to laugh. Even if everyone knew that everyone else knew, they may nevertheless be unsure about whether everyone else knew that everyone else knew; hence they would still be reluctant to be

these societies operate. The book goes on to discuss the various ways in which the details of these "social processes" are elaborated across cultures and across history.

[225] Berger and Luckmann, 1966, p34, p204.

[226] Presumably, to the extent that people benefit from others adopting the solution to coordination problems, it is in their interest to apply sanctions to those that fail to do so. However, I am not aware of any adaptationist work that has been done on the application of sanctions to deviants, as opposed to the application of punishment to free-riders. It would be interesting to investigate attitudes to punctuality and other aspects of etiquette from this perspective.

[227] Chwe, 2001, p10.

the first and possibly only one to laugh. In the story, the child that first points to the naked emperor and laughs creates common knowledge in the audience, because, through his actions, everyone now knows that everyone knows that everyone knows that the emperor is naked, and so they can all laugh together. Chwe goes on to investigate how certain public rituals, precedents, forms of communication and other institutions serve to establish common knowledge in social groups, and thereby coordinate their behaviour.[228] Importantly, for the argument of this chapter, Chwe notes that, "Common knowledge relies on people having a 'theory of mind,' an ability to understand the mental states of others. . . ".[229]

Finally, the possession of adaptations for coordination to mutual advantage would presumably explain why cooperative labour, and division of labour are human universals.[230]

In summary, there is no evidence to suggest that natural selection has removed any adaptations for coordination from the human line. On the contrary, the evidence suggests that humans have adaptations for predicting the behaviour of others, for signalling their intentions, for generating shared expectations about behaviour, and for caring for team-mates. And they use these abilities to form interest groups (teams), coordinate behaviour, and create, and adopt social norms.[231]

**Reciprocity**

As we saw in the previous chapter, extending mutualism to incorporate delayed exchanges creates additional problems that, in theory, can be solved by a tit-for-tat-like strategy. Although few non-human species

---

[228] Kurzban, Tooby and Cosmides have investigated the tendency of humans to form interest groups on the most arbitrary of characteristics – such as clothing and skin colour – which is perhaps merely because such cues are public, and hence common knowledge: Cosmides and Tooby, 1992; Kurzban, et al., 2001; Tooby and Cosmides, 1988; Tooby and Cosmides, 1996.
[229] Chwe, 2001, p17.
[230] Brown, 1991, p137.
[231] To quote Daniel Dennett, folk psychology is a "predictive strategy" that enables us to "cooperate on multi-person projects, learn from each other, and enjoy periods of local peace." Dennett, 1987, p11. And, as Schelling puts it, coordination involves "nothing more nor less than intuitively perceived mutual expectations." Schelling, 1960, p71. See also: Lewis, 1969, p27.

seem to have mastered delayed mutualism, various rudimentary components of reciprocity seem to be present in primates. In the course of human evolution, increased reliance of hominids on hunting seems likely to have provided the impetus for the evolution of adaptations for full-blown reciprocity. Hunting is a risky business that, compared to gathering, involves considerably more luck than effort. Successful hunts also tend to generate more food than can be consumed at one time by an individual hunter. This combination of risk and surplus creates the ideal conditions for the evolution of reciprocity, where a lucky hunter can 'lend' an unlucky hunter a meal, on the condition that should the situation reverse, the favour will be returned.

*Cheater-detection*

As we have seen, successful cooperation of this kind requires that free-riders ('cheats') are detected and punished. In order to test whether human psychology comprises adaptations designed to detect cheats, the evolutionary psychologist Leda Cosmides gave subjects different versions of the Wason selection task.

In this task, subjects are presented with a conditional rule of the form, *If P then Q*. Subjects are then presented with four cards with information about *P* on one side and information about *Q* on the other. The visible sides of the four cards display the information: *P*, *not-P*, *Q*, and *not-Q*. The subject's task is to indicate which of these cards must to be turned over in order to determine whether or not the rule has been violated. The correct answer is that cards *P* and *not-Q* should be turned over. The *P* card must be turned over in order to determine whether *Q* is on the other side; if *Q* is not on the other side, then the rule has been violated. And the *not-Q* card must be turned over in order to determine whether it has *P* on the other side; if it does have *P* on the other side, then the rule has been violated. (Turning over the *not-P* and the *Q* cards will not indicate whether the rule has been broken.)

Prior to Cosmides' work, it was known that very few people (less than 10%) choose the logically correct cards; and that performance varied across different rules. It had been assumed that this variation in performance could be explained with reference to the familiarity of the rule; in other words, people did better on rules with which they were familiar. Cosmides wondered whether the variation might be the result not of familiarity, but of the content of the rule. In particular, Cosmides suspected that people might be better at detecting violations of *social contract* type rules – of the kind "If you take the benefit, you must pay the cost" – than they are at detecting violations of logically equivalent rules that do not have the structure of a social contract.[232]

To this end, subjects were asked to look for violations of the same unfamiliar rule: "If a man eats cassava root, then he has a tattoo on his face". One group of subjects was told that this was a *descriptive* rule (in other words, men that live where cassava grows happen to have tattoos). Another group was told that this rule was a *social contract* (in other words, cassava root constituted a benefit, and the tattoo was a cost that had to be paid in order to eat it). In this experiment, 23% of subjects in the *descriptive* condition chose the logically correct cards ($P$ and *not-Q*); whereas a whopping 73% chose the correct cards in the *social contract* condition. These US results[233] have been replicated in Germany,[234] Hong Kong,[235] Italy,[236] the United Kingdom,[237] and the Shiwiar, a hunter-horticulturalist band living in the Ecuadorian Amazon.[238] Subsequent experiments showed that performance on *unfamiliar* social contracts was better even than performance on *familiar* descriptive rules – supporting the idea that content and not familiarity explain the difference.[239]

Cosmides also found that it was possible to manipulate, in predictable ways, the kind of errors that people made. In another experiment,

---

[232] Cosmides and Tooby, 1992.
[233] Cosmides, 1989; Cosmides and Tooby, 1992.
[234] Gigerenzer and Hug, 1992.
[235] Cheng and Holyoak, 1985.
[236] Girotto, et al., 1992.
[237] Johnson-Laird, et al., 1972.
[238] Sugiyamaa, et al., 2002.
[239] Cosmides, 1989; Cosmides and Tooby, 1989.

subjects were asked to look for violations of 'switched' social contracts that took the form "If you pay the cost, then you take the benefit". In this case, the logically correct answer remains *P* and *not-Q* (that is, one should look for people who paid the cost but did not take the benefit). But if people do not reason 'logically', but continue to for cheats (people who took the benefit without paying the cost) then they will choose *not-P* and *Q*. And indeed, in this round of experiments, 71% of subjects chose the latter, 'looking for cheaters', set of cards.[240] Gigerenzer and Hug had a similar result when they switched perspectives on the same social contract rule. They found that subjects chose different cards when looking for violations of the rule "If an employee works on the weekend, then that employee gets a day off during the week" according to whether they took the perspective of employer or employee. 'Employers' tended to look for cheating by 'employees' (*not-P* and *Q*); and 'employees' tended to look for cheating by 'employers' (*P* and *not-Q*).[241]

Cheng and Holyoak have argued that the finding that people are better at reasoning about social contracts than about descriptive rules is not sufficient to conclude that people have a special 'social contract' mechanism. It may be that people are just better at 'social rules' in general (including, for example, 'permission' rules that take the form "If an action is to be taken, then the precondition must be met").[242] Cosmides tested this objection by giving subjects permission rules, some of which were 'embedded' in a 'social contract' context. Again, only 21% of subjects chose the correct cards in the straight 'permission' condition, whereas 75% chose the correct cards in the 'permission *qua* social contract' condition – indicating that there was something special about social contracts that distinguished them from 'social rules' in general.[243]

---

[240] Cosmides, 1989; Cosmides and Tooby, 1989.

[241] Gigerenzer and Hug, 1992.

[242] Cheng and Holyoak, 1989.

[243] Cosmides, 1989. Subsequently, researchers have investigated whether, in addition to social-purpose mechanisms for social contracts, the 'domain' of social reasoning may also include special mechanisms for promises, warnings, threats, and precautions. Fiddick presents evidence for cognitive adaptations for dealing with hazards: Fiddick, forthcoming; Fiddick, 2003.

Laurence Fiddick reports that *deliberate* versus *accidental* violation of a rule also effects subjects' ability to identify violations of social contracts, but not of precautions. 80% of subjects chose the logically correct set of cards for intentional violations of social contracts; whereas only 45.5% chose the correct set of cards for unintentional violation of social contracts. Subjects chose the correct cards 82.6% and 79.2% of the time when dealing with intentional and unintentional violations of precaution rules. In other words, the intention to deceive triggers the 'cheat detection' mechanism; but intention is irrelevant to whether or not the 'hazard avoidance' mechanism is triggered. Fiddick notes that advanced 'theory of mind' would seem to be a pre-requisite for such attributions of intention, and hence for this kind of sophisticated cooperation.[244]

More recently, the theory that 'cheater-detection' is the product of a specific adaptation has been supported by the finding of a brain-damaged patient in whom reasoning about social contracts, but not precautions, is selectively impaired.[245]

*Behavioural economics*

Indirect evidence for adaptations for reciprocity comes from behavioural and experimental economics. This literature consists of illustrations of how everyday reasoning and decision-making departs in systematic ways from the predictions of standard applications of rational choice theory.[246] For example, it has been shown repeatedly that people tend to cooperate in one-shot prisoner's dilemmas when standard rational choice theory predicts that they will defect.[247] And it has been shown that people are willing to incur a cost in order to punish cheats, whereas standard rational choice theory suggests that such 'spite' is irrational and that it pays to 'let bygones be bygones'.[248]

---

[244] Fiddick, forthcoming; Fiddick, 2003.
[245] Stone, et al., 1997; Stone, et al., 2002.
[246] Kahneman, et al., 1982. For discussion, see: Samuels, et al., 2002; Samuels, et al., 1999.
[247] For reviews of the literature, see: Caporael, et al., 1989; Mansbridge, 1990; Ostrom, 1998.
[248] Fehr and Gachter, 2002.

These findings represent a puzzle for standard individual-level applications of rational choice theory.[249] The problem is not only why a rational agent would act in this way, but *how.* For example, how does a rational chooser resist the "temptation" to maximise his short-term interests and instead "bind" himself to courses of action that will maximise his welfare in the long-term? How can an agent rationally choose to be irrational – reducing the number of options available to him, or committing himself to things that would be counter to his interests were he to do them? Proposed solutions to these problems often invoke emotions as exogenous factors that interrupt the process of rational cognitive decision-making. This analysis suggests that emotions act like 'spanners in the rational works', transporting individuals far beyond where mean-spirited rationality could take them, to dizzying new heights of cooperation and ('genuine') altruism. . .[250]

However, altruism and spite cease to be anomalies, and indeed make perfect sense, if, as evolutionary theory suggests: human reasoning is not the product of rational choice, but is instead the product of evolved decision-rules, some of which have been designed to follow a tit-for-tat-like strategy;[251] and people tend to rely on these decision-rules, play single prisoner's dilemmas *as if* they were iterated, and remain relatively unmoved by experimental manipulations (such as assurances of anonymity and single iteration).[252] As the behavioural economist Thomas S. Ulen has noted: "evolutionary psychology offers the best explanation as to why human beings are imperfectly rational decisionmakers".[253]

An additional advantage of the evolutionary approach is that it allows us to tease apart various different phenomena that have been grouped under

---

[249] Frank, 1988; Nesse, 2001b; Schelling, 1978; Schelling, 1960,

[250] See, for example, Nesse, 2001a.

[251] As we have seen, evolutionary game theory makes different predictions about individual behaviour to classical game theory. In evolutionary theory, *natural selection*, not the individual, plays the role of rational chooser. Natural selection 'chooses' between alternative designs of organisms; and selection seems to have favoured a tit-for-tat-like design in the context of repeated prisoner's dilemmas.

[252] Price, et al., 2002 adopts this approach to explain various features of human punitive sentiment. What's more, people often make the kinds of 'mistakes' an evolutionist would expect: Palameta and Brown, 1999.

[253] Ulen, 2001, p936. See also: Gigerenzer and McElreath, 2003; Posner, 1998.

the heading of 'commitments' – phenomena such as initiation rites, wedding rings, and expensive business premises.[254] Reciprocity depends on, amongst other things, reducing uncertainty about a person's mobility ($w$) and discount rate ($d$). Therefore, in addition to "irrational" niceness and revenge, we should expect potential cooperators to attend to, and make available, honest signals of low mobility and low discount rate as a means of distinguishing themselves from less trustworthy individuals. These signals make it clear that it would be *in the interests of* the signaller to reciprocate cooperation; they do not signal unconditional cooperation, or a new and distinct form of altruism, as some of the literature suggests.[255]

The finding that reciprocity is a universal feature of human societies also provides circumstantial evidence for the existence of human adaptations for reciprocity. As Donald Brown reports, everywhere people "use reciprocal exchanges, whether of labor, of goods, or services, in a variety of settings. Reciprocity – including its negative or retaliatory forms – is an important element in the conduct of their lives."[256] What's more, the presence of adaptations for reciprocity might also help to explain cross-cultural variation in the cooperative practices of different societies. We saw above that increased reliance on hunting seems to have been an important impetus for the evolution of reciprocity. Hunting is a relatively risky activity – however lucky you were today, you may be less lucky tomorrow – and so pooling and sharing meat with fellow hunters acts as a kind of mutually-beneficial social insurance. In contrast, there is relatively little to be gained from entering into reciprocal exchanges of resources such as plant food. The amount of plant food gathered is largely a matter of effort, not luck, and so any redistribution would have merely taken from those who work hard and give to those who work less hard, producing no net benefit over time.[257] This aspect of the underlying logic of cooperation perhaps explains why capuchin monkeys and

---

[254] Nesse, 2001b.
[255] Nesse, 2001a
[256] Brown, 1991, p137. See also: Sahlins, 1972 and Singer, 1994, pp63-92.
[257] Cosmides and Tooby, 1992.

chimpanzees share meat but not plant food.[258] Assuming that human adaptations for reciprocity make a similar distinction between 'luck-based' and 'effort-based' goods, we should expect to see a similar pattern in the sharing practices of human societies. And this seems to be the case. Anthropologists have found that 'hunters' tend to put a premium on redistributive food-sharing, whereas 'gatherers' do not; that luck- and effort-based goods are distributed differently within the same societies; and that changes in the economy and ecology of hunter-gather societies lead to predictable changes in their sharing practices.[259]

In summary, people display many of the predicted features of a veteran reciprocator. And the notion that people are intuitive reciprocators, designed to cooperate under ancestral conditions, can explain how and why people depart from the predictions of classical rational choice theory. But there is a lot to be done to map out the full software package of human reciprocal altruism; evolutionary-minded research into other aspects of the biological basis of reciprocity – such as trust,[260] the intrinsic rewards of cooperation,[261] gratitude,[262] behavioural and facial cues of deception,[263] enhanced memory for the faces of cheats,[264] guilt,[265] and reputation[266] – remains at an early stage.

---

[258] McGrew and Feistner, 1992. de Waal reports that food typically shared between chimpanzees has the following characteristics: "1) Highly valued, concentrated, but prone to decay. 2) Too much for a single individual to consume. 3) Unpredictably available. 4) Procured through skills and strength that make certain classes of individuals dependent on others for access. 5) Most effectively procured through collaboration." de Waal, 1996, pp144-5.

[259] Cosmides and Tooby, 1992.

[260] Hormonal, physiological and neurological work on the biological basis of 'trust' suggests that the ability to 'open oneself up' to the possibility of exploitation is mediated by, and may be an evolutionary outgrowth of, the mechanism in mammals that enables mothers to tolerate the proximity of offspring. Grimes, 2003.

[261] Rilling, et al., 2002.

[262] Bonnie and de Waal, 2004.

[263] Various aspects of body language – such as facial expressions, posture, timbre of voice, respiration rate, or other cues of arousal such as sweating – may display anxiety and preparation for escape. Anxiety presumably reflects the anticipated costs of being caught cheating. This may explain why people are more inclined to trust and cooperate with people they have had a chance to inspect for such cues during face-to-face interactions. Ekman, 1985; Ekman, 1973; Ekman, et al., 1976.

[264] Mealey, et al., 1996; Pinker, 1997, pp272-4.

[265] Ketelaar and Au, 2003.

[266] Nowak and Sigmund, 1998.

**Conflict resolution: costly-signalling**

As we saw in the previous chapter, many social animals, including primates, resolve potentially injurious disputes through the display of costly or 'hard to fake' anatomical or behavioural signals that reliably indicate 'fitness' or "resource holding power". In some social species, the results of these interactions are used to form dominance hierarchies, that further reduce the incidence and costs of conflict. And, in addition to imposing a cost on the signaller, these signals can sometimes provide a benefit to the audience. Again, there is no reason to suppose that these adaptations have been erased over the past 6 million years. On the contrary, the evidence suggests that humans have retained the typical primate apparatus – anatomical, hormonal and psychological – for settling disputes and resolving conflict by means of the display of costly signals.

In men, as in other male mammals, genes on the Y chromosome turn embryonic asexual gonads into testes. Testes then produce a range of hormones, including testosterone, that masculinise the default female body plan. This difference in body plan and levels of circulating hormones leads to stereotypical male behaviour in boys – such as greater competitive 'rough and tumble' play. At puberty, testosterone activates changes that lead to increased size, height, muscle mass; to the development of secondary male sexual characteristics, such as growth of the genitals, larynx, jaw, beard, body hair; and to increased 'combativeness' and interest in the opposite sex. Once the basic male body plan is in place, circulating testosterone induces men to seek opportunities to compete with other men for status and resources.[267]

Whether it is in chickens, other primates or humans, the role of testosterone in these contests is the same. Testosterone rises in anticipation of a challenge, thereby boosting "coordination, cognitive performance, and concentration". Ethologically speaking, dominance

---

[267] Mazur and Booth, 1998. Levels of circulating testosterone peak in late adolescence and early 20s, and decline slowly thereafter, creating what Wilson and Daly call "the young male syndrome". Wilson and Daly, 1985.

contests in primates often involve "staring down" one's opponent: "[W]hen one animal is obviously the more powerful and assertive or the other appears fearful . . . a simple stare by the powerful animal, followed by the fearful animal's eye aversion or by its yielding something of value (perhaps food or a sitting place), may suffice. Sometimes a single contest is all that is needed to allocate ranks or to verify a pre-existing rank relationship, but often the outcome is settled only after a series of contests."[268]

After the contest, levels of testosterone remain high in the winners – causing them to feel elated, and to seek out further opportunities to compete – but fall in losers – causing them to feel dejected, and to withdraw from future challenges. Mazur and Booth comment:

> Within hours of this outcome, we assume Ego (the loser) experiences a drop in T[estosterone], reducing his assertiveness, diminishing his propensity to display the dominant actions associated with high status, and increasing his display of such submissive signs as stooped posture, smiling, or eye aversion (Mazur 1985). Faced with a new dominance encounter, Ego is more likely than before to retreat or submit. On the other side Alter, the winner, experiences the opposite effects: rising T, increased assertiveness, and a display of dominant signs such as erect posture, sauntering or striding gait, and direct eye contact with others. Alter may seek out new dominance encounters and is bolstered to win them.[269]

---

[268] Mazur and Booth, 1998. Mazur and Booth provide the following description of a typical face-to-face dominance contest: "Consider two strangers, Ego and Alter, whose eyes meet, by chance, across a room. Let us say that one of the strangers, Ego, decides to hold the stare. The chance eye contact now becomes a dominance encounter. Ego's stare makes Alter uncomfortable. Alter may then avert his eyes, thus relieving his discomfort while, in effect, surrendering, or he may stare back, making Ego uncomfortable in return. In the latter case, the staredown would continue, with each individual attempting to outstress the other until finally one person succumbed to the discomfort (and the challenger) by averting his eyes. The matter thus settled, the yielder usually avoids further eye contact, though the winner may occasionally look at the loser as if to verify his victory." (p359).

[269] Mazur and Booth, 1998, p359. Mazur and Booth also refer to a study of "sports fans who are not themselves participants in the physical competition": "Following the 1994 World Cup soccer

The result is, as Mazur and Booth note: "Like all primates, humans in face-to-face groups form themselves into fairly consistent dominance/status hierarchies so that higher-ranked members have more power, influence, and valued prerogatives than lower-ranked ones."[270]

Men seem also to compete for status through costly signals of altruism, such as bravery, generosity and 'public service'. For example, in Ronald C. Johnson's analysis of the *Attributes of Carnegie Medalists Performing Acts of Heroism and of the Recipients of Those Acts*, "About 92% of the [total of 676] acts of heroism were performed by males." And whereas women tended to rescue "relatives or people they knew", a higher proportion of men "rescued people they did not know".[271]

To take another example, several anthropologists have advanced sexual-selection of costly fitness-indicators as explanations of generosity, charity and their many cultural manifestations (such as potlatch feats, bonanzas, and festivals).[272] Hawkes et al, for example, argue that the sharing of big game among the Hazda does not conform to an 'exchange model' as would be predicted by mutualism or reciprocity. Unsuccessful hunters do not get less than others from kills, nor does there appear to be any exclusion or punishment of unsuccessful providers. On the contrary, a carcass appears to be a 'public good' simply because the hunter cannot exclude others from a portion, or it would not be cost-effective to do so. Indeed, the hunter does not seem to exert any influence over who gets what. Instead, sharing is conducted on a 'tolerated theft' or 'demand sharing' model. Only after the meat has been distributed do people 'own' it and can it become the subject of reciprocal exchanges.[273] If big game is a public good, then the question is what incentive do the hunters have to

---

tournament in which Brazil beat Italy, T[estosterone] increased significantly in Brazilian fans who had watched the match on television, and decreased in Italian fans."

[270] Mazur and Booth, 1998, p359.

[271] Johnson, 1996. Johnson correctly notes that the rescuing of strangers is not explained either by kin selection or reciprocity, but concludes that these acts of altruism therefore constitute evidence for group selection. He does not consider that they are perfect examples of costly signalling.

[272] If altruism is sometimes used to signal superior status, it would explain why, as George Orwell noted, "The man receiving charity always hates his benefactor; it is a fixed characteristic of human nature." (http://www.weeks-g.dircon.co.uk/quotes_by_author_o.htm).

[273] Hawkes, et al., 2001.

provide it? Why incur the costs of hunting, why not let others do the work and just turn up at the kill site? The answer, argues Hawkes, is that although the hunter does not exert any influence over the kill, everyone knows who killed it. Indeed, a big dead animal is especially good at broadcasting this information as it attracts individuals from far and wide. Thus the hunter benefits in terms of increased popularity and status among men, and increased sexual access to women. Hawkes reports that successful hunters are more often named as lovers, and have more surviving offspring. As such, Hawkes considers big game hunting as a costly signal of genetic quality and condition; it is the product of male status competition, but a competition that generates positives benefits for its audience.[274] Selection of such costly altruistic signals is consistent with the observation that 'generosity' is universally favoured among leaders.[275]

Human status competition seems to be an instance of the general primate model of dominance hierarchies in which individuals compete for status by solving collective-action problems, especially those involved in supporting other cooperative schemes – coordinating behaviour, distributing the spoils of collective action, punishing cheats, intervening to resolve conflicts, and so on – and in which the audience support high-status individuals (and their allies) on the basis of their ability to perform these functions. As de Waal observes: "In short, the dismantling of despotic hierarchies in the course of hominoid evolution brought an emphasis on leadership rather than dominance, and made the privileges of high status contingent upon services to the community (such as effective conflict management)."[276]

While there seems little doubt that men rely on a variety of physical and behavioural cues – of size, speed, strength, agility, bravery, charity – to settle disputes and compete for status (and seem to have invented a variety of conventional and highly ritualised ways of conducting these

---

[274] Smith et al have applied a similar analysis to turtle hunts. They point out that turtle hunting is inefficient relative to netting sardines, but that the successful capture of a turtle is the occasion for a feast of up to 200 people. Successful hunters gain in status and popularity, as well as sexual access to women. Smith, 1999; Smith and Bleige Bird, 2000.

[275] Brown, 1991. pp137-40. For more on leaders and leadership, see Foss, 2000.

[276] de Waal, 1996, p132.

contests, see below) there remains a question over the extent to which male costly signalling includes displays of intelligence, aestheticism, and creativity – the human equivalent of the peacock's tail or the nightingale's song. While there is little as yet by way of direct test of this idea, there is considerable supporting circumstantial and anecdotal evidence. The evolutionary psychologist Geoffrey Miller points out that in every cultural sphere – such as art, music and literature – men are responsible for around 10 times as much cultural production as women; that male cultural production peaks at the same time that testosterone and mating effort peaks (that is, during early adulthood); and that displays of intelligence, wit, creativity, form an important part of human courtship.[277]

The prediction implicit to costly-signalling theory is that, in the context of male-male competition, men should attend to, be intimidated by, and defer to cues of high-testosterone and high status – such as physical prowess, bravery, generosity, intelligence, creativity – in other males. Unfortunately, perhaps because the answer seems so obvious, this prediction has yet to be rigourously tested by contemporary evolutionary psychology. However, because male-male competition is ultimately about access to females, and because some male costly signals are designed to appeal directly to choosy females, a parallel prediction is that to the extent that women prefer to mate with 'winners' they will find cues of dominance and status sexually attractive. This prediction has been put to the test; and it has been found that women do find such cues attractive.[278] As Gregersen concludes from a study of almost 300 cultures: "for women the world over, male attractiveness is bound up with social status, or skills, strength, bravery, prowess, and similar qualities."[279] Not surprisingly, high-testosterone males also report more sexual partners.[280]

Research on costly signalling has tended to focus on male displays because evolutionary theory predicts that males will put more effort into

---

[277] Miller, 2000a.
[278] Buss, 1994; Ellis, 1992; Miller, 1998.
[279] Gregersen, 1982.
[280] "Although sexual activity in both sexes correlates with T levels . . . dominance appears to correlate with sexual attractiveness and number of sex partners in men but not women." Townsend, 1998.

producing large conspicuous displays to as many mates as possible. But to the extent that males invest anything at all in females and their offspring, we should expect males to be choosy too, and for females to compete to display the characteristics that males look for in mate. So whereas men compete to display attributes that women regard as attractive – such as correlates of good genes, health, intelligence, generosity, status and access to resources – women compete to display attributes that men find attractive – fertility (which has to do with underlying genetic quality as manifest in symmetry, as well as cues of youth and health) and, in the context of long-term mating, fidelity, kindness and intelligence. Together, these cues constitute cross-cultural standards of female beauty and 'feminine virtue'.[281]

Finally, in humans, as in other primates, cues of submission consist of the opposite of cues of dominance; and these signals are further ritualised in the context of the greeting, manners, etiquette and protocol used to lubricate the negotiation of formal dominance hierarchies.[282] And the inability to recognise and respond to cues of submission is one of the symptoms of psychopathy.[283]

**Conflict resolution: property**

In the previous chapter we saw that many species resolve disputes over territory, property and mates through the recognition of prior 'ownership'. Again, there is no reason to suppose that adaptations for recognising and respecting property have been removed during the course of human evolution. As Dennett notes, "many species manifestly do exhibit innate territoriality, and it is hard to think of what force there might be to remove such a disposition from our genetic makeup."[284]

---

[281] Buss, 1994; Miller, 2000a; Singh, 1993; Symons, 1995; Symons, 1979. This research presents evidence that a 0.7 waste-to-hip ratio, lighter-than-average skin colour, symmetry, low testosterone, and nulliparity are among the cues of female youth and fertility that form the basis of universal standards of female beauty.

[282] Ekman, 1985.

[283] Blair, 1997.

[284] Dennett, 1995, p487. Or, as Beagle notes, "Psychologically the acquisition impulse (or instinct), as it has been called, seems to be rooted very deeply and demands recognition. The existence of this instinct is a refutation of the opinion which considers the rights of property as

Ever since Darwin, there have been theorists who have argued that humans have some kind of 'acquisitive instinct' or 'hoarding instinct'. But these accounts have been correctly dismissed as too vague and lacking in detail to be useful in explaining the human psychology of property.[285] Besides, these accounts tend to refer only to the acquisition of goods, not to their recognition and respect by others. Unfortunately, there has been little attempt to test any modern evolutionary-game-theoretic predictions about the psychology of property.

One possible exception comes from the 'heuristics and biases' literature. The "endowment effect" refers to the finding that people value things that they own more than things that they don't own. In other words, they are willing to sell things for more than they would be willing to buy them. This represents something of an anomaly for standard applications of rational choice theory, because items 'cost' the same independent of whether they are in your possession or not. Behavioural economists attempt to explain this anomaly with "prospect theory" – the idea that losses loom larger than gains.[286] But this seems to do little more than relabel the phenomena. It may well be that the tendency to over-value goods in your possession is a reflection of the willingness to defend one's property against usurpers.

The psychology of 'ownership' also seems to govern human thinking about sexual proprietariness. In their article *The man who mistook his wife for a chattel*, Daly and Wilson document the various ways – from sexual jealousy to legal systems – in which men have tended to think of women as if they were property, or at least, have used the same kind of approach to settle disputes over 'ownership'.[287]

---

something conventional and artificial. Even with animals one finds the recognition of *meum* and *tuum* and that not only with regard to other individuals like the young of the family, but equally with regard to things. The bird claims the nest and even the whole tree as its own, and the dog defends its kennel with its life." Quoted in Litwinski, 1942, p36.

[285] Beaglehole, 1931; Dittmar, 1992.

[286] Kahneman and Tversky, 1979.

[287] Wilson and Daly, 1992.

The existence of a system of adaptations for recognising and respecting prior ownership would explain why people everywhere "have concepts of property, distinguishing what belongs – minimal though it may be – to the individual, or group, from what belongs to others",[288] and why some form of territoriality seems to be a feature of all human societies.[289]

This review of human conflict resolution draws our review of the evidence for adaptations for cooperation to a close. This brief review of the relevant literature gives us no reason to reject the default hypothesis that humans possess typical primate adaptations for cooperation, modified by several millions years of hunting on the African Savannah. We may tentatively conclude that humans remain in possession of adaptations for cooperation that include: adaptations for maternal and paternal investment, and for preferentially associating with and aiding kin; high-order theory of mind, language, and sympathy; adaptations for reciprocity, including assessing the trustworthiness of potential collaborators, and for detecting and punishing cheats; adaptations for settling disputes and forming dominance hierarchies through costly (and sometimes altruistic) signalling; and adaptations for recognising and respecting prior ownership.

**Extending the cooperative niche**

Before moving on to compare adaptations for cooperation with Hume's moral passions, there is one more evolved faculty to consider. No survey of human social psychology would be complete without discussing the ability that humans have to invent novel means of achieving their goals, and considering what effect this might have on human cooperation.

The evolutionary anthropologists Tooby and DeVore argue that whereas most plants and animals compete and cooperate with one another through naturally-selected genetic strategies, humans have evolved, in

---

[288] Brown, 1991, p140. We might expect the evolved psychology of property to generate different concepts of property under different ecological, economic and technological conditions that altered the mobility, defendability, divisibility, identifiability, and so on, of available goods.
[289] Wilson, 1978, pp107-8.

addition, a sophisticated ability to construct or imagine abstract cause-and-effect models of the world – a "causal or instrumental intelligence" – that enables them to model the effects of their own actions, and thereby devise new ways to achieve their goals. (We might think of this ability as an advanced version of the rudimentary folk physics and tool-using abilities found in other species.)[290] Instrumental intelligence has made it possible for humans to vastly increase the number and type of natural resources they can exploit, and increased their range of habitable environments. And, once invented, these novel manipulations can be pooled and passed onto others thanks to humans' enhanced inferential and communicative abilities ('theory of mind' and language).[291]

If humans can invent new ways of solving problems, then they can presumably invent new ways of solving collective-action problems. This is manifestly the case with a range of tools that can be seen as "extended phenotypes",[292] "outside-the-skin control mechanisms",[293] "extrasomatic scaffolding"[294] that support vast increases in the scope of human cooperation. For example, humans have extended their ability to coordinate their behaviour by inventing: written language, semaphore, morse code, telegraphs, telephones, radio, television, telecommunication satellites, the internet, road signs, maps, compasses, GPS, sundials, watches, clocks, diaries, and calendars. They have extended the scope of exchange and reciprocity through such "technologies of trust"[295] as: CVs, letters of recommendation, money,[296] written contracts, "mechanical cheater-detectors" such as "[c]ash registers tapes, punch clocks, train

---

[290] Whiten, (forthcoming).

[291] "Burrowing animals, underground storage organs, nuts, seeds, bone marrow, birds, fish, molluscs, tool accessible nests, plant foods whose toxins or other inhibiting secondary compounds can be neutralized through processing or cooking, quick animals that must be ambushed, animals whose capture requires close cooperation, tools, or intelligent trickery – all are made accessible by the ability to perform appropriate learned or invented manipulations." Tooby and DeVore, 1987, pp209-10.

[292] Dawkins, 1982.

[293] Clifford Geertz talks about symbols such as "words . . . gestures, drawings, musical sounds, mechanical devices like clocks" as "extragenetic, outside-the-skin control mechanisms . . . for ordering behaviour", for allowing humans to "find their bearings" Geertz, 1973, pp44-5. See also: Berger and Luckmann, 1966.

[294] Clark, 1997.

[295] Wright, 2000, p205.

[296] Money is a "formal token of delayed reciprocal altruism". Dawkins, 1976, p188.

tickets, receipts, accounting ledgers"[297], handcuffs, prisons, electric chairs, CCTV, branding of criminals, and criminal records. And they have extended the scope of conflict resolution through such technologies as medals, games and sports, titles, land registries, and so on.

We might apply the same analysis to the use of "rules". In theory, people could invent and recommend an endless variety of social rules: people must drive on the left; everyone must wear blue hats; murderers must be beheaded; people should drink only boiled water; people must eat bananas on Tuesdays… and so on. In practice, people will approve of and adopt only those rules that further their ends (including the social, cooperative and altruistic ends represented by their various adaptations for cooperation). Let's call solutions to coordination problems 'conventions', solutions to free-rider problems 'laws', and solutions to conflict resolution problems 'regulations'. Thus people approve of conventions, laws and regulations to the extent that they efficiently promote cooperation.[298] And we might add that costly-signalling alpha males shoulder a disproportionately large share of the burden of enforcing these social rules, and receive the approval and support of members of their societies to the extent that they do so efficiently and effectively; and also that social and political reform can be seen as the process whereby society is moved from one stable but sub-optimal set of rules to another more optimal set.[299]

And so, in addition to evolved intuitions about what they ought to do in order to bring about various cooperative outcomes, humans are equipped with a wide array of tools and rules that further these ends. This ability to generate and transmit novel techniques, including the ability to generate and transmit novel solutions to problems of cooperation, provides evolutionists with a rudimentary account and analysis of 'culture'. To

---

[297] Pinker, 1997, p504.

[298] This seems to be the case with conventions. Miller and Bresoff report that "rules are likely to be regarded as legitimate social conventions [*qua* "solutions to social-coordination problems"] if they are judged to be functional in meeting interests common to all members of a social body" Miller and Bersoff, 1988, p367. Posner has applied this analysis to law in general: Posner, 1973; Posner, 1981.

[299] For this last point see: Binmore, 1994a; Binmore, 1994b.

quote Steven Pinker: "Culture can be seen instead as a part of the human phenotype . . . The phenomena we call 'culture' arise as people pool and accumulate their discoveries, and as they institute conventions to coordinate their labours and adjudicate their conflicts."[300] Thus equipped, humans have navigated their way through the stormy waters of the past ten thousand years to generate what we know as human history.[301]

*An aside on cultural transmission and cultural evolution*

At this juncture it is worth mentioning that several theorists have argued that cultural transmission makes cultural artefacts subject to a process of evolution by natural selection;[302] and that cultural evolution provides an alternative explanation for the emergence of morality.

The basic idea of cultural evolution is that packets of cultural information or 'memes' – such as songs, limericks, catchphrases, techniques, habits, ideologies, institutions and so on – may vary slightly, and more memorable or useful memes might get picked up and spread further and faster than other less memorable or useful memes. In this way, the 'meme pool' – the ideas of a particular culture – may come to be full of memes that are good at getting themselves copied. And hence the design of cultural artefacts is not merely the product of human ingenuity, but also of a selective process equivalent to natural selection. As a result of this process, different cultures may come to be dominated by different tools and rules. Societies that come to adopt better memes may thrive at the expense of other societies, which may die out.[303] The differential success rate of societies provides an additional selective pressure driving cultural evolution.

---

[300] Pinker, 2002, p60. As this quote demonstrates, evolutionary and biological explanations of human behaviour are not alternatives to 'cultural explanations'; they constitute part of the explanation of what culture is. Furthermore, this 'instrumental' view of culture stands in contradistinction to accounts that would argue that 'culture' provides an autonomous source of moral values or norms.
[301] Diamond, 1999; Wright, 2000.
[302] Boyd and Richerson, 1985; Cavalli-Sforza and Feldman, 1981; Dawkins, 1976; Dennett, 1995; Lumsden and Wilson, 1981.
[303] This is sometimes called 'cultural trait group selection' or 'equilibrium selection'. See: Binmore, 1994a; Binmore, 1994b; Dawkins, 1980; Sober and Wilson, 1998.

Turning to the emergence of morality, some theorists begin with a pessimistic, Hobbesian account of human nature, and argue that cultural evolution has equipped humans with social, cooperative and moral *ends*. According to Dennett, memes have 'interests' in the same way that genes have interests; that is, they exhibit features designed to promote their replication. In this way, memes constitute an autonomous source of interests, goals and values and hence provide an alternative source of motivation for, and hence explanation of, human behaviour. The interests of memes can conflict with the interests of the genes of their hosts. In such cases memes can "redirect or exploit or subvert . . . [the] genetically endorsed designs" of their hosts in exactly the same way that a particular parasite takes over the nervous system of certain ants.[304] In other cases, however, memes share common interests with the genes of their host, and can therefore act as benign symbionts that provide benefits. For example, certain memes might lead their hosts to be more cooperative or altruistic, thereby solving various collective-action problems. Hence memes can counter-act some of our selfish tendencies and promote more beneficial cooperation; and individuals or societies 'infected' with such memes may 'do better' than uninfected rivals.[305] Other researchers, adopting a milder version of the thesis, have assumed that humans have a generalised tendency to be social, and to conform to local conventions, but that their biology leaves the *means* by which they cooperate open; cultural evolution is then invoked to explain the design of the means – the moral rules – that they adopt.

Clearly, arguments to the effect that morality is a product of cultural evolution must establish: that there is such a thing as cultural evolution; and that cultural evolution provides a better explanation of morality than does biological evolution. Below I argue that there is no such thing as cultural evolution in any strict sense, and so cultural evolution is not available as a candidate explanation of morality (or of anything else).

---

[304] Dennett, 1995, p472.
[305] Dennett, 1995, p470.

Does culture *evolve by natural selection* as opposed to merely changing gradually? In order to be subject to evolution by natural selection, an entity must undergo successive rounds of 'random variation and non-random replication'. Genes, for example, undergo random variation (mutation). And gene replication – by means of an exact template – is non-random by virtue of the fit between their phenotypic effects and the world. However, it is far from clear that there are such things as 'memes' that meet these criteria.

First, new cultural artefacts – tools and rules – do not appear, nor are they modified, at random. Cups, for example, did not start out as randomly-varying amorphous blobs and only later, through selective retention of some blobs, come to hold first a little bit of water, then a little bit more, and finally an entire pint. On the contrary, *people* invent, create and experiment with new designs; and once invented, artefacts tend to improve gradually over time because successive designers make improvements. Hence cultural artefacts exhibit 'non-random' or 'directed' variation. As a result, it is not clear what problem 'culture evolution' is intended to solve. Recall that in biology, the problem that natural selection solved was how to explain design in the absence of a designer. When it comes to culture, we have an explanation of design: the human designer. In order to *create* a problem in need of a solution, proponents of cultural evolution would need to point to a feature of an artefact that goes beyond what its creator intended, but might be explained as a product of the independent design process of cultural evolution. But no such examples are forthcoming. Instead, proponents of cultural evolution tend to point to artefacts or habits that persist despite harming their users (such as smoking); or they point to serendipitous features that had unintended consequences (such as, boiling water also purifies it). But we have plenty of explanations of why people are addicted to smoking – explanations that have to do with human physiology, and the existence of large cigarette corporations – that do not posit that smoking is a self-replicating entity. And while it is undoubtedly the case that discoveries and inventions have unanticipated consequences, and that societies that

make better discoveries 'do better' than societies that do not, it is not clear what cultural evolution adds to this observation.

Second, cultural transmission may well be non-random – in the sense that useful artefacts spread further and faster than useless ones – but it does not involve *replication* of the kind necessary to lead to the accumulation of beneficial design features. Cultural artefacts are not copied 'warts and all' from person to person as would be the case if transmission was as precise as the template-replication that genes undergo. Instead, social learning and cultural transmission involve the sophisticated inferential reconstruction of the goals of the actor, which includes the ability to correct for the actor's mistakes. Even infants correct for the errors that their teacher is making, and copy what the actor is trying to do and not what the actor actually does.[306] Hence memes exhibit non-random transmission but not non-random *replication*.

If cultural artefacts do not vary randomly, and are not replicated exactly, then they are not subject to natural selection, and cultural transmission provides no independent mechanism for explaining the design of tools and rules. As such, 'cultural evolution' seems to be little more than a convoluted redescription of the phenomena it seeks to explain. If Person A sees Person B enjoying an exotic fruit, and elects to try it herself, is it any more than a redescription of the phenomena to say that the 'meme for' eating that fruit has spread from Person A to Person B? If a newcomer adopts the local convention of driving on the left, what extra theoretical leverage is gained by claiming that the 'meme for' driving on the left has gained another host? If anything, the metaphor of cultural evolution serves only to obscure and distract attention from the actual motors of culture and cultural change: the many different psychological mechanisms, abilities and preferences involved in the invention and transmission of different types of cultural artefacts.[307] After all, the psychological mechanisms that enable a child to acquire local vocabulary are quite distinct from the mechanisms that create aversions to particular

---

[306] Pinker, 2002, p62.
[307] Daly, 1982; Miller, 2000b.

types of food, or the combination of folk physics and folk psychology that enable a person to inferentially reconstruct and recreate the purposes of a tool.

Clearly, if there are no such things as memes or cultural evolution, then this alternative account of morality does not even get off the starting blocks. We can leave for another day such questions as: how and why natural selection would have removed adaptations for cooperation from the human line; how 'cultural evolution' is supposed to account for the genetic, hormonal, anatomical, physiological and neurological evidence for specific adaptations for cooperation; or, in Dennett's case, how memes are supposed to commandeer human nervous systems, fusing their values with our own.[308]

**Hume's account of moral psychology**

As we have seen, Hume provided the classic statement of belief-desire psychology. He assumed that people would act in ways that they believed would satisfy their desires. But Hume argued, contrary to Hobbes, that human nature comprised some desires that promoted the welfare of others and the common good.[309] These attributes met with the approval of our moral sense – our sympathy with the common interest – and were thereby deemed 'moral'.[310] In this section I show that the human adaptations for cooperation discussed above correspond almost exactly to

---

[308] Incidentally, much of the appeal of Dennett's account of cultural evolution comes from the claim that our reliance on memes provides a way to break free from our biology: "This is our transcendence, our capacity to 'rebel against the tyranny of the selfish replicators'." But even if the account that Dennett gives of culture were true, it is not so obvious that he succeeds in liberating us; Dennett has, after all, compared memes to infestation by parasites. Indeed, in making us slaves to two masters, the triumph of memes seems something of a hollow victory. Dennett, 1995, p471. Undaunted, Dennett develops the idea that 'memes set us free' in: Dennett, 2003.

[309] As we have seen, Hume used a variety of synonyms for 'the common good': the "publick interest"; the "public good"; a "common end"; "the general interests of society"; "the good of mankind". Hume, 1739/1985, p532, p580, p590, p620, p628.

[310] The philosopher John Mackie notes: "Protagoras, Hobbes, Hume and Warnock are all at least broadly in agreement about the problem that morality is needed to solve: limited resources and limited sympathies together generate both competition leading to conflict and an absence of what would be *mutually beneficial cooperation*." And as Simon Blackburn notes: "the great naturalists, theorists who have sought to understand ethical thought as part of the natural world – notably Hobbes and Hume – are also inclined towards ethical rules whose authority eventually derives from promoting the common good, or, perhaps, avoiding the common bad." Blackburn, 1998, p46; Mackie, 1977, p111.

the passions, the sentiments and the character traits that Hume identified as promoting the common good. These include familial affection, sympathy, reciprocity, pride, humility, and respect for property. As we shall see, the specific social problems that these passions and traits are described as solving, and the way that they solve them, correspond almost exactly to the problems of cooperation and the adaptive solutions discussed above.

The main difference between the Darwinian and Humean accounts of moral psychology is that they draw the line between 'natural' and 'artificial' in different places. According to the Darwinian account, we have a range of adaptations for cooperation, and each of them may be 'artificially extended' through the invention of new tools and rules. According to the Humean account, human nature has a range of 'natural' virtues to which has been added a range of invented 'artificial' virtues. For example, whereas the Darwinian sees reciprocity and property as ancient and hard-wired aspects of human psychology, Hume regards them as relatively novel artificial inventions. However, the decision as to where to draw the line is an empirical one; Hume decided where to draw the line after surveying the evidence available to him. As we have seen, in pursuing his "anatomy of the mind", Hume adopted a principle of comparative parsimony, and considered evidence from nonhuman species to be relevant to determining whether certain passions were natural to humans: "where the structure of parts in brutes is the same as in men, and the operation of those parts also the same, the causes of that operation cannot be different, and that whatever we discover to be true of the one species, may be concluded without hesitation to be certain of the other".[311] But Hume mistakenly thought that animals "quickly lose sight of the relations of blood; and are incapable of that of right and property" and cannot take pride in "external objects".[312] (Hume had clearly never met a bee, a butterfly or a bower bird.) He therefore assumed that these traits must be artificially created by humans. I suggest that, with access to modern evidence, Hume would be willing to revise his initial conclusion

---

[311] Hume, 1739/1985, p375.
[312] Hume, 1739/1985, p377.

and view more of the virtues as natural (not least because such evidence would bolster his overall argument for the naturalness of morality).

In any case, let us see what Hume had to say about familial affection, sympathy, reciprocity, pride, humility, and respect for property.

*Kinship*

On the subject of kinship, Hume notes that families provide for the education of children, and contribute to the stability of society as a whole. Hume observes that families are held together by ties of natural affection: "the relation of blood produces the strongest tie the mind is capable of in the love of parents to their children, and a lesser degree of the same affection, as the relation lessens."[313] This natural affection for one's children gives rise to a duty of care for one's offspring: "We blame a father for neglecting his child. Why? because it shews a want of natural affection, which is the duty of every parent."[314]

More specifically, Hume noted that uncertainty of paternity would lead men and women to differ in their degree of commitment their children:

> [I]n order to induce the men to ... undergo cheerfully all the fatigues and expences [of child-rearing], they must believe, that the children are their own... Now if we examine the structure of the human body we shall find, that ... an error may easily take place on the side of the [man], tho' it be utterly impossible with regard to the [woman]. Men are induc'd to labour for the maintenance and education of their children, by the persuasion that they really are their own; and therefore 'tis reasonable, and even necessary, to give them some security in this particular.[315]

---

[313] Hume, 1739/1985, p401.
[314] Hume, 1739/1985, p530.
[315] Hume, 1739/1985, p621.

As a result of this asymmetry, argued Hume, chastity and sexual fidelity are valued more in women than in men: "In order, therefore, to impose a due restraint on the female sex, we must attach a peculiar degree of shame to their infidelity, above what arises merely from its injustice, and must bestow proportionable praises on their chastity." [316] In other words, feminine virtues such as fidelity and chastity are moral because they help to keep families together, which in turn benefits society.

*Coordination to mutual advantage*

Hume had relatively little to say about 'theory of mind' (unless one interprets the ability to take up the position of the "judicious spectator" in this context). Hume did, however, discuss sympathy and conventions *qua* solutions to coordination problems.

Hume regarded 'sympathy' as central to human moral psychology. For Hume, the human mind was such that feelings and emotions are inevitably transmitted from one to another: "The minds of all men are similar in their feelings and operations; nor can any one be actuated by any affection, of which all others are not, in some degree, susceptible. As in strings equally wound up, the motion of one communicates itself to the rest; so all the affections readily pass from one person to another, and beget correspondent movements in every human creature."[317] Sympathy enables us to feel another's pleasures and pains as our own. In the first instance, sympathy disposes people to be generous to "friends and family, or, at most, . . . [members of] their native country". But sympathy also establishes a general rule or habit by which we interact with others, and thus extends, in a weaker form, to people in general. The result is that "A propensity to the tender passions makes a man agreeable and useful in all the parts of life."[318]

---

[316] Hume, 1739/1985, p621.
[317] Hume, 1739/1985, pp626-7.
[318] Hume recognised that benevolence is dependent on there being *some* common interest. "[T]he fortifications of a city belonging to an enemy are esteemed beautiful upon account of their strength, tho' we cou'd wish that they were entirely destroy'd." Hume, 1739/1985, p637.

Sympathy was also central to Hume's account of 'artificial virtue'. As suggested above, in addition to the natural virtues, Hume argued that there were also a range of 'artificial virtues' that did not spring directly from human nature, but were the product of human artifice, invention and contrivance. "Mankind," noted Hume, "is an inventive species".[319] Foremost amongst these 'artificial virtues' were 'conventions'. Hume's account of convention begins with "a general sense of common interest" which "all the members of the society express to one another, and which induces them to regulate their conduct by certain rules". "When this common sense of interest is mutually expressed, and is known to both, it produces a suitable resolution and behaviour. And this may properly enough be call'd a convention or agreement betwixt us . . . ".[320] Hume illustrates conventions with two examples of cooperative labour: two men pulling on the oars of a boat, and two men working together to drain a meadow.[321] In the latter case: "Two neighbours may agree to drain a meadow, which they possess in common; because it is easy for them to know each others mind; and each must perceive, that the immediate consequence of his failing in his part, is, the abandoning the whole project."[322]

Artificial virtues, such as conventions, meet with our approval to the extent that they promote the common good, and thereby invoke our sympathy:

> Now as the means to an end can only be agreeable, where
> the end is agreeable; and as the good of society, where our
> own interest is not concerned, or that of our friends, pleases
> only by sympathy: It follows, that sympathy is the source of
> the esteem, which we pay to all the artificial virtues. . . .
> [S]ympathy is a very powerful principle in human nature, . .

---

[319] Hume, 1739/1985, p536.
[320] Hume, 1739/1985, p541.
[321] Hume, 1739/1985, p541, p542.
[322] Hume, 1739/1985, p590. Skyrms uses this example to illustrate the problem of Rousseau's 'stag hunt' Skyrms, 2004, p2.

. it produces our sentiment of morals in all the artificial virtues.[323]

As John Rawls notes, because the convention promotes the common interest, Hume treats it "as normative in that violations of it are objected to and criticized and the offending party is expected to accept these reproofs and suitable penalties".[324]

*Reciprocity*

Hume argues that the problems of 'delayed mutualism' are solved by another artificial virtue: reciprocity. Hume argued that selfishness and short-sightedness[325] often prevented people from taking advantage of certain forms of cooperation:

> Men being naturally selfish, or endow'd only with a confin'd generosity, they are not easily induc'd to perform any action for the interest of strangers, except with a view to some reciprocal advantage, which they had no hope of attaining but by such a performance. Now as it frequently happens, that these mutual performances cannot be finish'd at the same instant, 'tis necessary, that one party contented to remain in uncertainty, and depend upon the gratitude of the other for a return of kindness. . . . [But] we cannot depend upon their gratitude. Here then is the mutual commerce of good offices in a manner lost among mankind . . . [326]

---

[323] Hume, 1739/1985, p628. Hume continues: "had not men a natural sentiment of approbation and blame, it could never be excited by politicians; nor would the words laudable and praise-worthy, blameable and odious be any more intelligible, than if they were a language perfectly known to us, as we have already observed." (pp629-30). Or, as Hume puts it elsewhere, "a sympathy with public interest is the source of the moral approbation" of the convention. Hume, 1739/1985, p551.

[324] Rawls, 2000, pp60-1.

[325] "Men are always more concern'd about the present life than the future; and are apt to think the smallest evil, which regards the former, more important than the greatest, which regards the latter." "There is no quality in human nature, which causes more fatal errors in our conduct, than that which leads us to prefer whatever is present to the distant and remote…." Hume, 1739/1985, p577, p590.

[326] Hume, 1739/1985, pp571-2.

As a result, "both of us lose our harvest for want of mutual confidence and security."[327] People overcome this problem, argues Hume, by means of the institution of promising – a conventional sign of obligation – and we keep our promises in order to maintain our reputations for honesty. "Hence I learn to do a service to another, without bearing him any real kindness; because I forsee, that he will return my service, in expectation of another of the same kind, and in order to maintain the same correspondence of good offices with me or with others."[328] In this way, Hume identifies both the problem and the solution of the prisoner's dilemma.

*Conflict resolution: costly-signalling and hierarchy*

In another remarkable passage, Hume introduces pride, humility and admirable character traits (called 'virtues') with a discussion of the costly signalling of fitness indicators in other species:

> It is plain, that almost in every species of creatures, but especially of the nobler kind, there are many evident marks of pride and humility. The very port and gait of a swan, or turkey, or peacock show the high idea he has entertained of himself, and his contempt of all others. This is the more remarkable, that in the two last species of animals, the pride always attends the beauty, and is discovered in the male only. The vanity and emulation of nightingales in singing have been commonly remarked; as likewise that of horses in swiftness, of hounds in sagacity and smell, of the bull and cock in strength, and of every other animal in his particular excellency.[329]

---

[327] Hume, 1739/1985, p573.
[328] Hume, 1739/1985, p573.
[329] However, argued that animals could take pride only in their physical characteristics, not in their intelligence, or "external objects" Hume, 1739/1985, p376.

"All these are evident proofs," argues Hume, "that pride and humility are not merely human passions, but extend themselves over the whole animal creation."[330]

Hume proceeds to argue that the *causes* of pride and humility are "much the same in beasts as in us": "the same qualities cause pride in the animal as in the human kind; and it is on beauty, strength, swiftness or some other useful or agreeable quality that this passion is always founded".[331] In humans, Hume argues that attributes such as "beauty, strength, agility, good mien, address in dancing, riding, and . . . dexterity in any manual business or manufacture" as well as mental qualities such as "imagination, judgment, memory or disposition; wit, good-sense, learning, courage, justice, integrity" are all causes of "a genuine and hearty pride, or self esteem".[332] Such 'well founded' pride is "essential to the character of a man of honour" and is responsible for "*heroic virtue*" that benefits others. These virtues include: "[c]ourage, intrepidity, ambition, love of glory, magnanimity, and all the other shining virtues".[333]

Hume also notes that differences in ability give rise to hierarchies in which "certain deferences and mutual submissions" are required "of the different ranks of men towards each other . . . ". "'Tis necessary, therefore, to know our rank and station in the world, . . . to feel the sentiment and passion of pride in conformity to it, and to regulate our actions accordingly."[334] Humility, the converse of pride, is also valuable to the smooth running of society: "a just sense of our weakness, is esteem'd virtuous, and procures the good-will of everyone"[335]

---

[330] Hume, 1739/1985, p376.
[331] Hume, 1739/1985, pp376-7.
[332] Hume, 1739/1985, pp330-1, p648-9. The absence of these qualities was occasion for humility.
[333] Hume, 1739/1985, pp648-50. Adam Smith concurs: "We frequently, not only pardon, but thoroughly enter into and sympathize with the excessive self-estimation of those splendid characters in which we observe a great and distinguished superiority above the common level of mankind. We call them spirited, magnanimous, and high-minded; words which all involve in their meaning a considerable degree of praise and admiration." Smith, 1759/1976, p255.
[334] Hume, 1739/1985, p650. "A sense of superiority in another breeds in all men an inclination to keep themselves at a distance from him, and determines them to redouble the marks of respect and reverence, when they are oblig'd to approach him" (p441).
[335] Hume, 1739/1985, p642.

*Conflict resolution: property*

Finally, Hume discusses at length the problem of defending private property, and its solution.

There is, notes Hume, a scarcity of goods "in comparison of the wants and desires of men".[336] Hence, when men form society, "[t]he convention concerning the stability of possession is enter'd into, in order to cut off all occasions of discord and contention".[337] It is "a convention for the stability of possession, and for mutual restraint and forbearance".[338] Hume proceeds to discuss the means by which property is assigned, and argues that foremost among these is "first possession" or "occupation".[339] For Hume, this right to property is central to the formation of society, and accounts, in large part, for our ideas of justice and injustice. "After this convention . . . there immediately arise the ideas of justice and injustice; as also those of *property*, *right*, and *obligation*. . . . [A]fter the agreement for the fixing and observance of this rule, there remains little or nothing to be done towards settling a perfect harmony and concord."[340]

What is more, Hume anticipates 'prospect theory' by two hundred years when he observes that "Men generally fix their affections more on what they are possessed of, than on what they never enjoyed".[341]

In summary, it seems that evolutionists and Hume come to similar conclusions about the problems facing social animals, and about the kinds of solutions they tend to adopt. Adaptations for kin altruism, coordination, reciprocity and conflict resolution clearly encompass what Hume refers to as natural affection, feminine virtue, sympathy, convention, reciprocity, pride, heroic virtue, first possession, and so on.

---

[336] Hume, 1739/1985, p546.
[337] Hume, 1739/1985, p553.
[338] Hume, 1739/1985, p554.
[339] Hume, 1739/1985, p556. The other means of acquiring property are: *prescription*, which refers to the a historical record of first possession; *accession*, which is the right to own the produce of one's possessions, such as the apples from one's tree; and *succession*, which is the right to inherit property from one's kin. Property can also be *transferred by consent* – which serves to allocate property to those that desire it most. Hume, 1739/1985, pp556-8.
[340] Hume, 1739/1985, p542, p543.
[341] Hume, 1739/1985, p534.

**Conclusions**

Two conclusions follow from this correspondence between evolutionary psychology and Hume's thinking.

The first is that – thanks to evolutionary theory, animal behaviour, human evolutionary psychology, and the ethnographic record – Hume wins his argument with Hobbes. Human nature is not entirely selfish, but comprises desires aimed at the welfare of others and at the common good. For this reason, as Simon Blackburn observes, Hume would "have applauded contemporary alliances between philosophy and psychology, neuroscience and evolutionary theory".[342]

The second conclusion is that – thanks to Hume – the ethical import of the discovery of adaptations for cooperation becomes clear. Remember that, according to Hume, passions give rise to values; and passions aimed at the common good give rise to moral values. If 'moral passions' are revealed to be 'adaptations for cooperation', then it follows that 'moral values' are revealed as 'the proximate goals of these adaptations'. This would suggest that, as Hume envisaged, moral philosophy – the study of moral value – begins with the investigation of human nature and, as such, should be seen as a branch of biology, psychology and anthropology.[343]

In the next chapter I review the moral philosophy that emerges from this Humean-Darwinian account of morality by seeing how it tackles a range of problems in moral philosophy.

---

[342] Blackburn, 2000.

[343] As John Mackie puts it: "It is not for nothing that his work is entitled *A Treatise of Human Nature*, and subtitled, *An attempt to introduce the experimental method of reasoning into moral subjects*; it is an attempt to study and explain moral phenomena (as well as human knowledge and emotions) in the same sort of way in which Newton and his followers studied and explained the physical world." Mackie, 1980, p6. Or, as Ernest C. Mossner's introduction to the *Treatise* puts it, Hume's moral philosophy was "an attempt to establish a purely naturalistic ethics based on the facts of human nature". Hume, 1739/1985, p19.

# Chapter 5 – The problems of moral philosophy

*[E]thical philosophers intuit the deontological canons of morality by consulting the emotive centers of their own hypothalmic-limbic systems.*

                    – Edward O. Wilson[344]


## Introduction

The purpose of this chapter is to show how the Humean-Darwinian account of morality sketched in the previous chapters provides a rich deductive framework in which to locate, make sense of, incorporate and perhaps reconcile a wide variety of otherwise unruly, and apparently contradictory, positions in traditional moral philosophy. I show that the Humean-Darwinian account can incorporate meta-ethical theories about the nature of moral values (such as monism and pluralism, subjectivism and objectivism, universalism and particularism), as well as substantive or normative theories about the content of moral values (such as the feminist 'ethic of care', norm-following, social contracts, and virtue ethics). I also show that the Humean-Darwinian account of human morality provides a principled means of dividing up the problem space of applied ethics. I conclude that because the Humean-Darwinian account is consistent with what is known about human origins, human psychology and human behaviour, and because it can incorporate a wide variety of traditional ethical positions, it should be adopted as the default account of morality.


## Meta-ethics: What is the nature of moral value?

Meta-ethics can be divided into primary and secondary problems. Primary problems include: What is the nature of moral value? What kinds of things are they? Where do they come from? What are they made of? How do they work? Where do they fit into our picture of the universe? In the history of moral philosophy, there have been a variety of theological,

---

[344] Wilson, 1975, p563.

cosmological, biological and sociological answers to these primary questions (as well as nihilistic rejections of the premise that there are any such things as moral values). Some have argued that moral values are "divine commands" or "thoughts in the mind of God". Some have argued that moral values are the dictates of reason or a 'pure rational will'. Some have argued that moral values are the product of natural human desires. Some have argued that moral values are social conventions or cultural inventions. And others have argued that so-called moral values are merely the disguised 'interests of the stronger'.

David Hume treated the primary problem of 'the nature of morality' – and specifically, the question of whether morality is a product of certain desires – as an empirical problem to be solved by scientific method and the best available evidence.[345] The logic of Hume's argument runs as follows: Hume identifies acting morally with acting to promote the common good. He presents a 'subjective' theory of value, according to which values are the products of natural passions or desires. Hume then argues that human nature comprises a variety of passions that naturally dispose them to value and promote the common good. Thus Hume concludes that moral values are the product of certain natural human passions. (In addition, he argues that the moral passion of sympathy enables humans to approve of novel and 'artificial' means of promoting the common good.)

The Darwinian account of morality advanced in this thesis retains the structure of Hume's argument, but updates his ontology in two main ways. First, it uses the (evolutionary) game theory of cooperation to provide a rigorous account of 'the common good', and in doing so distinguishes several different types of cooperation. Second, it views 'passions', 'desires', 'sentiments' and so on as examples of 'adaptations'; and it identifies 'values' with the proximate goals of adaptations. So, the

---

[345] This approach to the problems of meta-ethics differs from much of the rest of contemporary meta-ethics, in which the *problem* is that we have conflicting intuitions about the operation of moral psychology, and the methods used to solve these problems include further introspection and intuition-mongering conducted within the framework provided by folk psychology. See: Jackson, et al., 2004; Smith, 1994. Compare: Doris and Stich, (forthcoming); Stich, 1993.

Humean-Darwinian account identifies morality with cooperation. It presents a theory of value according to which values are the proximate goals of adaptations. It shows that humans have adaptations for cooperation that lead them to be social, cooperative and altruistic under certain conditions. Thus it concludes that moral values are the products of certain natural human adaptations. (In addition, these adaptations for cooperation enable humans to approve of novel or 'artificial' means of cooperating.)

In this way, Hume and Darwin offer an entirely naturalistic account of the origins and ontological status of morality that is consistent with what we know about human origins, psychology and behaviour. This account of morality incorporates the 'natural desire' and 'cultural invention' meta-ethical theories; it has no need for such question-begging entities as a god or a 'pure rational will'; and it rejects nihilistic skepticism about the existence of moral value.

This Humean-Darwinian account also enables us to make sense of and incorporate a variety of secondary meta-ethical theories. Secondary problems include: Are moral values the subjective property of valuers, or are they objective properties of the things valued? How many moral values are there? Is there one 'crowning virtue' by which all the others can be measured, or are there many potentially incompatible moral values? Is there one moral system applicable to all humans, or do different societies have different but equally valid moral systems? The various possible answers are referred to as subjectivism and objectivism, monism and pluralism, and universalism and particularism, respectively.[346]

*Subject and object*

The Humean-Darwinian account of morality and moral value has subjective and objective elements.

---

[346] I refer to these as secondary problems because their solutions depend in part on the position one adopts on primary questions.

On the one hand, values – *qua* the proximate goals of adaptations – are clearly 'subjective' in the classical sense of being located in the minds of human valuers, and not 'out there' in the objective world. On the other hand, the adaptations that give rise to values are objective features of human nature – they are as 'objective' as giraffes' necks, peacocks' tails, or any other adaptation studied by evolutionary biology.[347] Hence, human values are not 'subjective' in the sense of being arbitrary, whimsical or 'merely a matter of personal opinion'. On the contrary, the adaptations involved are expected to be somewhat 'hardwired'; people can no more choose what moral values to have than they can choose which colours to see. In addition, these adaptations have been selected to 'latch on to' and to value certain reliably-occurring 'objective' features of the (social) world – namely, configurations of interests that constitute opportunities for cooperation, and traits and acts and inventions that realise such opportunities. Hence, questions such as "What kind of game are we playing (is this a coordination game or a prisoner's dilemma)? What would constitute a contribution to the common good in this context? Does this particular act contribute more to the common good than some other act?" and so on are objective questions that can be given objective answers. They can be pursued separately from evaluative questions to do with whether or not acting in a particular way (to promote the common good) is (morally) desirable.

From the Humean-Darwinian perspective, it is not clear what there is to gain by continuing to argue about whether adaptations are 'subjective' or 'objective'; adaptations are subjective in some ways, and objective in others. Now that we know what morality is, we can concentrate our energies on investigating how it works, while providing historians of moral philosophy with a theory of why it was that ethicists tended to divide along subject-object lines.

---

[347] As Isaiah Berlin puts it: "values are objective – that is to say, their nature, the pursuit of them, is part of what it is to be a human being, and this is an objective given. The fact that men are men and women are women and not dogs or cats or tables or chairs is an objective fact; and part of this objective fact is that there are certain values, and only those values, which men, while remaining men, can pursue" Berlin, 1969; Berlin, 1998.

A similar analysis can be applied to the tertiary question: What does a person mean when they say 'X is wrong'? Are they expressing something like a desire or an attitude? Or is the statement an expression of a belief about some matter of fact?[348] Subjectivists (emotivists, expressivists, internalists, non-cognitivists, and anti-realists) hold that moral statements are expressions of desire; objectivists (externalists, cognitivists and realists) hold that moral statements are expressions of belief. The Humean-Darwinian answer is: 'it depends'. Sometimes people say "X is wrong" in order to express their feelings, to advertise their convictions, or to reassure others of their humanity; sometimes, "X is wrong" expresses a belief or an expectation about what other people think, or makes a factual claim about the content of human moral sentiments;[349] sometimes "X is wrong" expresses a belief about the structure of the game, or the likely consequences of an act; sometimes "X is wrong" expresses a belief that X constitutes a relatively inefficient means to some more fundamental end; sometimes "X is wrong" expresses the mistaken belief that moral values are a property of external objects; and so on. The view that moral values are proximate goals of adaptations for cooperation is consistent with all these expressions. As always, in any given instance, the meaning of any signal depends on what the audience is uncertain about, and which aspect of that uncertainty the signaller is attempting to reduce.[350]

*The one and the many*

The Humean-Darwinian account of morality and moral value also has monist and pluralist elements.

On the one hand, the Humean-Darwinian account suggests that there are as many *values* as there are *adaptations*; and that there are as many *moral values* as there are *adaptations for cooperation*. There are many adaptations

---

[348] I call these tertiary questions because one's answers are largely the product of one's position on the secondary question of whether moral values are subjective or objective.
[349] Smith, for example, argues that moral statements are 'beliefs about desires'; that is, beliefs about what desires we would in fact have under certain ideal conditions. Smith, 1991. See also: Brink, et al., 1997.
[350] Presumably, the intended meaning can be revealed by asking the speaker to give his reasons for making the statement.

for cooperation. Therefore, there are many moral values. In this sense, the account is 'value pluralist' as defined by Isaiah Berlin: "there is a plurality of values which men can and do seek, and . . . these values differ. There is not an infinity of them: the number of human values, of values that I can pursue while maintaining my human semblance, my human character, is finite – let us say 74, or perhaps 122, or 26, but finite, whatever it may be."[351] It follows that, as value pluralists maintain, there are likely to be situations in which two or more incompatible values come into conflict, giving rise to apparently insoluble moral dilemmas.

On the other hand, what disparate moral values have in common is that they all constitute means of achieving essentially the same end, namely cooperation. In theory at least, the contribution that different values make to the common good provides a common metric that can be used to trade one value off against another. For example, if one is torn between keeping a promise and being heroic, it is in theory possible to calculate what contribution each value makes to the common good, and choose accordingly. (Although they still might come out equal, in which case the dilemma remains.) In this sense, the Humean-Darwinian account is 'monist' in Berlin's terms, because it suggests that values are commensurable.[352]

Again, on the Humean-Darwinian account, it is fruitless to argue about whether morality is monist *or* pluralist; it is monist in some ways, but pluralist in others. Instead, we can attempt to complete our catalogue of the moral values, and investigate how they fit together and what weightings they have been given by natural selection. Meanwhile, the Humean-Darwinian account can bequeath to historians of moral philosophy a theory of why it was that ethicists tended to divide themselves up into monists and pluralists.

---

[351] Berlin, 1969; Berlin, 1998.
[352] For further discussion of this issue, see Gray, 1995.

*Universal and particular*

Finally, the Humean-Darwinian account of morality has universalist and particularist elements.

Just as Hume thought that the moral passions were a universal part of human nature, the Darwinian update of Hume expects the adaptations for cooperation to be part of the 'species-typical' biology and psychology of *Homo sapiens*. However, the theory expects this psychology to consist of banks of sophisticated, condition-dependent, facultative adaptations that lead moral sentiments to develop differently – and generate different attitudes and behaviour – under different conditions. For example, the psychological mechanisms governing paternal investment are sensitive to cues of paternity; and cross-cultural differences in 'certainty of paternity' might be expected to lead to predictable cross-cultural differences in the obligations of husbands to 'their' children. To take another example, adaptations for reciprocity are expected to be sensitive to the degree of luck versus effort that goes into the acquisition of a good, and to be engaged only when the exchange of goods yields a net advantage to the participants. Different ways of making a living – for example, hunting versus gathering – might lead members of different societies to express predictably different attitudes towards sharing and redistribution.[353]

For this reason, several theorists have suggested that the psychological mechanisms underpinning morality can be compared to the psychological mechanisms underpinning language – in other words, there might be a universal 'moral grammar' that generates different surface moralities in different environments.[354]

---

[353] Cosmides and Tooby, 1992; Henrich, et al., (forthcoming). Compare Michael Walzer who, in *Spheres of Justice*, observes that "the prescriptions . . . that attach children to their maternal uncles" change over time, and that different societies assume that "different goods ought to be distributed for different reasons, in accordance with different procedures". He attributes these differences to differences in "cultural understandings" and "social meanings" that are "the inevitable product of historical and cultural particularism". Thus Walzer takes variation in moral practice as evidence against universalist theories and in favour of relativist theories of the nature of morality, and argues against the possibility of effective cross-cultural moral judgments Walzer, 1983, p228, p6.

[354] See: Charvet, 1981, p190; Chomsky, 1988, pp153-4; Harman, 2000, Ch. 13; Rawls, 1971, p47.

So, depending on one's frame of reference, the Humean-Darwinian can appear universalist and/or particularist.

On one hand, the Humean-Darwinian theory is *universalist* because it posits that there are rules – embodied in species-typical adaptations for cooperation – that underpin all possible human moral systems. And the theory suggests that these rules could, in principle, be used to evaluate the moral worth of an action under any given set of conditions, in any given society, just as in linguistics, the rules of language embodied in 'universal grammar' can be used to determine the grammaticality of any sentence in any linguistic community.

On the other hand, the theory is *particularist* because it expects different, equally valid sets of moral values in different places. And the theory suggests that it makes no sense to evaluate acts performed under one set of conditions according to a set of values expressed under a different set of conditions, just as it makes no sense to use the rules of English grammar to determine whether a Chinese sentence is grammatical.

Once again, on the Humean-Darwinian account, it is redundant to argue about whether morality is universalist *or* particularist; it is universalist in some ways and particularist in others. Instead, we can continue to map the countless ways in which adaptations for cooperation are context-dependent, and begin to explain how moral systems develop differently under different conditions. And we can furnish historians of moral philosophy with a theory of why it was that ethicists tended to divide themselves up along universalist-particularist lines.

Incidentally, the significance of the idea that there may be universal human moral values that develop in different ways under different conditions is not lost on moral relativists. David Wong, for example, recognises that "some of the striking differences in moral belief across societies may not be rooted in differences in fundamental values but in the fact that these values may have to be implemented in different ways

given the varying conditions that obtain across societies".[355] Nevertheless, Wong stakes his moral relativism on the claim that there are "deep and wide" value differences that go beyond what is accounted for by a 'universal grammar' approach. His candidate "deep and wide" difference is the one that exists between the family-based 'virtue-centered' moralities of the East and the contract-based 'rights-centred' moralities of the West.[356] Of course, any such claim is vulnerable to advances in our understanding of 'moral grammar'. The question of whether any particular value differences are really 'deep and wide', or whether they are merely the product of yet-to-be-discovered variables in the underlying moral grammar, is an empirical question that cannot be settled until we have a more detailed explication of the underlying content and structure of moral psychology.

Now that we have looked at how the Humean-Darwinian account of morality tackles the problems of meta-ethics, we may proceed to look at how it tackles the problems of substantive ethics.


### Substantive ethics: What is the content of moral values?

Substantive ethics is the attempt to answer such questions as: What is the content of our moral values? What does morality demand of us? Which things are right and wrong?

According to Hume, acting morally consists of acting to promote the common good. Humans promote the common good by: taking care of their children, sympathising, adopting conventions, returning favours, being heroic, respecting private property, and so on. Therefore, acting morally requires that you: take care of your children, sympathise, adopt conventions, return favours, be heroic, respect private property, and so on. According to the Darwinian update of Hume, acting morally consists of cooperating. Humans cooperate by: caring for family members, adopting local norms, keeping promises, punishing cheats, resolving

---

[355] Wong, 1991, pp444-5.
[356] Wong, 1984; Wong, 1991.

disputes peacefully, respecting prior ownership, and so on. Therefore, acting morally requires that you: care for family members, adopt local norms, keep promises, punish cheats, resolve disputes peacefully, respect prior ownership, and so on.

This Humean-Darwinian account of the content of morality provides a rich deductive framework in which to locate, make sense of, incorporate and perhaps reconcile a wide variety of disparate, and apparently contradictory, positions in traditional normative ethics. Below I show, first, that cooperation, in one form or another, has been identified as the *raison d'etre* of morality by a wide variety of moral philosophers. Second, I show that the various different ways that humans cooperate, as specified by the Humean-Darwinian account, have been identified by various different traditions in moral philosophy – such as the feminist 'ethic of care', norm-following, social contract theory, and virtue ethics – as constitutive elements of human morality.

*Cooperation and the common good*

The Humean-Darwinian assumption that morality can be identified with cooperation is shared by a wide variety of moral philosophers. Plato saw justice as an *efficient division of labour*, in which each individual, and each part of the soul, stuck to the role that it was good at, to the mutual benefit of all.[357] Aristotle argued that justice is "what is for the benefit of the whole community" or "to the *common advantage*."[358] St Augustine argued that the human law consists of "an ordered concord of civic obedience and rule in order to secure a kind of *co-operation* of men's wills for the sake of attaining the things which belong to this mortal life".[359] Aquinas observed: "If then a group of free men is directed by a rule to the *common good* of the group, his government will be right and just . . . ".[360] Bishop Butler suggested that "there is a *public end* and interest of society which

---

[357] Plato, 1974.
[358] Aristotle, 1992, p207, 1160a10-14.
[359] Augustine, 1998, p945.
[360] Aquinas, 1988, pp15-6.

each particular is obliged to promote, is the sum of morals."[361] Bertrand Russell notes that "[M]en's desires conflict, and 'good' is, to my mind, mainly a social concept, designed to find issue from this conflict."[362] John Rawls notes that: "The circumstances of justice may be described as the normal conditions under which human *cooperation* is both possible and necessary."[363] David Wong comments: "Human beings have needs to resolve internal conflicts between requirements and to *resolve interpersonal conflicts of interest*. Morality is a social creation that evolved in response to these needs".[364] John Mackie observes that: "Protagoras, Hobbes, Hume and Warnock are all at least broadly in agreement about the problem that morality is needed to solve: limited resources and limited sympathies together generate both competition leading to conflict and an absence of what would be mutually beneficial *cooperation*."[365] The economist and social theorist Henry Hazlitt observes that: "Social *cooperation* is the foremost means by which the majority of us attain most of our ends. It is on the implicit if not the explicit recognition of this that our codes of morals, our rules of conduct, are ultimately based. 'Justice' itself . . . consists in observance of the rules or principles that do most, in the long run, to preserve and promote social cooperation."[366] And the philosophers Daniel Hausman and Michael McPherson are 'tempted' "to think of the normative principles governing individual interactions as human contrivances to adjudicate conflicts of interest and to secure the benefits of *cooperation*".[367]

Some moral philosophers have been even more specific, and have mentioned all four of the types of cooperation that evolutionary theory leads us to expect. In his influential *The Methods of Ethics*, Henry Sidgwick writes that 'common sense' suggests that there are four kinds of moral duties: "(1) duties arising out of comparatively permanent relationships not voluntarily chosen, such as Kindred . . . : (2) those of similar

---

[361] Butler, 1856, IX.
[362] Russel, 1927, p230.
[363] Rawls, 1971, p126.
[364] Wong, 1984, p175.
[365] Mackie, 1977, p111.
[366] Hazlitt, 1964.
[367] Hausman and McPherson, 1996, p151, p186.

relationships voluntarily contracted, such as Friendship: (3) those that spring from special services received, or Duties of Gratitude; and (4) those that seem due to special need, or Duties of Pity."[368] And it is to these four types of cooperation that we now turn.


*Kin altruism and family values*

The obligations that one has to aid one's children and one's wider family have been considered constituent elements of human morality the world over. In his survey of the world's moral systems, Edward Westermarck observed that:


> There is one duty so universal and obvious that it is seldom mentioned: the mother's duty to rear her children, provided they are suffered to live. Another duty – equally primitive, I believe, in the human race – is incumbent on the married man: the protection and support of his family. . . . The parent's duty of taking care of their offspring is, in the first place, based on the sentiment of parental affection. That the maternal sentiment is universal in mankind is a fact too generally admitted to need demonstration; not so the father's love of his children. . . . [T]he paternal sentiment is hardly less universal than the maternal, although it is probably never so strong and in many cases distinctly feeble.[369]


More recent anthropological surveys have tended to agree about the ubiquity of the duties of the good mother. Edel and Edel observe that "the moral obligation for a mother to take care of her children" is "a common moral obligation which is deeply grounded in humans needs, biological, physiological and social . . . . This is a universal imperative, and very

[368] Sidgwick, 1962, p248. Here I am interpreting 'duties due to special need' in terms of costly signals of altruism.
[369] Westermarck, 1906, quoted in Westermarck, 1994, pp61-3.

pervasive in its moral implications, for it serves the groundwork not just for restrictions, but for positive ideals and virtues."[370]

The Humean-Darwinian account of moral values can also explain why women might feel that they have a different 'ethic' to men in some areas. As we have seen, evolutionary theory expects the sexes to differ to the extent that they have faced different adaptive problems during their evolutionary history. For example, women's reproductive success has depended more upon making discriminating decisions when it comes to investing in offspring than has the reproductive success of men, and as a result they have a more exacting psychology for solving such problems. If moral reasoning involves the execution of evolved adaptive decision rules, then evolutionary theory should lead us to expect sex differences in moral reasoning to the extent that men and women have faced different collective-action problems in their evolutionary histories. And if the proximate goals of these adaptations constitute moral goals, then there should be some disagreement between men and women about how one ought to behave. Charles Darwin was the first person to draw out the implications of his theory for sex differences in moral values. Darwin expected that the adaptive problems facing women – primarily, childcare – would select for sympathy and selflessness, whereas the adaptive problems facing men – primarily, competition with other men – would select for the virtues of patriotism and courage. Darwin saw moral progress as the process whereby female sympathy came to predominate over the male virtues, and began to extend from the family to the tribe to the nation to all of humanity and then to all living things.[371]

In recent years, many feminist writers have (inadvertently) re-invented this argument, although without the biological underpinnings. The basic message of "feminist ethics" is that there is a female way of relating to others that is distinct from the way that males relate to others. Furthermore, it is argued, this mode of relating to others provides an alternative set of values from which to critique the values of

---

[370] Edel and Edel, 1959/1968, p34.
[371] Darwin, 1871, p326. For a discussion, see: Arnhart, 1998, pp143-149.

contemporary, 'male-dominated' society.[372] For example, Carol Gilligan and others have argued that women are more likely to adopt an 'ethic of care' – that is, distributing goods according to the needs of the recipients, and the relationships that one has with them – as opposed to an 'ethic of justice' – that is, distributing goods according to abstract rules and principles.[373] This female mode of moral reasoning is suggestively similar to the way in which a mammalian mother would relate to her offspring. This theme is apparent in work with titles such as *Maternal Thinking* and *Caring: A feminine approach to ethics and education.*[374]

Turning to "the duty of assisting brothers and sisters and more distant relatives", Westermarck observes that: "Among the Aleuts . . . a brother 'must always aid his brother in war as well as in the chase, and each protect the other; but if anybody, disregarding this natural law, should go to live apart . . . such dishonourable conduct would lead to general contempt.' . . . I presume that these examples of fraternal relations may, on the whole, be regarded as expressive of universal facts."[375]

Westermarck goes on to comment that duties to kin reach their apotheosis in the moral philosophy of Confucius, according to whom "the love which brother should bear to brother is second only to that which is due from children to parents."[376] The moral relativist David Wong, and the social theorist Francis Fukuyama concur that obligations to one's kin constitute moral duties according to Confucianism. Wong observes that Confucianism makes "the family and kinship groups the models for the common good", and that the Chinese "modelled ideal political relations after the hierarchical structure of the family".[377] And Fukuyama observes that: "The central core [of the 'Confucian personal ethic'] was the apotheosis of the family – in Chinese, the *jia* – as the social relationship to

---

[372] Grimshaw, 1991.
[373] Gilligan, 1982.
[374] Noddings, 1978; Ruddick, 1980.
[375] Westermarck, 1906, quoted in Westermarck, 1994, pp61-3. "Aleut" is the name given to inhabitants of the islands off Alaska. Dumond, 1987.
[376] Westermarck, 1906, quoted in Westermarck, 1994, pp61-3.
[377] Wong, 1984, p163; Wong, 1991, p445.

which all others were subordinate. Duty to the family trumped all other duties…".[378]


*Coordination to mutual advantage*

Successful mutualism involves identifying others with whom you have common interests, and coordinating your behaviour with theirs. Iterated mutualisms can also lead to the formation of strong affiliations between unrelated individuals: friendships. These ways of cooperating – the identification with another's goals, coordination by means of shared expectations or 'norms', and the formation of friendships – have figured prominently in a number of moral philosophies.

The tendency to identify with others' goals and adopt them as one's own – to exhibit sympathy, compassion, empathy, fellow-feeling, pity – has been central to a number of prominent moral philosophies.[379] In the previous chapter we saw that Hume regarded sympathy as the pre-eminent moral passion, responsible for our general 'agreeable' attitude towards others, and also for our approval of the artificial virtues. Darwin argued that many animals felt sympathy with their fellow creatures, and that such sympathy was the "foundation-stone" of a "moral sense or conscience".[380] Most famously, Adam Smith made sympathy central to his *Theory of Moral Sentiments*.[381] Smith argued that, contrary to Hobbes, humans are not indifferent to the joys and sorrows of others, but are disposed to "enter into the motives" of others and be affected accordingly. Smith referred to this tendency as pity, compassion or sympathy.[382] According to Smith, sympathy works by imagining that we are in the other's place, and "conceiving what we ourselves should feel in

---

[378] Fukuyama, 1996, p85.

[379] For reviews, see Goldman, 1993; Harman, 2000, Ch. 11. These reviews seem to suggest that further investigation into the human capacity for 'mindreading' may reveal that theory of mind makes possible other apparently unique features of human moral psychology, such as the ability to see oneself as others see you – to conceive of an impartial spectator, or an impartial observer – and hence to make impartial evaluations of one's own conduct.

[380] Darwin, 1871, pp70-2.

[381] Smith, 1759/1976.

[382] "Pity and compassion are words appropriated to signify our fellow-feeling with the sorrow of others. Sympathy, though its meaning was, perhaps, originally the same, may now, however, without much impropriety, be made use of to denote our fellow-feeling with any passion whatever." Smith, 1759/1976.

the like situation".[383] Sympathy, according to Smith, is more sophisticated than merely copying the emotions present in another ('emotional contagion'). Sometimes an observer experiences emotions that the observed does not (for example, when we feel sorry for a unwitting cuckold, or embarrassed for another's unacknowledged *faux pas*). And sometimes the observer does not experience the emotions that the observed does (for example, when we think the person is 'over-reacting', or where his anger is not justified). Smith argues that the exercise of sympathy gives rise to the virtues of humanity and benevolence: "And hence it is, that to feel much for others and little for ourselves, that to restrain our selfish, and to indulge our benevolent affections, constitutes the perfection of human nature; and can alone produce among humankind that harmony of sentiments and passions in which consists their whole grace and propriety."[384]

Turning to coordination through norms, the moral philosopher Allan Gibbard has done most to place norm-following in an evolutionary framework.[385] Gibbard argues that people possess "biological adaptations for coordination" – "psychic mechanisms that were shaped in the course of human evolution" – that enable them to identify and adopt norms and conventions.[386] These norms consist of implicit or explicit expectations about behaviour that serve to coordinate individuals to mutual advantage.[387] This tendency, argues Gibbard, is central to human morality: "The key to human moral nature lies in coordination broadly construed."[388] Conversely, the failure to adopt such norms constitutes deviance and invites sanction. As Lewis observes, a convention is "a socially enforced norm: one is expected to conform, and failure to

---

[383] "By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him. . ." Smith, 1759/1976.
[384] Smith, 1759/1976.
[385] "I want to understand acceptance of norms as a natural, biological phenomenon. . . . If there is such a thing as governance by norms, there must be psychic mechanisms that accomplish it, and we can ask about their biological function. That function, I want to suggest, is to *coordinate*. . . . " Gibbard, 1997, p186.
[386] Gibbard, 1990, p71, p79.
[387] Gibbard, 1990, p65.
[388] Gibbard, 1990, p26.

conform tends to invoke unfavourable responses from others" in the form of "poor opinions . . . reproaches, punishment, and distrust".[389]

Incidentally, the abundance of norms in human social life has led some to conclude that morality consists of nothing but norms, social constructions, or 'cultural meanings'.[390] This mistake has an ancient pedigree. As Aristotle commented: "In what is just by convention . . . it makes originally no difference whether it is fixed one way or another, but it does make a difference once it is fixed, for example, . . . that a sacrifice shall consist of a goat but not of two sheep . . . . Now, some people [Sophists] think that everything [that is] just exists only by convention, since whatever is by nature is unchangeable and has the same force everywhere – as, for example, fire burns both here and in Persia – whereas they see that notions of what is just change. But this is not the correct view, although it has an element of truth."[391] Not every problem of social life is a coordination problem, hence not every solution to a social problem is a coordination solution.[392]

Finally, friendship. A number of moral philosophers have seen friendship as an important part of the moral life. For Aristotle, friendship revolves around the mutually advantageous pursuit of some common interest:[393]

> We count as a friend (1) a person who wishes for and does
> what is good or what appears to him to be good for his
> friend's sake; or (2) a person who wishes for the existence
> and life of his friend for the friend's sake. . . . (3) a person
> who spends his time in our company and (4) whose desires

---

[389] Lewis, 1969, p99. Traditional sociological accounts of deviance tend not to distinguish between out of equilibrium play in coordination games, the choice of a frequency-dependent equilibrium, and defection due to a dominant incentive; and their analyses suffer as a result. Goode, 1978/97. For example, not all traditional deviance is pathological. I use the term deviance to refer to out-of-equilibrium play in coordination games; under this definition frequency-dependent play and defection are equilibrium choices, and therefore not deviations.

[390] For example, see: Walzer, 1983.

[391] Aristotle, 1962, V, vii, 19-28.

[392] As Lewis observes, not all social regularities are conventions. There are also "rules, laws of nature, generalisations, mathematical truths, strategic maxims, hypothetical imperatives, some threats and warnings, social contracts, overwhelming sanction, or acts for which there are independent instrumental reasons" and so on Lewis, 1969, p100.

[393] Aristotle, 1962, VIII, 1155a 3-4.

are the same as ours, or (5) a person who shares sorrow and joy with his friend. This quality, too, is most frequently found in mothers. By one or another of these sentiments people also define friendship.[394]

The moral importance that Aristotle attaches to friendship is reflected in the fact that he devotes two books of his *Nichomachian Ethics* to the subject – more space than is given to any other topic. Cicero, meanwhile, devoted an entire dialogue to the issue. In *Laelius: On Friendship*, Cicero defined friendship as "a complete identity of feeling about all things in heaven and earth: an identity which is strengthened by mutual goodwill and affection" and argued that the origins of friendship were "primeval and noble, something emanating directly from the actual processes of nature". Cicero urged his readers "to place friendship above every other human concern that can be imagined! Nothing else in the whole world is so completely in harmony with nature, and nothing is so utterly right, in prosperity and adversity alike". Friendship was "the noblest and most delightful of all the gifts the gods have given mankind"; it "ranks as the finest and most glorious of all possible connexions, embodying the highest good that nature has to offer."[395] And, as a final example, in the 20th century, GE Moore ranked friendship as one of "the most valuable things that we can know or imagine". Indeed, Moore went so far as to say that "it is only for the sake of these things [friendship and beauty]. . . that any one can be justified in performing any public or private duty. . . . [They] form the rational ultimate end of human action".[396]

*Reciprocity*

In addition to overcoming coordination problems, potential cooperators must also overcome free-rider problems. In the previous chapters we saw a variety of ways of doing this, culminating in the 'tit-for-tat' strategy of reciprocal altruism, and we reviewed the evidence for adaptations for

---

[394] Aristotle, 1962, IX, 1166a, 1-10.
[395] Cicero, 1971, p187, p191, p191, p185, p201, p217.
[396] Moore, 1903, pp188-9. Peter Singer and J B Schneewind interpret 'pleasurable human intercourse' and 'personal affection' as friendship, and so shall I. Schneewind, 1991, p153; Singer, 1994, p217.

reciprocity in human psychology. The presence of these adaptations can explain why reciprocity is everywhere considered a moral ideal, why moral philosophers have seen trust, commitment, social contracts, punishment, guilt and forgiveness as important aspects of human morality.

Reciprocity involves exposing oneself to the risk of defection. The tendency to take this risk, and open with a nice move, forms the basis of ethical discussions of 'trust'.[397] To take just one example, Annette Baier argues that "Trust is acceptance of vulnerability to harm that others could inflict, but which we judge that they will not inflict", and that we have to rely on trust if we are "to enter and sustain mutually beneficial relationships of trust-involving cooperation".[398] Game theorists employ a similar analysis.[399]

It follows that attempts to reduce uncertainty and bolster trust should have a moral feel to them. One of the ways of reassuring others about one's cooperative intentions is to possess and display signals of a low discount rate ($d$); and indeed, many philosophers have seen a low discount rate as an important component of moral character – in other words, they have recognised that 'patience is a virtue'. John Stuart Mill, for example, in the context of higher and lower pleasures, attributes "election for the nearer good" to "infirmity of character".[400] Henry Sidgwick considered the proposition "I ought not prefer a present lesser good to a future greater good" to be a self-evident moral maxim.[401] Allan Gibbard argues that the capacity to forego short-term pleasures in favour of long-term gains is the product of a "normative control system" that is "peculiar to human beings" and which overrides the "animal control system".[402] And Robert Frank argues that such "self-control devices" make

---

[397] Baier, 1995; Fukuyama, 1996; Hollis, 1998; O'Neill, 2002.
[398] Baier, 1995, p152. For good measure, she adds: "Fidelity is certainly the virtue of those who do not let down others when they have encouraged them to trust them . . . " (p167).
[399] Bacharach and Gambetta, for example, argue that "the primary problem of trust" in social dilemmas is a function of "uncertainty about the payoffs of the trustee". Bacharach and Gambetta, 2001, p1. See also: Hausman, 1997.
[400] Mill and Bentham, 1987, p281.
[401] Sidgwick, 1962, p383.
[402] Gibbard, 1990, p57.

cooperation and morality possible.[403] Conversely, the tendency to heavily discount the future is typically referred to pejoratively as "'impulsivity' and 'short time horizons' or . . . impatience, myopia, lack of self control, and incapacity to delay gratification."[404]

As noted in the previous chapter, the obligation to reciprocate – to keep your promises, to repay your debts, to return favours, and so on – is a feature of the ethical systems of all known societies.[405] It also forms the backbone of many moral philosophies. For example, when asked for a single word that could sum up morality, Confucius answered: "Reciprocity perhaps? Do not inflict on others what you yourself would not wish done to you."[406] More generally, the problem of delayed social exchange and its reciprocal solution provide the logical framework for the 'social contract' tradition in ethics, according to which moral duties and obligations arise as a result of mutually-advantageous agreements entered into by agents acting in their own self-interest. Hobbes provides the classic statement of this position; but one can find the same social-contract view of justice being articulated (by the character 'Glaucon') as early as Plato's *Republic*.[407]

Deterring cheats, by punishment or exclusion, has also been a significant component of substantive moral philosophies. Bishop Butler argues, in the eighth of his *Fifteen Sermons*, that the desire to punish wrongdoers is a moral sentiment.[408] Our indignation and desire to punish, says Butler, is

---

[403] Frank, 1992; Frank, 1990; Frank, 1988.

[404] Steep discounting is not necessarily "dysfunctional". As Daly and Wilson observe: "Steep discounting may be a 'rational' response to information that indicates an uncertain or low probability of surviving to reap delayed benefits, for example, and 'reckless' risk taking can be optimal when the expected profits from safer courses of action are negligible." Wilson and Daly, 1997.

[405] Brown, 1991, p137.

[406] Confucius, 1994.

[407] Hobbes, 1651/1958; Plato, 1974. Indeed, one could argue that Plato anticipated the entire game-theoretic analysis of cooperation and morality. 'Socrates' argues that life is a coordination problem (a division of labour), and that justice ensues when everyone performs their allotted role. 'Glaucon' argues that life is a social contract, and that justice ensues when everyone abides by its terms. 'Thrasymachus' argues that life is a hawk-dove game, and that justice ensues when the weak defer to the strong. 'Socrates' also recognized that his plan for the perfect society could be derailed by the allegiances that individuals have to their kin; hence he proposed to dissolve traditional families among the Philosopher Kings.

[408] Butler, 1856, VIII.

raised by the contemplation of "injury and injustice"; and the purpose of resentment, he says, is to "prevent and to remedy such injury, and the miseries arising from it". "It is to be considered as a weapon put into our hands by nature, against injury, injustice and cruelty." Butler adds: "The good influence which this passion, has, in fact, upon the affairs of the world, is obvious to every one's notice. Men are plainly restrained from injuring their fellow-creatures by fear of their resentment". John Locke notes that punishment is the only basis on which "one man may lawfully do harm to another". "[E]ach transgression," says Locke, "may be *punished* to that *degree*, and with so much *severity*, as will suffice to make it an ill bargain to the offender, give him cause to repent, and terrify others from doing the like".[409] John Stuart Mill observes, "to deter by suffering from inflicting suffering is not only possible, but the very purpose of penal justice".[410] And the legal theorist Richard Posner observes: "Although Aristotle and Kant obviously had no opportunity to read *The Origin of Species*, their ideas about remedial justice – the justice of sanctions for transgressions – are rooted in a view of human nature, as quintessentially vengeful, that is highly compatible with a Darwinist view."[411] And so on. In the words of Daly and Wilson:

> From the perspective of evolutionary psychology, this almost mystical and seemingly irreducible sort of moral imperative [retribution] is the output of a mental mechanism with a straightforward adaptive function: to reckon justice and administer punishment by a calculus which ensures that violators reap no advantage from their misdeeds. The enormous volume of mystico-religious bafflegab about atonement and penance and divine justice and the like is the attribution to a higher, detached authority of what is actually a mundane, pragmatic matter: discouraging self-interested competitive acts by reducing their profitability to nil.[412]

---

[409] Locke, 1992, p740-1.
[410] Mill, 1868/1999, p182.
[411] Posner, 1990, p331.
[412] Daly and Wilson, 1988, p256.

The motivation to make amends after a breach (guilt), and the motivation to re-accept someone into a cooperative scheme after the appropriate costs have been paid (forgiveness) have also figured prominently in moral philosophies.

Allan Gibbard captures the essence of guilt by contrasting it with shame.[413] Both guilt and shame, he says, are emotional adaptations that "respond preeminently to threats to one's place in cooperative schemes". "Shame stems from things that indicate a lack of the abilities, powers, or resources one needs if one is to be valued for one's cooperation and reciprocity. Guilt stems from things that indicate insufficient motivation." Shame – as indicative of failure – is met by others with disdain, indifference, neglect and abandonment. And shame motivates one to withdraw, develop one's powers, and then display them. Guilt – as indicative of transgression, or cheating – is met by others with anger and punishment in an attempt to alter motivation. And guilt motivates attempts to placate others through apology and restitution. Shame, then, would seem to stem from the inability to engage in mutually-beneficial exchange, and perhaps from a lack of the traits needed to resolve conflicts in one's favour. Guilt, meanwhile, stems from a failure to reciprocate.

Forgiveness is a common theme of ethical codes. Hobbes' sixth 'law of nature' states that "a man ought to pardon the offences past of them that repenting, desire it".[414] And as Bishop Butler notes, the purpose of forgiveness is to prevent a "slight offence" becoming "the occasion of entering into a long intercourse of ill offices".[415] Recent philosophical literature on forgiveness follows a similar, and predictable, line. For Downie, a "logically necessary condition of forgiveness" is that the forgiver has been injured by the forgivee.[416] Kolnai defines forgiveness as "re-acceptance" or "reestablishing a relationship" that involves "a venture of trust".[417] In this light, "to forgive is a virtue and inability to forgive, or at

---

[413] Gibbard, 1990, pp138-9.
[414] Hobbes, 1651/1958, p633.
[415] Butler, 1856, IX.
[416] Downie, 1965, p128.
[417] Quoted in Richards, 1988, p77, p92.

least unwillingness to try, is a vice"; "the forgiving spirit is always in itself to be morally approved of"; "an inability to forgive constitutes a moral defect".[418] And so on.

*Conflict resolution: virtue*

We saw how the display of costly signals of fitness could be used to settle or forestall conflicts through face-to-face encounters and through the establishment of stable social hierarchies. These costly signals can be indicators of physical and mental prowess, and can confer benefits on their audience; and in some species, the costly signals seem to include fair and effective leadership. In the previous chapter we reviewed the evidence for homologous adaptations in humans; and we saw that Hume regarded such traits as 'shining' and 'heroic' virtues.[419] Costly signals have been celebrated as 'virtues' by other moral philosophers, including Aristotle, Machiavelli, and Nietzsche.[420]

According to Aristotle, self-control, courage, generosity, honour, truthfulness, wit, modesty, righteous indignation, and magnificence (given to public works) were the principal examples of human 'excellence' or 'virtue'. "Praise is proper to virtue or excellence, because it is excellence that makes men capable of performing noble deeds."[421] Aristotle regarded magnanimity – "greatness of soul" – as the highest of

---

[418] Downie, 1965, p128.

[419] The word 'virtue' comes from the Latin for 'proper to a man', as in 'virile'. "Appelata est enim a viro virtus: viri autem propria maxime est fortitudo." ("The term virtue is from the word that signifies man; a man's chief quality is fortitude.") Cicero, 1945, I, ix, 18.

[420] Does the praise of 'virtue' suggest that might is right? No. First, competition for scarce resources creates problems for which 'excellences' (and 'first possession') are among the solutions. Perhaps the world would be a better place if resources were not scarce, there was no competition, and hence no need for ritual contests and private property. But that is not the world that we live in; and until it is, it hardly makes sense to criticize solutions on the grounds that the problem is unpalatable. A better line of criticism would be to propose alternative, superior, means of settling disputes. Second, excellences and private property solve problems only in the domain of 'chicken'. There is no suggestion that might is generally right, in other domains such as reciprocity or coordination. Different games have different rules, and strategies from one game are not expected to bring about cooperation in another. Third, the inequality manifest in, for example, a dominance hierarchy, may be corrosive to other forms of cooperation that require players to be on a more equal footing. In such circumstances, intervention to redress inequality – and therefore reduce the reliance upon virtue and private property – would be justified.

[421] Aristotle, 1962, p28.

all virtues.[422] The characteristics of the magnanimous man read like the attributes of an idealised alpha male, who competes for status through acts of altruism. The magnanimous man, says Aristotle, is characterised by "a slow gait . . . , a deep voice, and a deliberate way of speaking."[423] Moreover, the magnanimous man will do good turns, but: "is ashamed to accept a good turn, because the former marks a man as superior, the latter as inferior. . . . For the recipient is inferior to the benefactor, whereas a high-minded man wants to be superior. . . . It is, further, typical of a high-minded man not to ask for any favours, or only reluctantly, but to offer aid readily."[424] Aristotle's view of virtue was common in the ancient world. Alisdair MacIntyre reports that, in Homeric societies, virtue (*aretê*) referred to "excellence of any kind"; examples include: physical strength, courage, intelligence, cunning, sense of humour, fidelity, glory (*kudos*), and prosperity.[425]

This classical view of virtue was resuscitated by Machiavelli. For Machiavelli, a Prince's *virtú* involved the display of 'manly qualities' and accomplishments such as strength, skill, prowess, courage, fortitude, audacity, wisdom, valour, virtuosity and civic spirit. These qualities helped to convey a Prince into power; and if they are used not to further the Prince's own interests, but to further the common good, then 'power' is transformed into 'glory'.[426]

And, in *The Genealogy of Morals*, Nietzsche celebrated a 'master morality' that involved the exhibition of such virtues as "bravery, skill, beauty, fertility, strength, pride, leadership, stoicism, sacrifice, tolerance, mercy, joy, humor, grace, good manners, and the creation of social norms".[427] As Geoffrey Miller observes:

---

[422] Aristotle, 1962, IV, iii, 24a.
[423] Aristotle, 1962, IV, iii, 25a.
[424] Aristotle, 1962, IV, iii, 24b. As Aristotle says, "it is hard to be truly high-minded and, in fact, impossible without goodness and nobility".
[425] MacIntyre, 1981, p111, Ch. 10.
[426] Machiavelli, 1988.
[427] Miller, 2000a, pp337-8.

What is striking here is that Nietzsche's virtues sound remarkably like sexually selected fitness indicators. . . . He wrote of virtue as 'a luxury of the first order' which shows 'the charm of rareness, inimitableness, exceptionalness, and unaverageness'. By their luxuriant excess, virtues reveal 'processes of physiological prosperity or failure.' For Nietzsche, virtue was what the strong and healthy could afford to display.[428]

Nietzsche famously contrasted his account of 'master morality' with 'herd' or 'slave morality' that celebrated turning the other cheek, meekness, humility, quietism, stoicism, sacrifice, and patience (of which Christianity was the predominant example). The Humean-Darwinian account seems to have a place for these virtues too. As we have seen, the attributes of the 'superior' cannot settle disputes peacefully without a complementary set of attributes of the 'inferior'. We should therefore expect to find in the literature, not only celebrations of the hawkish virtues of the 'strong', but also praise for the relevant dove-ish virtues of the 'weak'. This might be one way to interpret the 'Christian' virtues, and to explain the traditional antipathy between these and classical virtues.

Finally, costly signals of altruism also seem to encompass what John Rawls refers to as 'superogatory acts', that is: "acts of benevolence and mercy, of heroism and self-sacrifice . . . [that] are not required, though normally they would be were it not for the loss or risk involved for the agent himself". Accounting for these acts, says Rawls, is an issue "of first importance for ethical theory".[429]

---

[428] Miller, 2000a, pp337-8. Conversely, according to Nietzsche, "'Vice is a somewhat arbitrarily limited concept designed to express in one word certain consequences of physiological degeneration'." If we replaced 'physiological degeneration' with 'mutation', this would be a fairly good description of virtue acting as indicators of heritable fitness (Miller, personal communication).

[429] Rawls, 1971, p117.

*Conflict resolution: property*

In addition to costly signalling, in previous chapters we saw how 'prior ownership' could be used to settle disputes about resources. And, it is the case that – from the commandment 'Thou shalt not steal' to maxims such as 'property is nine-tenths of the law' – the right of private property has been a perennial feature of moral and legal systems.

John Locke provides the classic account and defence of private property. In his *Second Treatise on Government*, Locke argues that people have a natural right to their bodies and to the means of sustaining their lives. For this reason, they have a natural right of possession over the fruits of their labour: "[E]very man has property in his own person. This nobody has any right to but himself. The labour of his body, and the work of his hands, we may say, are properly his. Whatsoever, then, he removes out of the state that nature hath provided and left it in, he hath mixed his labour with, and joined to it something that is his own, and thereby makes it his property".[430] Locke proceeds to argue that it is the principal job of government to protect private property. The protection of property ensures that people will be rewarded for their labour; hence the institution of property provides an incentive to industry, and as a result people produce a surplus that can be traded, thereby promoting the common good.[431]

Incidentally, Locke's defence of private property provides a good example of how evolved intuitions can run into trouble in the modern world. Locke assumed the resources available in the state of nature, such as food, were perishable; and he restricted the right of private property to only so much of a resource that could be consumed without waste. The result was a natural limit on the degree of inequality. However, Locke recognised that with the invention of gold and silver money that does not perish, resources could be stockpiled indefinitely. Locke defended the resulting inequalities by highlighting the economic benefits that a

---

[430] Locke, 1992, p746.
[431] Locke was not the first, or the only, person to defend private property. Cicero, for example, had made a similar argument 700 years earlier. Cicero, 1991, pp92-5.

conventional means of exchange can provide. Others, notably Marx and Marxists, have used Locke's own 'labour theory of value' to argue that, under conditions of gross economic inequality, labourers are no longer rewarded in proportion to their efforts.[432] In this way, intuitions about property – that were designed to resolve conflicts under one set of (ancestral) conditions – may no longer operate as intended under modern, technologically-advanced conditions.

In summary, the Humean-Darwinian account of morality suggests that moral values include promoting the common good by caring for family members, adopting local norms, keeping promises, punishing cheats, resolving disputes peacefully, and respecting prior ownership. This whistle-stop tour of substantive moral philosophy suggests that, irrespective of their various meta-ethical starting points, a diverse range of moral philosophers and ethical traditions have come to similar substantive conclusions about the content of morality. One way of interpreting this convergence is that moral philosophers have, in the first instance, been trying to articulate and make sense of the content of their moral intuitions; and that different theorists have articulated different aspects of their moral intuitions and the moral psychologies of those around them, perhaps as a result of the different ways that their moral psychologies have developed under different conditions in different times and places. If this is the case, then the Humean-Darwinian account has the potential to show why these different strands of traditional moral philosophy are not alternatives; and it provides a principled framework for weaving them into a comprehensive account of moral value. In addition, the Humean-Darwinian account suggests that, in the future, we can supplement these second-hand reports of moral values by going directly to the source, investigating moral psychology as we would any other aspect of psychology.

---

[432] For example, see: MacPherson, 1962.

**Applied ethics: How do people make moral decisions?**

Finally, the Humean-Darwinian account of morality can provide an account of what is involved in moral reasoning and moral argumentation, and it can explain why there are such things as moral disagreements and moral dilemmas, and what they are likely to be about. Hence it offers an over-arching theoretical framework from which the problems of 'applied ethics' can be derived.

*Moral reasoning and decision-making*

According to the Humean-Darwinian account, to a first approximation, human moral psychology consists of a suite of evolved decision-rules designed to solve the kinds of collective-action problems that were recurrent in the lives of our ancestors. Hence moral reasoning is no different in kind to reasoning in other domains; it is merely the information-processing carried out by our adaptations for cooperation. Decision rules render a person uncertain about what to do, how to act, whether to approve, and so on; and this uncertainty is resolved with reference to the conditions specified by the rule.[433] In the simplest case, moral reasoning can be represented as a syllogism in which one's moral value constitutes the Major Premise. For example:

> **P1**:     If a person is your elder, then you ought to respect him.
> <u>**P2**:     Bill is your elder.</u>
> **C**:     Therefore, you ought to respect Bill.

In this case, if the moral rule is "Respect your elders" then in order to decide whether to respect any particular individual you will have to answer the question "Are they my elder?" The answer to any particular question depends on the facts of the matter specified as relevant by the rule. Of course, "respect your elders" is a fairly simple moral rule to apply – it requires only one 'bit' of information. Moral reasoning becomes more

---

[433] As in other domains of instrumental reasoning, what constitutes a 'good reason' for acting in a particular way is a function of the prior rule. For example, that 'there is milk in the fridge' is a good reason to go to the fridge only if you want to drink milk. The Humean-Darwinian perspective suggests that there is no such thing as a free-standing reason.

difficult as the moral rule that you wish to apply becomes more complicated – but the principle is the same.[434] Much of applied ethics can be seen as the attempt to articulate our moral intuitions and decision-rules, and make explicit the information processing that they perform.[435]

We should be able to derive from our account of the various adaptations for cooperation predictions about what moral reasoners will be uncertain about, and hence what the salient content of moral discussions will be. For example, Cosmides and Tooby suggest that, in the context of social exchange and redistribution, intuitions regarding luck and effort continue to determine what counts as relevant:

> Consider the political and moral debate concerning the homeless in the United States. . . . One persistent theme among those who wish to motivate more sharing is the idea 'there but for fortune, go you or I.' That is, they emphasize the random, variance-driven dimensions of the situation. The potential recipient of aid is viewed as worthy because he or she is the unlucky victim of circumstances, such as unemployment, discrimination, or mental illness. On the other hand, those who oppose an increase in sharing with the homeless emphasize the putatively chosen or self-caused dimensions of the situation. Potential recipients are viewed as unworthy of aid because they 'brought it on themselves': They are portrayed as able-bodied but lazy, or as having debilitated themselves through choosing to use alcohol and other drugs. The counterresponse from those

---

[434] These kinds of rules could be extended indefinitely and hierarchically. "The principle involved in an 'ought'-statement may be a highly specific, complex and detailed one . . . [M]oral principles do not need to be as general as 'Never tell lies': they can be more specific, like 'Never tell lies except when it is necessary in order to save an innocent life, except when . . . except when . . . and except when . . . In a morally developed person the exceptions may get too complex to get formulated in words." Hare, 1991, p457.

[435] To put it mildly, this will be no easy task. As the editors of one volume put it: "To hope to condense the subtle moral expertise encoded by a biologically realistic neural network into a tractable set of summary principles or moral rules is in all probability a quite hopeless task. In this respect, moral expertise looks set to follow the pattern of expertise in general. Long volumes fail dismally to capture the knowledge of the expert chess player. The knowledge of the successful moral agent is probably no less complex, and no more amenable to brief linguaform summation." May, et al., 1996, p7.

who want to motivate more sharing is to portray drug use not as a choice but as a sickness, and so on.[436]

Furthermore, as in other domains, moral decision-making can take place collaboratively, in a moral argument or discussion, where information relevant to the decision is sought, elicited and offered by others. Information about Bill's age, for example, could come from a third-party. Moral discussions create the opportunity for strategic manipulation among actors with non-identical interests; for this reason, it may be useful to model such interactions as signalling games. Theories of the content of moral intuitions, combined with an appreciation of the structure of signalling games, provides a rigorous framework in which to analyse the form and content of rhetoric and oratory.[437]

*Moral disagreements and dilemmas*

The majority of the time, moral decision-making takes place without fuss or incident; people have no trouble deciding that it is right to help old ladies across the road, or that it is wrong to murder the innocent. However, the Humean-Darwinian account leads us to expect (at least) four instances in which people will experience difficulties in making moral decisions.

First, indecision. In the simplest case, an individual could fail to acquire the information necessary to make a decision, and hence remain in a state of uncertainty. For example, one could fail to ascertain Bill's age, and therefore remain uncertain about whether to respect him or not.

Second, internal conflict. Because individuals pursue multiple goals or values, there will inevitably be situations in which these values come into conflict. At the point at which an individual is indifferent between values, they will remain in a state of uncertainty and experience a 'dilemma'. When one or more of the values involved is a moral value, the person will

---

[436] Cosmides and Tooby, 1992.
[437] Dawkins and Krebs, 1978; Sperber, 2001.

experience a moral dilemma. For example, Bill might be an elder, and also a cheat, requiring that he be punished and respected. Or, to take Sartre's famous example, one could feel a duty to attend to one's ailing mother, and also an incompatible duty to go off and fight the Nazis.

Third, conflict with others. Moral disagreements will arise among individuals when they have different values, or when people have different information, or both.[438] Adaptations for cooperation, like other adaptations, are expected to exhibit a normal distribution in the population, and to exhibit systematic age and sex differences (see Chapter 7). We also expect moral psychology to develop and operate differently under different conditions, thereby giving rise to cross-cultural differences in the manifestation of moral values. As a result, we should expect there to be individual differences in moral values; and individuals with different values will not be able to agree on what they ought to do, even when they have access to the same information. (They may also disagree about what constitutes relevant information.) A similar analysis pertains when individuals with the same values have access to different information, although in these cases there is a more tractable means of settling the dispute.

Fourth, evolutionarily-novel environments. Human moral intuitions were designed to solve Stone-Age collective-action problems; there is no guarantee that they will be able to properly frame, engage with, or generate stable conclusions about all the apparent problems created by modern technology and the modern world in general.[439] It may be the case, for example, that intuitions designed to distribute meat after a successful hunt are not the best guides as to how to distribute healthcare or unemployment benefit.[440] More specifically, we might expect our

---

[438] Arnhart provides an extensive account of this kind of moral disagreement. Arnhart, 1998.
[439] For further discussion, see: Pinker, 2002.
[440] For example, the discussions of "brute and option luck" – sometimes referred to as "chance and choice" – illustrate how intuitions regarding luck versus effort offer little guidance when employed outside of their original domain. One stable conclusion of much recent literature is that, to the extent to which the talents with which one is born are the product of a genetic lottery, a certain amount of redistributive taxation is permitted. However, to the extent that one's income is a product of one's effort, redistribution is not warranted. But these intuitions begin to break down a) when one's genetic endowment is seen not as a product of luck, but of the effort one's parents

ancient moral intuitions to generate *false positives* and *false negatives*. False positives occur in situations in which we intuit that there is a collective-action problem, and engage our moral psychologies, when in fact no such problem exists. Perhaps the controversies surrounding recent developments in biotechnology, such as cloning, or stem-cell research, are examples of false positives. False negatives occur in situations in which we fail to intuit that there is a collective-action problem, and hence do not engage our moral psychologies, when in fact such a problem exists. Perhaps the continuing degradation of the environment – a 'tragedy of the commons' on a global scale – is an example of a false negative.

A correct account of moral psychology should be able to predict and explain why and where certain moral dilemmas and disagreements arise. Hence moral disagreements and dilemmas are a useful source of evidence for moral psychology, in much the same way that optical illusions are a useful source of evidence when studying the psychology of vision.

**Conclusion**

I have followed Hume and Darwin in assuming that morality is a natural phenomenon that can be studied by the natural sciences. I have argued: that human morality is an adaptation to social life; that it consists of a range of social, cooperative and altruistic dispositions that promote cooperation; and that moral reasoning can be treated in the same way as reasoning in other domains. This parsimonious account of morality is consistent with what we know about human origins, human psychology and human behaviour; it re-structures the problems and solutions of moral philosophy; and it incorporates – and provides a principled structure for – a wide variety of theories in traditional meta-, substantive and applied ethics. For these reasons, the Humean-Darwinian account should be adopted as the default explanation of morality, and it should constitute the theoretical framework within which future research in moral psychology, moral philosophy and ethics takes place.

---

put into earning the money necessary to fund genetic engineering; and b) when we discover that 'effort' is heritable. Dworkin, 2000, grapples with these conflicting intuitions.

In the final chapter I review some of the directions that this research could take; they include broadening and deepening our understanding of human adaptations for cooperation, filling the gaps in the existing literature, and assessing the extent to which evolved moral sentiments constitute optimal solutions to collective-action problems. But first, I shall have to deal with the objection that such a project is ill-conceived from the outset, on the grounds that any attempt to find a naturalistic base for ethics commits a fallacy.

## Chapter 6 – Who's afraid of 'the naturalistic fallacy'?

> [T]he 'is'/'ought' gap, *and* the naturalistic fallacy *are perhaps better seen as warnings than as outright barriers, reminding us of ways in which the project can fail, and indeed often has failed. But they should not warn us off the project altogether, since the need to ask how morality fits with our best empirical understanding of ourselves and our place in nature and history arises from within normative moral thought itself.*
>
> – Peter Railton[441]

**Introduction**

This thesis has argued for a Humean-Darwinian meta-ethic that sees moral values as the proximate goals of adaptations for cooperation. The previous chapter outlined the moral philosophy that follows from this starting point.

Few philosophers, however, have been willing to pursue the implications of this Darwinian update of Hume's meta-ethics for fear that the thesis commits something called "the naturalistic fallacy".[442] As Michael Ruse puts it: "Everybody knows (or 'knows') that it [evolutionary ethics] has been the excuse for some of the worst kinds of fallacious arguments in the philosophical workbook . . . It has been enough for the student to murmur the magical phrase 'naturalistic fallacy', and then he or she can move on to the next question, confident of having gained full marks on the exam."[443] Although, as I will show, there are many versions of the naturalistic fallacy, it has come to be synonymous with the widespread belief that (someone, somewhere, has demonstrated conclusively that) the natural and the normative inhabit two entirely separate realms, and never the twain shall meet.[444] One consequence of this view is that contemporary moral philosophy has tended to neglect the empirical

---

[441] Railton, 2000, p126.
[442] The work of Michael Ruse is the notable exception to this rule. See, for example: Ruse, 1995.
[443] Ruse, 1995, p223.
[444] Held, 2002.

sciences.[445] Another consequence is that contemporary moral philosophers have tended to look somewhere other than 'nature' for their account of moral value.

But what is "the naturalistic fallacy"? And does it really present a lethal challenge to a Humean-Darwinian meta-ethic? The purpose of this chapter is to show that such fears are misplaced. None of the arguments that go by the name of "the naturalistic fallacy" constitute arguments against the Humean-Darwinian meta-ethics presented above.


**The naturalistic fallacies**

The first thing that anyone wishing to investigate the naturalistic fallacy discovers is that there is not one argument that goes by this name; there are many. A survey of the literature reveals not one but (at least) eight alleged mistakes that carry the label "the naturalistic fallacy":

1. Moving from is to ought (Hume's fallacy).
2. Moving from facts to values.
3. Identifying good with its object (Moore's fallacy).
4. Claiming that good is a natural property.
5. Going 'in the direction of evolution'.
6. Assuming that what is natural is good.
7. Assuming that what currently exists ought to exist.
8. Substituting explanation for justification.

For example: 1) Daniel Dennett says that the naturalistic fallacy involves the derivation of "'ought' from 'is'".[446] 2) Peter Singer assumes that the naturalistic fallacy involves "defining values in terms of facts".[447] 3) Charles Pigden, meanwhile, argues that the naturalistic fallacy consists of confusing "the property of goodness with the things that possess that

---

[445] Doris and Stich, (forthcoming).
[446] Dennett, 1995, p467. And Frans de Waal claims that the naturalistic fallacy "has to do with the impossibility of translating 'is' language (how things are) into 'ought' language (how things ought to be)". de Waal, 1996, p38.
[447] Singer, 1981, p74.

property or with some other property that good things possess".[448] 4) Simon Blackburn states that the naturalistic fallacy "consists of identifying an ethical concept with a 'natural' concept".[449] 5) Robert Wright claims that the naturalistic fallacy involves "drawing values from evolution or, for that matter, from any aspect of observed nature".[450] 6) Dylan Evans claims that "[a]rguing that something is good *because it is natural* is called the 'naturalistic fallacy'".[451] 7) David Buss states that "the naturalistic fallacy . . . maintains that whatever exists should exist".[452] 8) William Rottschaefer claims that one version of the naturalistic fallacy is committed when "causal explanations are confused with or substituted for justifying reasons".[453]

My concern here is not with which is the 'true' fallacy. Rather, my aim is to bring all the versions of the argument into the light, so that we may examine them one by one, and thereby establish whether they constitute serious obstacles to a Humean-Darwinian meta-ethic. Let's begin with Hume's fallacy.

## 1. Moving from is to ought (Hume's fallacy).

Although Hume's theory is primarily a meta-ethical account of the nature and ontological status of morality, it segues into normative or substantive ethics in the following way. The ends supplied by the passions provide the first premises of chains of means-end reasoning. For example, if you

---

[448] Pigden, 1991, p426.

[449] Blackburn, 1994, p255. Larry Arnhart maintains that the naturalistic fallacy is the mistaken attempt to "to define 'good' in natural terms" Arnhart, 1998, p82. Anthony O'Hear claims that to deny that "goodness could consist in no natural property" is to commit the naturalistic fallacy. O'Hear, 1985, p258.

[450] Wright, 1994, p330.

[451] Emphasis added. Evans and Zarate, 1999, p163. According to Anthony Flew, to commit the naturalistic fallacy is to offer "some supposedly neutral descriptive statement about what is allegedly natural as if it could by itself entail some conclusion about what is in some way commendable". Flew, 1978, p148. Matt Ridley hedges his bets: "The naturalistic fallacy . . . is to argue that what is natural is moral: deducing an 'ought' from and 'is'." Ridley, 1996, p257. As does Joshua Greene: the naturalistic fallacy is "the mistake of identifying that which is natural with that which is right or good (or more broadly, the mistake of identifying moral properties with natural properties)." Greene, 2003, p847.

[452] Buss, 1994, p16. John Beckstrom concurs: "The fact that something exists and has existed for a very longtime does not call for society's endorsement. To say that it does is to commit the 'naturalistic fallacy'." Beckstrom, 1993, p2.

[453] Rottschaefer, 1997, p8. This fallacy is sometimes rolled together with the 'the genetic fallacy'. Baggini and Fosl, 2003, pp89-92.

*want* to get to Grand Central Station, and if the Number 4 train takes you there, then you *ought* to take the Number 4 train. It follows from this view of human psychology – and in particular, from this instrumentalist account of reason – that, in the absence of any passions, desires, or ends, reason alone cannot tell you what you ought to do.[454] Hume applied the same argument to morality. Moral passions provide the first premise of moral reasoning: if you *want* to punish cheats, and if Derek cheated, then you *ought* to punish Derek. In the absence of moral passions, reason alone cannot tell you what you morally ought to do.[455] As Hume put it: "Morals excite passions, and produce or prevent actions. Reason itself is utterly impotent in this particular. The rules of morality, therefore, are not conclusions of our reason."[456]

At the beginning of his account of the moral passions – in a chapter entitled *Moral distinctions not deriv'd from reason* – Hume pauses to castigate rationalist moral philosophers who neglect to mention the passions, and who proceed instead as if one *could* derive ought statements using reason alone. Reading such philosophers, Hume is "surpriz'd to find, that instead of the usual copulations of propositions, *is* and *is not*, I meet with no proposition that is not connected with *ought* or *ought not*."[457] Hume points out that these rationalist philosophers have not explained how reason is capable of such a feat. According to Hume, reason – or 'the understanding' – works by 'relating' 'simple' ideas to one another in order to form 'complex' ideas. There are seven 'relations': "resemblance, identity, relations of time and place, proportion in quantity or number, degrees in any quality, contrariety, and causation".[458] Rationalist philosophers, says Hume, have not "observ'd" or "explain'd" any additional relation that would allow them to move directly from is to ought, nor have they shown how such a move is possible by means of any of the other relations: "[A]s this *ought*, or *ought not*, expresses some new

---

[454] "Reason alone can never be a motive to any action of the will". "[R]eason alone can never produce any action, or give rise to volition". Hume, 1739/1985, p460, p462.

[455] Or, as Alisdair MacIntyre puts it: "we can connect the facts of the situation with what we ought to do only by means of one of those concepts which Hume treats under the heading of the passions" MacIntyre, 1959, pp465-6.

[456] Hume, 1739/1985, p509.

[457] Hume, 1739/1985, p521.

[458] Hume, 1739/1985, p117.

relation or affirmation, 'tis necessary that shou'd be observ'd and explain'd; and at the same time that a reason should be given for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it."[459]

Of course, one can take issue with Hume's account of psychology. But the question under consideration here is whether this standard point about the components of means-end reasoning constitutes an argument against a Humean-Darwinian meta-ethic. And the answer is clearly: no. The claim that "means-end reasoning must include a statement of ends" is not an argument against the claim that "values (ends) are the product of certain natural passions". It is simply the case that Hume's straightforward point about the logic of moral arguments is neutral with regard to meta-ethical arguments about the nature or ontological status of moral values; and it is perfectly consistent with Hume's own naturalistic view of the nature of moral value.

## 2. Moving from facts to values

Despite Hume's naturalistic approach to ethics, and despite it being the case that Hume's simple logical point about moving 'from is to ought' is entirely neutral with regard to the ontological status of moral values, Hume is widely regarded as having demonstrated a fundamental flaw in naturalistic ethics. How did this happen? Part of the explanation is that Hume's dictum is often presented in terms of the impossibility of moving from 'facts to values'; and that this formulation tends to conflate 'values' and 'oughts'. Let *value* refer to the first premise of a means-end syllogism, let *fact* refer to the second premise, and let *ought* refer to the conclusion. Hume argued, first, for a naturalistic account of *value*; and, second, that in the absence of *values*, *facts* alone are insufficient to arrive at *oughts*. However, if one presents this second argument – about moving 'from facts to oughts' – in terms of the impossibility of moving 'from facts to *values*' it can look *as if* Hume's second argument contradicts his first.

---

[459] Hume, 1739/1985, p521.

So, for example, Peter Singer takes E. O. Wilson to task for committing the mistake of moving 'from is to ought' – specifically, for moving from the premise "Our genes came from a common pool and will return to a common pool" to the conclusion that "we ought not to do anything which imperils the human gene pool".[460] Singer correctly points out that this argument is invalid as it stands – Wilson has neither made explicit, nor defended as a moral value, the premise: "We ought not to do anything which imperils the long-term survival of our genes". Then, invoking nothing more than Hume's "unbridgeable gulf between facts and values", Singer asserts that "ethical premises are not the kind of thing discovered by scientific investigation" and hence "[n]o science is ever going to discover ethical premises inherent in our biological nature".[461] But clearly, the demonstration that Wilson has neglected to provide or defend a particular value-premise does not constitute an argument against the more general Humean meta-ethical thesis that 'values are the products of passions that are inherent to human nature'. Singer's assertion does not follow from Hume's dictum, and Singer provides no further arguments in its support.

### 3. Good is identical with its object (Moore's fallacy)

As we have seen, Hume compared moral values to "sounds, colours, heat and cold" which "are not qualities in objects, but perceptions in the mind."[462] George Edward Moore's moral philosophy is exactly the same as Hume's in this regard.

Moore argued that *good* is an "object of thought" that is evoked by certain features of the world (such as "the pleasures of human intercourse and the enjoyment of beautiful objects").[463] Moore compared *goodness* to *yellowness*, in that both are subjective psychological entities and not objective features of the world. Moore sought to establish that *good* was

---

[460] Singer, 1981, p80.
[461] Singer, 1981, p73, p77. More specifically, Singer argues that no "facts about our evolutionary history, our biology or the origins of altruism", "can compel me to accept any value, or any conclusion about what I ought to do" (p75).
[462] Hume, 1739/1985, pp520-1.
[463] Moore, 1903, p188.

an object of thought – and not an objective feature of the world – by employing the famous "open question" argument. According to this argument, whenever one claims that *good* is identical with external object X, it is always possible to ask "Is X good?" Hence *good* and X are not the same thing because we can conceive of them separately. We have, according to Moore, "two different notions before our minds".[464] Moore also argued that *good* was a "simple" – as opposed to a "complex" – object of thought, in the sense that it was not composed of any parts.[465] Hence it was not possible to "define" *good* – by which Moore meant render a complex idea into the simple ideas of which it is composed.[466] *Good* is a "simple, indefinable, unanalysable object of thought".[467]

It follows from this account of *good* as an "object of thought" that it is a mistake to identify the thought *good* with the states of the world that evoke the thought, for the same reason that it is a mistake to identify the thought *yellow* with the objects in the world (for example, light of a certain wavelength) that evoke the thought. Moore called the mistake of identifying an object of thought with its object a *fallacy*. And if the object – with which one mistakenly identified the thought – happened to be a natural object, as opposed to metaphysical entity, then the error became the "*naturalistic* fallacy".[468]

The question here is not whether Moore made a good case for his moral philosophy.[469] Rather, the question is whether Moore's naturalistic fallacy constitutes an argument against Humean-Darwinian meta-ethics. Does Moore's contention that "*good* is a subjective psychological entity and not an objective feature of the world" contradict the Humean-Darwinian thesis that moral values "are not qualities in objects, but perceptions in the mind"? No, it does not. Hume and Moore begin with exactly the same premise: that "*good* is a subjective psychological entity and not an

---

[464] Moore, 1903, p16.
[465] Baggini and Fosl, 2003, p204.
[466] "[A] definition states what are the parts which invariably compose a certain whole; and in this sense 'good' has no definition because it is simple and has no parts." Moore, 1903, p9.
[467] Moore, 1903, p21.
[468] Moore's fallacy is not peculiar to ethics: "The naturalistic fallacy has been quite as commonly committed with regard to beauty as with regard to good." Moore, 1903, p201.
[469] For a critical overview, see: Sturgeon, 2003.

objective feature of the world". And both argue that the opposite view is mistaken. The main difference is that Hume calls the opposing view "vulgar" whereas Moore calls it "fallacious".

## 4. Good is a natural property

Thus far, Moore's moral philosophy is entirely neutral with regard to the ontological status of moral values. However, it is widely believed that Moore went on to argue that *good* – *qua* object of thought – was non-natural in the sense of being not part of the natural world. Hilary Putnam, for example, assumes that Moore demonstrated that: "Good was a 'non-natural' property, i.e. one totally outside the physicalist ontology of natural science."[470] This has led to a literature in which some assume that it is a fallacy to claim that moral goodness is part of the natural world, and in which others point out that to begin by supposing that good is a non-natural property is merely to beg the question.[471]

If Moore had conclusively demonstrated that *good* was a non-natural property in this sense, then Humean-Darwinian meta-ethics would indeed be in trouble. However, Moore made no such argument. And there is no such argument to be found in the pages of *Principia Ethica*. This mistaken interpretation of Moore can be traced to his somewhat idiosyncratic use of the term "natural". For Moore, the opposite of "natural" is "intuitive", not "supernatural".[472] Moore used "natural" to

---

[470] Putnam, 1981, p206.
[471] As Michael Ruse comments: "Why should one claim that goodness is a non-natural property? Surely that is to presuppose the very point at issue?" Ruse, 1995, p230. And as WK Frankena observes, "the charge of committing the naturalistic fallacy can be made, if at all, only as a conclusion from the discussion and not as an instrument of deciding it." Frankena, 1939, p465.
[472] "Naturalism has had two meanings in twentieth-century discussions relating to ethics. In a wider sense 'naturalism' has stood for a general philosophical point of view the essence of which is a denial of supernaturalism . . . . But since the publication of G. E. Moore's *Principia Ethica* in 1903 . . . 'naturalism' has had also a more special meaning when used in ethics; it has meant any ethical theory which holds that an ethical judgment is simply a true or false ascription of a definable and natural (or empirical) property to an action, object, or person. Opposed to ethical naturalism in this usage is not supernaturalism, but a pair of positions: (a) intuitionism, also called 'non-naturalism', which holds that an ethical judgment is a true or false ascription to something of an indefinable and non-natural (or non-empirical) property, and (b) non-cognitivism, which in its extreme form claims that an ethical judgment is not a true or false ascription of any property to anything, but something very different, like an interjection, a command, a wish, a resolution, or a prescription." Frankena, 1957, p457.

refer to properties of the external world.[473] He contrasted "natural" with "intuitive", which he used to refer to properties of the mind – including "objects of thought" such as *good*. Hence when Moore claims that *good* is not a natural property, he is simply restating the point that *good* is an intuitive "object of thought" and not an objective feature of the outside world. Moore made no further arguments to the effect that moral goodness or intuitions were supernatural. On the contrary, Moore maintained that *good* was an "object of thought", and that "in so far as . . . thoughts did exist, they too are natural objects".[474] Neither did Moore see the "naturalistic" part of his fallacy as crucial: "Even if [*good*] were a natural object, that would not alter the nature of the fallacy nor diminish its importance one whit. All that I have said about it would remain quite equally true: only the name which I have called it ["the *naturalistic* fallacy"] would not be so appropriate as I think it is. And I do not care about the name: what I do care about is the fallacy."[475]

Much of the confusion surrounding "the naturalistic fallacy" stems from the fact that Moore made several further distinct arguments against Herbert Spencer's version of evolutionary ethics.[476] If Moore's classic naturalistic fallacy presents no obstacles to a Humean-Darwin meta-ethic, then perhaps one of his other arguments against evolutionary ethics does.

**Fallacies 5, 6, & 7**

In his critique of Spencer, Moore argued that: the direction of evolution is not necessarily good (5); and that the moral worth of something is not decided by its being natural (6). Perhaps as a spin-off of fallacy number five, some people have also claimed that it is a fallacy to argue that what currently exists (perhaps because it is 'natural') ought to exist (7). Under this view, statements such as "evolution is a progressive force, and we

---

[473] As Sturgeon puts it, for Moore: "Natural objects and qualities . . . are objects of experience (sec. 25, pp. 38-39), of perception; they are what we can 'touch and see and feel' (sec. 66, pp. 110-11)." "[N]atural qualities are ones that we can know about by 'empirical observation and induction' (sec. 25, p. 39)." Sturgeon, 2003, p541.

[474] Moore, 1903, p41.

[475] Moore, 1903, p14.

[476] In this paper I will not be concerned with the question of whether Moore accurately represented Spencer's moral philosophy. See MacIntyre, 1966, p251.

*ought* to help it on its way – and prevent humankind from becoming 'less evolved' – by practicing various forms of eugenics"; "males are naturally more promiscuous than females, and therefore it is *morally acceptable* for them to be more promiscuous"; or "males have always been more promiscuous than females, therefore one *ought not* to change the natural order of things", are taken to be examples of fallacious or invalid arguments.

Let us accept, without further argument, that 'fallacies' 5, 6 & 7 are valid objections to spurious uses of evolutionary theory. The question here is whether any of these objections constitute arguments against the Humean-Darwinian meta-ethical thesis that "moral values are the products of certain natural human passions". And again, the answer is: no. First, according to the orthodox evolutionary theory on which the Humean-Darwinian thesis is based, evolution has no direction. And so the notions of 'moving in the direction of evolution' or of being 'more or less evolved' have no content, and play no role in the thesis. Second, the Humean-Darwinian thesis argues that 'all moral values are natural phenomena', but it does not argue, and nor does it follow, that 'all natural phenomena are moral' or even 'all natural values are moral'. As we have seen, according to the Humean-Darwinian thesis the test of whether a passion is moral is whether it promotes 'the common good', not whether it is natural.[477] And besides, to a naturalist, all possible states of the universe are equally natural, and therefore 'naturalness' cannot act as a criterion of anything.[478] Third, according to the Humean-Darwinian thesis, the 'moral goodness' of an institution has to do with whether or not it promotes the common good, not whether it exists or not. A biological explanation of a given state of affairs does not justify it any more than a sociological explanation does.

---

[477] As Hume puts it: "[N]othing can be more unphilosophical than those systems, which assert, that virtue is the same with what is natural, and vice with what is unnatural. For in the sense of the word, Nature, as opposed to miracles, both vice and virtue are equally natural . . .". Hume, 1739/1985, p526.

[478] Radcliffe Richards, 2001, p246.

This brings us on to the final version of the naturalistic fallacy: the objection that Humean-Darwinian ethics can explain, but not justify moral values.

## 8. Explanation and justification

Let us, for the sake of argument, suppose that the Humean-Darwinian thesis is correct, and that moral values are the projections of certain human passions. Some critics of this thesis have argued that even if the naturalistic or evolutionary theories may be able to *explain* and *describe* the nature and content of moral value, they cannot *justify* those moral values. Since the true business of ethics involves the justification of moral values, so the argument goes, there are certain aspects of ethics that will forever remain beyond the reach of naturalistic approaches. In short, there is more to ethics than evolution.

For example, Christine Korskgaard writes: "When we seek a philosophical foundation for morality we are not looking merely for an explanation of moral practices. We are asking what *justifies* the claims that morality makes on us." She continues: "People who take up the study of moral philosophy do not merely want to know why those peculiar animals, human beings, think that they ought to do certain things. We want to know what, if anything, we *really* ought to do."[479] Korskgaard calls this "the normative question". And, as Paul Farber notes, "to go beyond description, to enter the arena of the normative, that is, to say what ought to be, involves an important shift that requires justification. Here, sociobiology provides no new basis, no new foundation, no new hope."[480]

However, objections of this kind rest on two faulty assumptions.

First, the normative objection misconstrues the Humean-Darwinian argument. It supposes that the Humean-Darwinian thesis is merely an

---

[479] Korsgaard, 1996, pp9-10, p13. Emphasis added.
[480] Farber, 1994, p156

explanation or description of what people happen to do or think. If this were the case, then the call for justification would be in order: one could attempt to justify particular opinions and practices with reference to some higher-order moral value. However, like all meta-ethical theories, the explicandum of the Humean-Darwinian thesis is not 'what people happen to think or do'. Rather, the explicandum is the nature and ontological status of moral values.[481]

Second, the normative objection assumes that the question "Are moral values morally justified?" is capable of a coherent answer (and that it is the job of meta-ethics to provide such an answer). However, this assumption would seem to be unwarranted. To recall an earlier distinction, you can justify *ought* statements with reference to certain *values*, but what are you supposed to justify *values* by? What you need is some super-ordinate criterion of justification. And yet it is exactly the existence of such super-ordinate criteria that any meta-ethical theory denies. If there were super-ordinate criteria, then one would have a different meta-ethical theory (and you would be back to square one). Clearly, one cannot go on justifying statements forever, one must come to a stop somewhere. And where one comes to a stop constitutes one's meta-ethical theory. Theologians stop at divine commands, relativists stop at social conventions, Humeans stop at certain passions. In the words of Hume:

> Ask a man *why he uses exercise*; he will answer, *because he desires to keep his health*. If you then enquire, *why he desires health*, he will readily reply, *because sickness is painful*. If you push your enquiries farther, and desire reason *why he hates pain*, it is impossible he can ever give any. This is an ultimate end, and is never referred to any other object. . . . And beyond this it is an absurdity to ask for a reason. It is impossible there can be a progress *in infinitum*; and that one thing can always be a reason why another is desired.

---

[481] If moral values *were* the products of human passions, then the Humean-Darwinian thesis would, in addition, partly explain why people thought they ought to do certain things.

Something must be desirable on its own account, and because of its immediate accord or agreement with human sentiment and affection.[482]

Hume, and numerous other philosophers, have concluded that when you get to this point, the call for justification is "senseless", "unanswerable" or "mistaken", and all that remains to be done is explain the origins and ontological status of moral values.[483] Korsgaard provides no reason for overturning this conclusion. She begins by clearly stating the difficulties that one encounters when attempting to answer the normative question:

So we are faced with a dilemma. . . . If we try to derive [the authority of morality] from some supposedly normative consideration, such as gratitude or contract, we must in turn explain why that consideration is normative, or where its authority comes from. Either its authority comes from morality, in which case we have argued in a circle, or it comes from something else, in which case the question arises again, and we are faced with an infinite regress.[484]

And, faced with the option of either circularity or infinite regression, she opts for circularity. Korsgaard's answer to the normative question is "reflective endorsement". Reflective endorsement is the view that values are justified if, having reflected upon them, we approve or endorse them. Because Korsgaard is a Humean to the extent that she takes moral values

---

[482] Hume, 1777, pp244-5.

[483] Thomas Nagel has written: "Denial that justification is the appropriate final defence of ethics suggests the familiar view that the question 'Why should I be moral?' is senseless or in principle unanswerable. Strictly the suggestion is correct . . . I believe that an explanation can be discovered for the basic principles of ethics, even though it is not a justification. . . . Psychology . . . may therefore be the appropriate field in which to make progress in ethical theory." Nagel, 1970, p5. Michael Ruse writes: "evolution *explains* (not justifies) morality in the sense of showing where it came from. Furthermore, once such an explanation is given, one sees that the traditional call for justification is mistaken. There can be no ultimate support for morality in the sense of reasoned absolute foundations." Ruse, 1987, p427. "The view of reason adopted by sociobiology is Humean . . . an instrument that allows us to calculate the best means to the attainment of our ends. . . . What it cannot do, however, is evaluate the ends finally accepted as ultimate, for these are given by the passions and, at this level, reason is a slave of the passions. And where do these basic passions come from? Evolutionary biology surely has, at least, part of the answer to this question." Murphy, 1982, p100.

[484] Korsgaard, 1996, p30.

to be the expression of a moral sense, reflective endorsement involves the moral sense approving of itself. In other words, the criterion that Korsgaard proposes as a measure of whether moral values are justified is those same moral values. So, morality is moral if: "The moral sense approves of its own origins and workings and so approves of itself."[485] Unfortunately, Korsgaard does not explain how the values embodied in the moral sense can be used as a standard of justification when it is exactly the justifiability of these values that is at issue.

In case this is all rather abstract, let me re-state the point by way of an analogy. Suppose that instead of being about morality and why people find certain things morally good and bad, this dissertation had been about sweetness, and why people find certain things sweet and certain things sour. The Humean-Darwinian would have argued that humans have an evolved digestive system that distinguishes between good and bad sources of nutrition and energy; and that the human 'sweet tooth' is an evolved preference for foods with high sugar-content over foods with low sugar-content. If one accepted this premise, it would make no sense to complain that evolution may have *explained why* humans find certain things sweet, but it cannot tell us whether these things are *really* sweet or not. It follows from the premises of the argument that there is no criterion of sweetness independent of human psychology, and hence this question cannot arise. Of course, one may object to the premise, and claim that there really is such a thing as sweetness that human psychology is latching on to. But then this would be a different argument. One would have to make the case for an independent criterion of sweetness, and not merely assume it. Applying this to morality: The Humean-Darwinian argues that humans are equipped with a suite of adaptations for cooperation, that these adaptations constitute what have been called the moral passions or moral sentiments, and that these adaptations determine what people deem morally good and bad. If one accepts this argument, it makes no sense to complain that evolution may have *explained why* humans find certain things morally good, but it cannot tell us whether these things are *really* morally good or not. It follows from the premises of

---

[485] Korsgaard, 1996, pp62-3.

the argument that there is no criterion of 'moral goodness' independent of human psychology, and hence this question cannot arise. Of course, one may object to the premise, and claim that there really is such a thing as 'goodness' that human psychology is latching on to. But then this would be a different argument. One would have to make the case for an independent criterion, and not merely assume it.[486]

In summary, the normative question mistakes the Humean-Darwinian thesis for an argument about what people believe, when it is in fact an argument about the ontological status of moral values; and the normative question presupposes, incorrectly, that it makes sense to morally evaluate moral values. Given that it is not possible to answer a senseless question, the failure of Humean-Darwinian meta-ethics to answer the normative question does not count against the thesis. In time, the 'failure' of Humean-Darwinian ethics to provide a justification of moral values may come to be seen as a strength rather than a weakness.

**Conclusion**

This chapter has discussed eight different versions of the "naturalistic fallacy", and shown that none of them constitute obstacles to Humean-Darwinian meta-ethics. Of course, there may be other versions of the naturalistic fallacy, or other arguments altogether, that succeed in establishing that moral values inhabit a realm distinct from the natural, rendering Humean-Darwinian and other naturalistic meta-ethics untenable. What would such an argument look like? Presumably, such an argument must establish that there is something other than the natural world, that moral values reside there, and that such entities can somehow exert an influence this world. Any such arguments would seem to be mystical, dualist, transcendental, supernatural or theological in flavour.

---

[486] One of the things that seems to unite the various naturalistic fallacies is that they are motivated by a disagreement with the basic premise of ethical naturalism; they dispute the claim that values are merely the product of natural human passions. 'The naturalistic fallacy' is often invoked in an attempt to soften up evolutionary ethical naturalism before a replacement meta-ethic is proposed. However, one needs to make an argument in favour of these alternative sources of value, and not merely smuggle it in under the guise of a logical fallacy.

In the absence of any such arguments, we can conclude that the Humean-Darwinian version of ethical naturalism remains a "live option".[487]

The naturalistic fallacy, by contrast, seems to have become something of a superstition. It is dimly understood and widely feared, and its ritual incantation is an obligatory part of the apprenticeship of moral philosophers and biologists alike. But if the arguments presented above are correct, then it is surely time to dispense with this superstition. To that end I make the following recommendation: Whenever someone uses the term "naturalistic fallacy", ask them "Which one?", and insist that they explain the arguments behind their accusation. It is only by bringing the 'fallacy' out into the open that we can break the mysterious spell that it continues to cast over ethics.

---

[487] As Charles Pigden comments: "[T]here is no need for naturalists to evade the arguments of Moore and Hume . . . Insofar as they are valid, Hume's arguments, and Moore's too, are compatible with naturalism. Formal attempts to refute naturalism having failed, it remains a live option." Pigden, 1991, pp427-8.

## Chapter 7 – The future of ethics

*Someone may ask, 'What is the difference, then, between moral philosophy and moral psychology?' Well, who says there is an interesting difference? Surely, the issue is of interest only to university administrators.*

　　　　　　　　　– Gilbert Harman[488]


### Introduction

David Hume argued that human nature is equipped with a range of moral passions that give rise to human moral values; and he suggested that morality could be studied like any other part of the natural world. This thesis has argued that, with a little help from Darwin, Hume was basically right. Human nature comprises a suite of adaptations for cooperation; these adaptations constitute human moral sentiments, and their proximate goals constitute human moral values; hence human morality can be studied like any other branch of natural history.

The purpose of this chapter is to preview some of the directions that moral philosophy, pursued in a Humean-Darwinian spirit, might take in the future. I suggest that, first, we should aim to increase our knowledge of human adaptations for cooperation. And second, we should evaluate the design of these adaptations to see whether they provide efficient solutions to collective-action problems.


### More problems

The goal of "evolutionary moral psychology" is to arrive at a full circuit-diagram of human adaptations for cooperation. Chapter 4 reviewed the progress that has been made so far towards this goal (see also Appendix 3). We saw how evolutionary theorists use game theory to identify distinct problems of cooperation, use these problems to derive predictions about the design of possible adaptive solutions, and then go into the field

---

[488] Harman, 2003, p15.

to look for the predicted design features in humans and other organisms. And we reviewed the evidence that showed that, like many other species, humans seem to have adaptations for kin altruism, coordination to mutual advantage, reciprocity and conflict resolution.

Future research in this area will attempt to be more specific about the particular adaptive problems that fall within each of these broad categories, and thereby attempt to arrive at a complete list of human moral sentiments. Again, this process begins by generating hypotheses about additional collective-action problems that were recurrent in the lives of human ancestors. Examples of problem areas that have yet to be investigated from this perspective include: the problem of partner choice in cooperative 'markets' and the evolution of confidence and cues of competence; bargaining problems and the evolution of a sense of fairness; the problems of cooperation between the sexes and the evolution of sexual morality; the problem of alpha males and the evolution of political psychology; and the problems of intergroup relations and the evolution of an in-group/out-group psychology.

*Markets and partner choice*

In most extant models of the evolution of cooperation players are paired at random. Models in which individuals can choose with whom to cooperate are at an early stage.[489] Given that many animals can exercise choice over cooperative partners, we might ask: What criteria do individuals use when selecting cooperative partners? How do individuals cultivate and advertise attributes that meet these criteria? How do individuals ensure that they get 'picked for the team', and are not simply redundant? What is the role of reputation in partner choice? And so on. Most discussions of partner choice have focused on uncertainty about incentives, and hence trustworthiness; but we should also expect individuals to be uncertain about the 'competence' of other partners (especially in cases of mutualism where incentive is not at issue) and hence to the degree of 'confidence' one player should have in another.

---

[489] Noë, et al., 2001.

*Bargaining problems*

Bargaining problems arise in the context of deciding how to divide up the spoils of a cooperative venture among participants. What constitutes a stable solution depends on such factors as: what contribution different players make to the venture (what they 'bring to the table' in terms of skill or effort), how indispensable are those contributions, what alternative opportunities players have, and the costs of bargaining, in terms of the 'decay' rate of the good and the 'impatience' or discount rate of the players.[490] Bargaining problems are more difficult to solve when players are asymmetric, and vary along one or more of these dimensions. (So, the default solution might be 'equal shares', unless one of these other conditions could be established.) Existing theoretical work on bargaining, including the evolution of a 'sense of fairness', has tended to be conducted under the heading of reciprocity.[491] But evolutionary theory suggests that bargaining problems may arise in other contexts, such as mutualism ('fair shares'), the allocation of resources to kin (especially from parents to offspring – 'that's not fair!'),[492] and in the context of conflict resolution ('fair play'). To the extent that different games have different rules, we might expect distinct 'senses of fairness' for each one. An evolutionarily-informed experimental economics that, instead of relying on a tried-and-tested instrument (the 'ultimatum game'), made explicit the adaptive problem and solution it is looking for, might help to tease such adaptations apart.

*Sexual morality*

Numerous games of conflict and cooperation are played out in between males and females in the context of sexual reproduction.[493] Given that

---

[490] For a limpid introduction to bargaining problems, see: Dixit and Skeath, 1999, Ch. 16.

[491] Trivers, 1983. See also Rawls: "The main idea [of fairness] is that when a number of persons engage in a mutually advantageous cooperative venture according to rules, and thus restrict their liberty in ways necessary to yield advantages for all, those that have submitted to these restrictions have a right to a similar acquiescence on the part of those who have benefited from their submission. We are not to gain from the cooperative labors of others without doing our fair share. . . . Each person receives a fair share when all (himself included) do their part." Rawls, 1971, p112.

[492] Trivers, 1974.

[493] Dawkins, 1976, Ch. 9.

sexual reproduction is an evolutionarily ancient problem with significant evolutionary consequences, we might expect human males and females to have specific intuitions regarding relations between the sexes. Humans may have a distinct set of intuitions governing their attitudes to: length of courtship, promiscuity, commitment, parental investment, infidelity, sexual jealousy, mate desertion, polygamy, infanticide, abortion, incest,[494] domestic violence, sexual harassment, and rape. To the extent that these intuitions are designed to solve recurrent problems of cooperation, we might say that they form the basis of human 'sexual morality'.

*Political psychology*

As we have seen, in several primate species, individuals – usually males – compete for status not only through brute force, but also by cultivating allies, and by acting in ways that benefit other members of the group. In chimpanzees, dominant males lead collective hunts and raids into neighbouring territories, distribute some of the spoils of collective hunts, and intervene impartially to settle disputes. It seems as if individuals gain and maintain support (especially from the constituency of females) by being willing and able to shoulder some of the costs of maintaining collective-action schemes. Previous accounts of 'chimpanzee politics' have been content to document these phenomena, and to provide blow-by-blow accounts of the 'court intrigue' that ensues as one coalition displaces another at the top of the dominance hierarchy.[495] It would be interesting to go further and investigate whether this account of 'primate politics' can shed any light on the intuitions (and perhaps philosophical concepts) that humans have about politics and political leadership – concepts such as authority, political obligation, sovereignty, charismatic leadership, impartiality, tyranny, justice, the 'monopoly of the legitimate use of force' and so on. A 'primate political theory' might be illuminating both in providing a causal explanation of the origin of such intuitions (homology), and/or in providing a relatively simple model in which to examine basic political processes (analogy).[496]

---

[494] For recent work on the evolution of incest avoidance, see: Lieberman, et al., 2003.
[495] de Waal, 1982.
[496] Curry, 2003.

*International relations*

As we have seen, whereas Hobbes argued that individuals are inherently selfish, and that cooperation is an artificial invention, Hume argued that humans are inherently social, and that cooperation is a result of various natural social and moral dispositions. I have argued that Hume was correct when it comes to modelling relations between *individuals*. However, it remains an open question as to how best to model relations between *groups*. One view, standard in the field of international relations, is that human societies inhabit a Hobbesian state of nature, and that cooperation between groups is an artificial invention.[497] An alternative, 'Humean', view of international relations would make the following argument. First, human groups lived in close proximity to one another for a sufficient period of time to lead to the evolution of a specific set of intuitions governing relations between groups (or between members of those groups). Second, the opportunities for *competition* with other groups may have led to the familiar psychology of in-group/out-group distinctions, xenophobia, and warlike aggression.[498] But, third, opportunities for *cooperation* with other groups may have led, in addition, to a more benign psychology of tolerance, live-and-let-live, 'internationalism', mutual aid, coalition-formation, conflict resolution, and so on, that are not merely instances of adaptations for inter-individual cooperation operating on a larger scale.[499]

These and other problems can be used to develop a more detailed and comprehensive account of human social and moral psychology.

**A new framework**

In addition to providing a more fine-grained analysis of the problems of cooperation, the evolutionary approach makes a number of broader predictions about moral psychology, which provide the framework in

---

[497] Bull, 1977/1995.
[498] Tooby and Cosmides, 1988; Wrangham and Peterson, 1996.
[499] See Kurzban, et al., 2001 for one example of research into 'coalitional', 'inter-group' or 'corporate' psychology.

which future empirical research in moral psychology will take place. For example, the evolutionary approach suggests that: morality is moderately heritable; morality is the product of a number of distinct neuroanatomical structures that humans share with closely-related species; and morality exhibits systematic sex and age differences. The evolutionary approach also promises to explain: how moral psychology develops differently under different environmental conditions; how moral values vary cross-culturally in predictable ways; and how moral psychology goes wrong. Below I preview what form these research programs will take; and I contrast them with existing approaches to moral psychology, which operate with different, incompatible, pre-Darwinian theories of psychology.

*Morality is moderately heritable*

The proportion of variance in a trait in a given population that is due to differences in genes is referred to as the 'heritability' of that trait. The heritability of height, for example, is around 70%. This means that 70% of the differences in height between people in a given population is due to differences in genes. The heritability of most behavioural and personality traits appears to be around 50%.[500] We should expect adaptations for cooperation, and hence moral values, to exhibit a similar heritability. And this appears to be the case. In a review of the literature, Steven Pinker comments: "Study after study has shown that a willingness to commit antisocial acts, including lying, stealing, starting fights, and destroying property, is partly heritable . . . ".[501]

However, existing heritability studies have tended to look for variance not in the specific adaptations that evolutionary theory suggests are components of moral psychology, but in the larger more loosely-defined, a-biological, social psychology categories such as 'pro-' and 'anti-

---

[500] Harris, 1998; Plomin, 1990.

[501] Pinker, 2002, p50. Studies of 'lying and stealing' are, at the same time, studies of 'telling the truth and respecting others' property'. It would be interesting to know to what extent behavioural genetics explains the other end of the curve; that is, why some people become priests, traffic wardens, police officers, magistrates, judges, triage nurses, sports referees, and so on. In what ways do their moral psychologies differ from the average?

sociality' or 'individual- versus community-orientation'. The alternative, Darwinian approach to moral psychology would investigate the heritability of the performance of actual adaptive mechanisms, such as: maternal and paternal sentiment; offers made in ultimatum games; skill at solving coordination problems using various orders of 'theory of mind' (that is, problems requiring first-, second-, third-, nth-order 'theory of mind'); discount rate and tendency to cooperate in prisoner's dilemmas; accuracy and speed of detecting cheats in a Wason Selection Task; 'punitivity'; willingness to forgive; display of costly-signals of fitness; tendency to detect and respond to facial cues of distress or submission; and so on.

In addition to testing the performance on these 'benchmarks' tasks, we could look for individual differences in the relative strength or weighting of one moral sentiment as compared to another, or of a moral sentiment compared to a nonmoral sentiment through the construction of suitably excruciating moral dilemmas. (It should be possible, by altering the variables of particular scenarios, to find the point at which individuals were indifferent between conflicting moral intuitions, and between moral and nonmoral intuitions, and thereby assess their relative strength. Such dilemmas would play on conflicting loyalties to family, friends, teams, associates and strangers.)[502] Such tests would tap into the various adaptations that comprise human moral psychology and thereby reveal the heritability of these underlying adaptations. We might find that some areas of moral value – such as the display of virtue – exhibit greater heritable variation than others.[503]

Assuming that there are differences in the 'genes for' morality, we might proceed to ask "Where are these genes?" and "Why do they differ?". We are a long way off from being able to match genotype with phenotype, especially for traits as complicated as morality, but in theory it should be possible to identify the genes involved. If we were able to do this, we

---

[502] The boardgame *Scruples* provides one example: You are making the school run with your own child and your best friend's child, but your new car has only one seatbelt. Which child do you put in the seat with the belt?
[503] Miller, 2004.

could then look for homologous genes in other species, and thereby build up a family tree of morality – a true genealogy of morals. We might, for example, be able to tell when and in which species different aspects of moral psychology first evolved, what they evolved from, and with which extant species humans share them. (In turn, evidence from genetic studies might converge with evidence from comparative neuroanatomy to give us a complete picture of the evolution of moral sentiments. See below.)

Turning to the question of *why* genes differ, such differences could merely reflect a normal distribution, produced by mutation and noise. Alternatively, differences in genes could reflect the selection of distinct adaptive 'morphs' in the population. For example, Linda Mealey has argued that some instances of sociopathy are the product of a genetic polymorphism – "frequency-dependent, genetically-based individual differences in employment of life strategies". Mealey argues that the game-theoretic prediction that a small proportion of 'natural born hawks' can survive by preying on the majority 'doves', explains the high-risk/high-reward, anti-social behaviour that is characteristic of "a small, cross-culturally similar, and unchanging baseline frequency of sociopaths".[504]

*The neuroanatomical location of moral sentiments*

If morality is the product of a number of distinct adaptations, then we might expect different parts of the brain to 'light up' when attempting to solve different collective-action problems.[505] Evolutionary theory itself makes no predictions about the specific neuroanatomical location of particular adaptations. But, as Hume noted, comparative and phylogenetic evidence can.[506] Studying the brains of closely-related species might generate predictions about where to find homologous adaptations in human brains. For example, we could place other primates in typical social and cooperative situations – we could play distress calls

---

[504] Mealey, 1997, p163.
[505] In general, see: Blakemore, et al., 2004; Duchaine, et al., 2001. More specifically, see: Rilling, et al., 2002; Sanfey, et al., 2003; Stone, et al., 2002.
[506] Clutton-Brock and Harvey, 1984; Hume, 1739/1985, pp375-6.

of an infant to a rhesus mother, get orangutans to play coordination games, steal a chimp's banana, or disrespect an alpha-male gorilla – and see which parts of their brain are activated. We could then place humans in equivalent situations, and look to see whether the same areas light up.[507] The discovery of functional homologies between nonhuman adaptations for cooperation and human moral sentiments would obviously provide a very firm evidentiary foundation for the Humean-Darwinian account of morality.

James Rilling has recently begun to use this comparative method to establish the neuroanatomical location of mechanisms governing mate guarding, or "jealousy".[508] Rilling allowed a dominant male rhesus to mate with a female. Rilling then put the dominant male behind a one-way mirror, from which the male could watch as a subordinate male mated with the same female. Not surprisingly, the dominant male became very agitated. Rilling was then able to conduct a PET-scan on the dominant male to discover the location of his putative 'mate guarding system'. Rilling now knows where to look for a homologous system in humans. The prediction would be that if you subjected a human male to the same situation – albeit by means of an imaginary scenario – then the same parts of his brain would become activated. If they did, then this would constitute *prima facie* evidence that sexual jealousy was an ancient instinct, and not merely an artificial social construction.[509]

This comparative approach differs from most extant attempts to investigate moral psychology using brain-scanning technology. Instead of using predictions derived from evolutionary or phylogenetic considerations, most current brain-scan research tends to test whether moral decision-making is a product of 'emotion' or 'reason'.[510] As such,

---

[507] We might, at a later date, attempt to establish homologous neuroanatomical structures across wider groups than primates – perhaps including other social mammals (dolphins, wolves), or social vertebrates in general (ravens, cleaner fish).
[508] Rilling and Kilts, 2003.
[509] See also: Buss, 2000a.
[510] Greene and Haidt, 2002.

modern technology is being used to test the ancient, mediaeval, and folk theories of psychology that are *au courant* in moral philosophy.[511]

Moreover, researchers attempt to test the theory not by giving individuals distinct 'ecologically valid' problems and scenarios, but by giving moral dilemmas that arise in evolutionarily-novel environments.[512] As discussed in the previous chapter, a moral dilemma may not be the best way to tease out the components of a moral psychology, because such dilemmas are likely to arise at exactly the point at which individuals are indifferent between outcomes, and where their intuitions are most unstable, conflicted and/or weak. This is not to say that moral dilemmas are of no use in the investigation of moral psychology. Like optical illusions, moral dilemmas can be used to reveal the gaps between the adaptations. It's just that in order to make sense of moral dilemmas – or optical illusions – we must locate them in a broader theory of the underlying mechanisms that give rise to them. This alternative approach to moral dilemmas would begin by establishing distinct adaptive systems, and would then construct scenarios in which these systems come into conflict or interfere with one another.

*Sex differences in morality*

Generally speaking, evolutionary theory leads us to expect that organisms that have faced different adaptive problems will exhibit different adaptive solutions. This applies to individuals within a particular species as much as to members of different species. So, for example, evolutionary theory leads us to expect that, to the extent that men and women have faced different adaptive problems, they will

---

[511] The problem of sticking with theories that are long past their 'sell by date' was elegantly diagnosed by one of the early pioneers of cybernetics: "Some of today's outstanding problems about the brain and behaviour come to us from mediaeval and earlier times, when the basic assumptions were very different and often, by today's standards, ludicrously false. Some of these problems are probably wrongly put, and are on a par with the problem, classic in mediaeval medicine: what are the relations between the four elements and the four humours? This problem, be it noticed, was never solved – what happened was that when chemists and pathologists got to know more about the body they realised that they must ignore it." What is required instead is "a clear demonstration that they should not be asked". Ashby, 1956, p191.
[512] The results of these studies may also be confounded by invoking intuitions about tool-use, as in 'the trolley problem'.

exhibit different adaptations. And, to the extent that men and women have faced different problems of cooperation over evolutionary time, they will exhibit different adaptations for cooperation, and hence different moral values.

For example, we might expect the expression of kin altruism to loom larger in the psychologies of women, for whom social relations have involved a greater degree of interaction with close kin (and others with whom their interests are closely tied). We might also expect the expression of reciprocal altruism and conflict resolution to loom larger in the psychologies of men, for whom social relations have involved a greater degree of interaction with non-kin and those with whom their interests are in conflict. We might also expect the display and selection of particular 'virtues' to reflect the cross-cultural sex differences in behaviour and mate-choice criteria.[513] We have already seen that men seem more willing to perform risky heroic acts to save the lives of strangers; and that in women, chastity and fidelity are more highly valued.[514]

This approach to sex differences differs from existing approaches, which tend to assume that boys and girls start with identical minds, and that any resulting differences between them are the arbitrary products of differential socialisation, or of some kind of Freudian psychodrama.[515]

*Age differences in morality*

Evolutionary theory also leads us to expect that individuals will exhibit systematic age differences in moral values. Different stages of an individual's life – infancy, childhood, adolescence, sexual maturity, parenting, grandparenting – present different adaptive challenges, and we might expect different psychological mechanisms to come on- and off-

---

[513] Cross-cultural sex differences in mate-choice criteria: Buss, 1994. See also: Low, 1989.
[514] One prediction might be that men and women hold men more accountable than women for failing to perform heroic acts. That is, attitudes to the moral status of the omission versus the commission of particular moral acts may vary according to the sex of the actor.
[515] Chodorow, 1978.

line at different times to deal with them. And this includes adaptations for various kinds of cooperation.

For example, we might expect individuals to have different adaptations: for signalling viability and need to mothers when infants; for competing with siblings for parental investment; for fine-tuning moral competences through play during childhood; for identifying with a peer group; for negotiating relations with the opposite sex following sexual maturity; for climbing status hierarchies through the display of sexually-selected fitness indicators; for adopting local conventions; and so on.[516] These adaptations might come on-line (and go off-line) according to a pre-specified developmental sequence – similar to the emergence of secondary sexual characteristics. To quote Tooby and Cosmides: "just as teeth and breasts are absent at birth and develop later in an individual's life history, perceptual organization, domain-specific reasoning mechanisms, the language acquisition device, motivational organization and many other intricate psychological adaptations mature and are elaborated in age-specific fashions that are not simply the product of 'experience'."[517] Alternatively, the development of certain adaptations might be triggered if individuals enter into long-term sexual relationships, if they have status and property to defend, if they have children, or grandchildren of their own, or if they ascend to the rank of alpha status and are called upon to resolve disputes.

This evolutionary approach to developmental moral psychology differs from the dominant approach, which tends to assume that moral development in children is largely a function of general growth in the capacity for sympathy, coupled with a general increase in intelligence that enables the child to learn, comprehend, and eventually criticise, increasingly-sophisticated moral rules.[518]

---

[516] "To understand why the form of moral judgement undergoes systematic changes with development, evolutionary theory directs us to attend to its functions; this more than anything, is missing in cognitive-developmental models". Krebs, 1998, p364.
[517] Tooby and Cosmides, 1992, p81.
[518] Kohlberg, 1981; Piaget, 1997; Rawls, 1971, Ch. 13.

One prediction that combines sex and age differences relates to life-history strategy and attitudes to abortion. Life-history strategy predicts that females value offspring as a function of the females' age. That is, younger females, who have the opportunity to reproduce again in the future, discount offspring at a higher rate than do older females, for whom each offspring may be their last. Could these attitudes towards the value of the lives of infants be reflected in, and largely constitute, attitudes towards abortion? One evolutionary prediction is that women's approval of abortion will decline as a function of age; whereas men's approval will not (or not by as much).

*Morality develops differently under different environmental conditions*

If about half the variance in moral personality traits can be traced to difference in genes, then this leaves about half to be explained by non-genetic factors. A Darwinian account of psychology can help here by generating hypotheses about which aspects of the environment particular adaptations for cooperation are likely to be sensitive to (and how they are likely to respond).

As we have seen, adaptations for cooperation tend to be 'conditional'; hence specifying the design of such an adaptation necessarily involves specifying what states the adaptation will adopt under different environmental conditions. As discussed above, this conditionality might be manifest as 'triggering'. In addition, there may be many 'critical periods' during (childhood) development in which distinct moral competences are fine-tuned, similar to the critical periods in the development of language.

For example, evolutionary theory leads us to expect that adaptations for kin altruism might be calibrated by the viability of one's offspring, or to the size and distribution of one's family group.[519] Adaptations for mutualism may be calibrated by one's skills and abilities – what one has

---

[519] Lieberman, et al., 2003.

to offer.[520] Adaptations for reciprocity may be sensitive to average mobility and discount rates, or life-expectancy.[521] Adaptations for conflict resolution may be calibrated by feedback on entering contests during adolescence, by the gradient of local status hierarchies, or by the availability of certain types of resources.[522] And so on.

In this way, the Darwinian account of moral sentiments suggests the development of moral psychology is influenced by 'the environment' in many more ways that are currently envisaged by standard social psychological theories that restrict their focus to rote-learning from parents or imitation of role models.[523]

Incidentally, it would be strange, from an evolutionary perspective, if natural selection had designed children to fine-tune their cooperative strategies entirely, or largely, in response to the behaviour of their parents. How one is treated by one's parents as a child – let alone as an infant – seems to hold relatively little predictive information about how one will be treated by other people, several decades later, in one's adult social environment. More reliable information about one's future social environment may come from assessments of one's own abilities, strengths and weaknesses, and how these abilities compare to those with whom one will be competing, cooperating, meeting and mating for the rest of one's life – one's peer group – as well as background assessments of local ecological or economic conditions. Recent research in behavioural genetics has indeed undermined the standard 'nurture assumption' that one's parents and one's home are the most important environmental influences on a child's developing personality, including their developing moral character. Parents influence their children primarily through the genes that they give them. Parenting style and home environment seem

---

[520] Tooby and Cosmides, 1996.
[521] Wilson and Daly, 1997.
[522] Economists and social psychologists have found that, for men, adolescence height is a better predictor than adult height of adult self-confidence, dominance, 'social poise', and income: Jones, 1957; Persico, et al., 2001. This suggests that there is a 'critical period' during adolescence in which males calibrate their 'conflict resolution' strategies.
[523] Pinker, 2002.

to have negligible effects on adult personality.[524] But which other factors may be important has yet to be fully investigated.

*Abnormal psychology and pathology*

The emerging picture of moral psychology is one of a vast and intricate decision-tree, with hundreds and thousands of branches, decision nodes, variables, switches, sensitivities, subroutines, feedback loops and so on. A full account of this vast circuitry is a necessary pre-requisite for understanding the countless ways in which it can go wrong. There will be as many different types of moral psychopathy as there are mechanisms that can go wrong multiplied by the number of different ways that they can go wrong. (Genetic mutation, developmental extremes, poisoning, hormonal imbalances, a bump on the head or countless other interventions could all be responsible for knocking out or getting wires crossed in various parts of this intricate structure, which in turn, go wrong in countless ways.) In this way, the evolutionary account of moral psychology creates the possibility of a vast list of distinct pathological conditions, for which current terms such as 'sociopath' or 'psychopath' may merely be imprecise catch-all labels.

Research in the area of evolutionary moral psychopathology has only just begun.[525] But already, work by James Blair seems to have identified one of the mechanisms that can fail. We saw earlier that many animals forestall or settle conflicts by attending to costly signals of fitness and submission cues. James Blair has argued that problems with this 'violence inhibition mechanism' lead to one of the typical patterns of psychopathy.[526]

Again, it is an open question as to whether some or all of these abnormal psychologies are the product of pathology, or whether they represent polymorphisms.

---

[524] Harris, 1998.
[525] See, for example: Murphy and Stich, 2000; Nesse, 2001b.
[526] Blair, 1997.

*Moral values exhibit predictable cross-cultural variation*

So far I have been discussing the Darwinian approach to individual differences in moral value. Aggregating these theories of individual difference constitutes the first step in explaining cross-cultural differences in moral value.

For example, if individuals of different ages and sexes have different adaptations for cooperation, and hence different moral values, then societies composed of different proportions of such individuals can be expected to exhibit different moral systems. Consider, for example, demographic factors such as changes in the age-structure and operational sex ratio of a population. What would happen if a society experienced a war that decimates males, had a baby boom, or had a disproportionately-old population. Perhaps a baby boom produces an outbreak of maternal sentiment, followed twenty years later by intense costly-signalling as a swollen age-cohort enters a fiercely competitive mating market, followed ten years later by a suffocating preponderance of risk-averse conservatism… and so on.[527]

In addition, given that individuals rely on several kinds of adaptations for cooperation – for kin altruism, mutualism, reciprocity, and conflict resolution – we might be able to identify, first in theory, and then empirically, conditions under which different modes of social interaction become predominant, and categorise societies accordingly. Factors that might alter the balance between different modes of social interaction might include: dispersal of families, mobility of the population, type of economy, efficiency of trade, and degree of inequality. For example, we might predict that social order in a society in which family relations were disrupted – by migration, urbanisation, or industrialisation for example – would come to be characterised more by reciprocal ties than by kin, and vice versa.[528] As a society becomes more unequal, we might expect its

---

[527] See Buss, for a discussion of the effects of a baby boom on the mating market Buss, 1994, p202.
[528] In the West, was the shift from 'status to contract' in the modern era the point at which the rewards to reciprocal relations outweighed the rewards to kin relations? Was this shift an effect of

social institutions to become increasingly those designed to manage conflicts and competition; we might expect the display of sexually-selected virtues to increase when male-male competition is particularly intense, or, for women, when hypergamy pays particularly well.

We should also expect 'conditional' adaptations operating under different conditions to produce cross-cultural differences in attitudes and behaviour. We have seen, for example, how psychological mechanisms governing paternal investment are sensitive to certainty of paternity, which in turn can lead to society-wide differences in family structure. In turn, certainty of paternity depends on such factors as time spent away from the home base, sex ratios, degree of male-male inequality, ecological distribution of resources. We have also seen how cognitive adaptations for social exchange are likely to trade luck-based goods but not effort-based goods.[529] Mechanisms governing incest aversion are triggered by the presence of siblings in the household; this predicts that China's one-child policy will have produced a nation for whom the thought of sibling incest is not repugnant… And so on.

What moral philosophy needs is a cross-cultural survey of moral psychology of the kind conducted by David Buss into sexual psychology. Buss took evolutionary predictions about sexual psychology and mate preferences and tested them on 10,000 individuals in 37 cultures.[530] A survey of moral attitudes could provide a provisional account of 'universal moral grammar'; that is, it would help in discovering cross-cultural universals, as well as testing theories of how moral systems vary from place to place.

---

technology making it possible for an individuals to acquire more resources in their lifetimes than they could inherit from their families?

[529] Other factors influencing the preponderance of social exchange may include: "degree of kinship; status or rank; whether a relationship is long-term or short-term; whether one is in daily contact . . . or only rare contact; whether storage [of the goods] is possible; whether the group is sedentary enough for inequalities in wealth to accumulate; whether gaining a resource requires close, interdependent cooperation; whether people are trading different resources or dividing the same resource; whether an external, consensual definition of 'equal portion' is feasible; whether an individual can control access to a resource, and thereby 'own' it; and so on." Cosmides and Tooby, 1992, p285.

[530] Buss, 1989.

As yet, evolutionarily-informed tests of moral reasoning have not progressed further than testing *whether or not* people are kin altruists, as opposed *to what extent* they are kin altruists as compared to reciprocators. O'Neill and Petronovich, for example, tested American and Taiwanese children on a version of the trolley problem that required them to choose between saving family and friends, saving a stranger, or saving a nonhuman animal. Not surprisingly, for all groups, "it was preferable to save the humans at the expense of the nonhumans, and kin and friends at the expense of strangers".[531]

The evolutionary approach to cross-cultural diversity in moral values provides a broader framework for – and potentially incorporates – existing accounts, which tend to focus exclusively on historical descriptions of local conventional solutions to coordination problems (or assume that all differences or changes are the product thereof).

In summary, by providing more specific predictions about the contents of moral psychology, and by providing a broader integrated framework in which research can take place, the evolutionary approach promises to bring moral psychology into the 21st century, and vastly increase our understanding of how human moral sentiments work.

**Re-evaluating morality**

Once we have the full circuit diagram of moral sentiments, we will be in a position to: a) examine the extent to which human moral sentiments (continue to) constitute optimal solutions to collective actions problems; and, to the extent that they fall short, b) to design interventions – education, incentives, institutions, policies – to make up the difference.

Adaptationist theorising has a role to play here also. Remember that adaptationists begin by deriving optimal or evolutionary stable solutions to adaptive problems, and then proceed to use these 'ideal solutions' to

---

[531] O'Neill and Petrinovich, 1998, p363. The article goes on to review a number of similar findings by other researchers.

guide the search for actual adaptations. But as well as being of heuristic value, these 'ideal solutions' can provide a benchmark against which to evaluate the performance of particular adaptations.

Evolutionists expect adaptations to fall short of any theoretical optimal solution for a variety of reasons. Natural selection cannot wipe the slate clean and redesign adaptations from scratch; it can work only with the available genetic variation; there are always trade-offs to be made between the demands of different adaptations; and so on.[532] But perhaps most relevant to assessing human moral sentiments is the problem of 'time lag': some adaptations designed to solve collective-action problems in the Pleistocene may not provide the best guides for solving collective-action problems in the modern world. As the moral philosopher Henry Sidgwick observed: "however completely adapted the moral instincts of community may be at some particular time to its conditions of existence, any rapid change of circumstances would tend to derange the adaptation, from survival of instincts formerly useful, which through this change become useless or pernicious."[533]

Biotechnology provides one possible example of the way in which ancient moral psychology fails to cope with the modern world. Steven Pinker has argued that our attributions of agency – sometimes articulated as a ghost in the machine – may work well for negotiating social relationships most of the time. But this intuition seems incapable of resolving the problems generated by modern medicine and reproductive biotechnologies, such as abortion, euthanasia, cloning, animal rights, environmental degradation and so on – or rather, it makes decisions about whether to permit certain activities turn on such arcane issues as "At what point does the soul enter the body?" A proper understanding of the nature of moral intuitions may lead us to reframe the problem, or more radically, question whether the issue really is a moral one and not merely an example of a 'false positive'.

---

[532] Dawkins, 1982, Ch. 3.
[533] Sidgwick, 1962, p465.

We might conclude that 'bioethics' consists of the range of problems that our moral intuitions are incapable of solving.[534]

International relations may provide another example. Revenge and honour may have been necessary to secure order in small bands of hunter-gatherers armed only with sticks and stones; but is it still a good rule to use in populations of billions armed with ICBMs, anthrax and dirty bombs? The point is not to prejudge the issue. Natural selection has had several million years 'research and design' experience of building systems of ethical rules, and moral philosophers have a lot to learn from it. Perhaps the problem of international order in the modern world is sufficiently similar to the problem of order in prehistory for revenge to continue to be a good strategy, at least as good as the proposed alternatives. The point is that we cannot simply assume that this is the case because it feels right.

Once we have identified the ways in which our moral sentiments let us down, and assuming that we nevertheless want to promote cooperation, we can proceed with proposals for novel interventions to make up the difference. Peter Singer has suggested that, in the distant future, it may be possible to genetically modify morality, "so that instead of living in societies constrained by our evolutionary origins, we can build the kind of society we judge best".[535] Less drastically, we might engage in the more familiar practice of attempting to design incentives and institutions that artificially extend and sustain evolutionarily-novel levels of human cooperation.[536]

A curious but inevitable result of this re-examination of our moral sentiments is that, as Michael Ruse notes, once we recognise that ancient moral rules may not be "the ideal strategy for long-term survival and reproduction in an era of high technology" it could become the case that

---

[534] Pinker, 2002, pp223-9.
[535] Singer, 1999.
[536] I hesitate to give 'a' footnote to the concept of policy-making. But the following represent the best recent general accounts of using game theory to design policy in the light of evolutionarily-informed accounts of human nature: Binmore, 1994a; Binmore, 1994b; Popper, 1945; Posner, 1981; Rawls, 1971. More specifically, see: Crawford and Salmon, 2004.

we "owe it to our children to be immoral". Or, in other words, "a better understanding of biology might incline us to go against morality".[537] What's more, if policy-makers suggest novel ways of tackling social problems that depart from our evolved intuitions, the proposed solutions might strike us, intuitively, as non-moral or, in some cases, immoral.[538] That, after all, is exactly what we should expect. There is an analogy here with human dietary preferences. Our evolved tastes for sugars, salts and fats were good guides in ancestral conditions; but they no longer work as intended in a world where these foods are abundant. Given our desire to be attractive, and to live long healthy lives, we try to overcome these preferences. Nutritionists can prescribe a healthier diet than can our beleaguered taste buds, but we shouldn't expect the new diet to taste as good. It is a moot point whether you would want to call this new set of rules 'moral', any more than the new diet is 'tasty'.

Incidentally, the project outlined in this chapter – of first arriving at an accurate description of our evolved moral sentiments, and then assessing them according to 'optimal' cooperative strategies – is in many ways similar to John Rawls' view of moral philosophy. First, Rawls accepts the view that human moral sentiments are the product of evolution by natural selection. Rawls writes: "The theory of evolution would suggest that [human moral psychology] is the outcome of natural selection; the capacity for a sense of justice and the moral feelings is an adaptation of mankind to its place in nature".[539] Referring to Triver's paper on reciprocal altruism, Rawls adds: "We can also see how the system of moral feelings might evolve as inclinations supporting the natural duties and as stabilizing mechanisms for just schemes."[540] Second, Rawls argues that the first job of the moral philosopher is to accurately describe human moral faculties. Rawls writes: "[O]ne may think of moral philosophy at first . . . as the attempt to describe our moral capacity; or, in the present case, one may regard a theory of justice as describing our sense of

---

[537] Ruse, 1995, p256, p283.
[538] "[A] scientific ethics might be equally counter-intuitive." Collier and Stingl, 1993, p59,
[539] Rawls, 1971, p503.
[540] Rawls, 1971, p504.

justice."[541] Rawls regards the problem of describing our sense of justice as equivalent to the problem, in linguistics, of describing "the sense of grammaticalness that we have for the sentences of our native language".[542] Third, Rawls sets up a model of ideal cooperation – through the apparatus of 'the original position' – that can be used to evaluate our moral intuitions and political institutions. Rawls suggests that we should be willing to revise our intuitions and institutions in order to bring them into line with the rules generated by the model (a process he called 'reflective equilibrium').[543]

**Conclusion: pastures new**

David Hume's contribution to moral philosophy consists not of a particular psychological theory, but in his 'bold conjecture' that morality is a part of the natural world, and can be studied using the tools of natural science. With this move, moral philosophy ceased to consist of a profusion of mysteries, and became instead a series of problems.

This thesis has presented some of the progress that has been made on these problems since Hume's day. We have seen how recent advances in evolutionary biology, game theory, animal behaviour, decision-making, and neuro-imaging have shed new light on the mechanisms underlying human moral reasoning and moral behaviour. Evidence from all these sources suggests that human psychology is in part composed of a suite of "adaptations for cooperation" – including adaptations for caring for kin, for coordinating behaviour with others, for making and enforcing

---

[541] Rawls, 1971, p46.

[542] Rawls, 1971, p47.

[543] If morality consists of a suite of decision rules, then it should be possible to program them into a computer or to include them among the control-systems of robots. This has already happened in a relatively simple way in the context of artificial-life and evolutionary game theory, where different strategies for playing nonzero-sum games are let loose to compete and cooperate with one another. But in time we should be able to program a computer with a more comprehensive repertoire of human moral sentiments. Danielson, 1992; May, et al., 1996. Such a computer would provide a novel way for us to test our theories of the content and organisation of moral psychology, in much the same way that artificial intelligence in general has been a proving ground for philosophy of mind: Harvey, 2000. Such a computer program could form the basis of a 'Moral Turing Test', or a cybernetic version of Rawl's reflective equilibrium. We would know that we had the theories correct if we couldn't distinguish the computer's answers to our moral questions from the answers given by a human, or if the output of the program meshed with our moral intuitions.

agreements, for detecting and punishing cheats, and for settling disputes peacefully. This thesis has argued that these adaptations for cooperation constitute what have been referred to as "human moral sentiments" – sentiments such as maternal love, sympathy, shame, trust, revenge, guilt, and magnanimity – and showed that this evolutionary approach to human morality makes sense of, and has the potential to incorporate, a wide variety of traditional positions in moral philosophy. In this chapter, I have looked at some of the problems that remain, at some of the novel problems that this approach to morality generates, and considered how scientists and philosophers might go about solving them.

This exciting approach to morality promises to be a rewarding non-zero sum game for all those willing and able to participate.
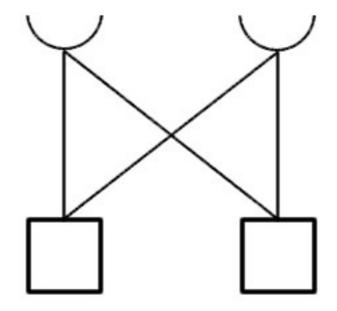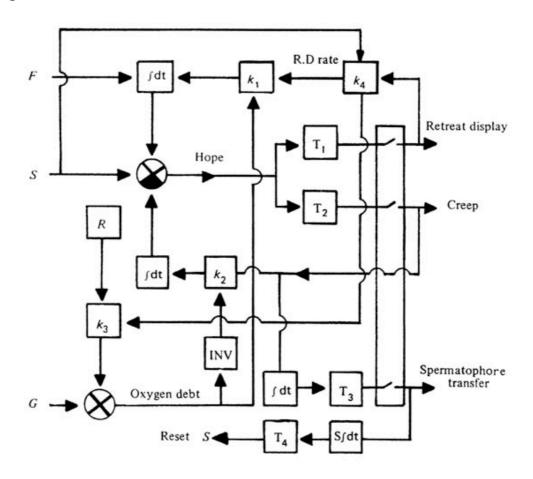
## Appendix 1: How to build an animal

**Figure 1: A simple animal**



The 'cups' at the top of the diagram are photo-detectors. They are linked to the square 'motors' at the bottom. Assume that blue light falling on the detectors activates the motor(s) indicated by the crossed lines; and that red light activates the motor(s) indicated by the straight lines. Assume also that if both circuits are activated, the red circuit takes priority over the blue. You now have a mobile system that will turn and move towards blue objects, and that will turn away from red objects. For example, a blue light source in the top left hand corner of the page will activate the right-hand motor, causing the system to turn towards the blue light. At the point at which both detectors are activated equally the system moves directly towards the source of the blue light. (Adapted from Brooks, 1999.)

**Figure 2: A control model for the courtship of the smooth newt**

(Holland, O. and D. McFarland (2001). <u>Artificial Ethology</u>. Oxford, OUP, p10.)

# Appendix 2: Nonzero-sum games

## Table 1: The principal categories of nonzero-sum games

**Coordination**

|   | A | B |
|---|---|---|
| A | *1, 1* | 0, 0 |
| B | 0, 0 | *1, 1* |

**Prisoner's dilemma**

|   | A | B |
|---|---|---|
| A | 2, 2 | 0, 3 |
| B | 3, 0 | *1, 1* |

**Chicken**

|   | A | B |
|---|---|---|
| A | 2, 2 | *1, 3* |
| B | *3, 1* | 0, 0 |

## Table 2: Coordination games

**Simple coordination**

|   | A | B |
|---|---|---|
| A | *1, 1* | 0, 0 |
| B | 0, 0 | *1, 1* |

**Division of labour**

|   | A | B |
|---|---|---|
| A | 0, 0 | *1, 1* |
| B | *1, 1* | 0, 0 |

## Table 3: Prisoner's dilemmas[544]

**Prisoner's dilemma**

|   | C | D |
|---|---|---|
| C | R, R | S, T |
| D | T, S | *P, P* |

**Prisoner's dilemma**

|   | C | D |
|---|---|---|
| C | 1, 1 | -1, 2 |
| D | 2, -1 | *0, 0* |

**Repeated**

|   | C | D |
|---|---|---|
| C | R+*wd*R, R+*wd*R | S+*wd*P, T+*wd*P |
| D | T+*wd*P, S+*wd*P | P+*wd*P, P+*wd*P |

**Assurance**

|   | C | D |
|---|---|---|
| C | 2, 2 | -1, 1 |
| D | 1, -1 | *0, 0* |

## Table 4: Hawk-Dove/Chicken

**Hawk-Dove**

|   | D | H |
|---|---|---|
| D | V/2, V/2 | *0, V* |
| H | *V, 0* | (V-c)/2, (V-c)/2 |

**Hawk-Dove (V<c)**

|   | D | H |
|---|---|---|
| D | 2, 2 | *0, 4* |
| H | *4, 0* | -1, -1 |

---

[544] A game is a prisoner's dilemma if T>R>P>S, and R>(S+T)/2. A hawk-dove game in which V>c is also a prisoner's dilemma Dixit and Skeath, 1999, p341.

# Appendix 3: An overview of research on human adaptations for cooperation

| | Genetic | Hormonal | Develop-mental | Neuro-logical | Behavioural | Anthropo-logical |
|---|---|---|---|---|---|---|
| **Kinship** | | Hrdy, 1999; Panksepp, 2000 | | Platek, 2003 | Platek, et al., 2002 | Brown, 1991; Daly and Wilson, 1988 |
| **Coordination to mutual advantage** | | Grimes, 2003 | Baron-Cohen, 1995; Hoffman, 2001; Pinker, 1994 | Baron-Cohen, 1995; Decety and Chaminade, 2002; Pinker, 1994 | Baron-Cohen, 1995 | Brown, 1991 |
| **Reciprocity** | | | | Rilling, et al., 2002; Stone, et al., 1997; Stone, et al., 2002 | Cosmides and Tooby, 1981; Frank, 1988; Price, et al., 2002 | Brown, 1991 |
| **Conflict resolution (CA)** | | Mazur and Booth, 1998; Miller, 2000a | | Blair, 1997 | Miller, 2000a | Brown, 1991; Miller, 2000a |
| **Conflict resolution (UA)** | | | | | Kahneman, et al., 1982 | Brown, 1991; Wilson and Daly, 1992 |
| **Other (Sex, disgust)** | | | Lieberman, et al., 2003 | | | |

# Bibliography

Alcock, J. (1998). *Animal Behavior: An Evolutionary Approach*. Sunderland, Massachusetts, Sinauer Associates, Inc.

Alexander, G. M., and M. Hines (2002). "Sex Differences in Response to Children's Toys in Nonhuman Primates (*Cercopithecus Aethiops Sabaeus*)." *Evolution and Human Behavior* **23**(6): 467-9.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York, Aldine de Gruyter.

Alexander, R. D. (1979). *Darwinism and Human Affairs*. Washington, University of Washington Press.

Allee, W. C. (1931). "Co-Operation among Animals." *American Journal of Sociology* **37**(3): 386-98.

Anscombe, G. (1981). Modern Moral Philosophy. *Ethics, Religion and Politics*. Minneapolis, University of Minnesota**:** 26-42.

Aquinas (1988). *On Politics and Ethics*. New York, London, WW Norton & Company.

Ardrey, R. (1966). *The Territorial Imperative: A Personal Inquiry into the Animal Origins of Property and Nations*.

Aristotle (1962). *Nichomachean Ethics*. Engelwood Cliffs, NJ, Prentice-Hall.

Aristotle (1992). *The Politics*, Penguin Books.

Arnhart, L. (1998). *Darwinian Natural Right: The Biological Ethics of Human Nature*, SUNY Press.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. London, Chapman & Hall.

Augustine (1998). *The City of God against the Pagans*, Cambridge University Press.

Aureli, F., and F. B. M. de Waal, Eds. (2000). *Natural Conflict Resolution*. Berkeley, University of California Press.

Avital, E., and E. Jablonka (2000). *Animal Traditions : Behavioural Inheritance in Evolution*. Cambridge, Cambridge University Press.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York, Basic Books.

Bacharach, M. O. L., and D. Gambetta (2001). Trust in Signs. *Trust in Society*. K Cook. New York, Russell Sage Foundation.

Baggini, J., and P. S. Fosl (2003). *The Philosopher's Toolkit: A Compendium of Philosophical Concepts and Methods*. Oxford, Blackwell Publishing.

Baier, A. (1995). *Moral Prejudices: Essays on Ethics*, Harvard University Press.

Baker, R. (1996). *Sperm Wars: The Science of Sex*. New York, Basic Books.

Barkow, J. H., L. Cosmides, and J. Tooby, Eds. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York, Oxford University Press.

Baron-Cohen, S. (2003). *The Essential Difference*.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, Massachusetts, MIT Press.

Baron-Cohen, S. (1988). "Without a Theory of Mind One Cannot Participate in a Conversation." *Cognition* **29**(1): 83-4.

Beaglehole, E. (1931). *Property: A Study in Social Psychology*. London, Allen and Unwin.

Beckstrom, J. H. (1993). *Darwinism Applied: Evolutionary Paths to Social Goals*. Westport, Connecticut, Praeger Publishers.

Berger, P. L., and T. Luckmann (1966). *The Social Construction of Reality*. London, Allen Lane The Penguin Press.

Berlin, I. (1969). *Four Essays on Liberty*. London and New York, Oxford University Press.

Berlin, I. (1998). "Isaiah Berlin on Pluralism." *New York Review of Books* **XLV**(8).

Betzig, L. (1993). "Where Are the Bastard's Daddies?" *Behavioural and Brain Sciences* **16**(2): 284-5.

Binmore, K. (1994a). *Game Theory and the Social Contract, Vol 1: Playing Fair?* Cambridge, Massachusetts, MIT Press.

Binmore, K. (1994b). *Game Theory and the Social Contract, Vol 2: Just Playing*. Cambridge, Massachusetts, MIT Press.

Blackburn, S. (1994). *Oxford Dictionary of Philosophy*. Oxford, Oxford University Press.

Blackburn, S. (1998). *Ruling Passions*. Oxford, Clarendon Press.

Blackburn, S. "Why We Pull Together: Hume's Understanding of Habit and Sentiment in Human Life." *Times Literary Supplement*, February 25th 2000.

Blair, R. J. (1997). A Cognitive Developmental Approach to Morality: Investigating the Psychopath. *The Maladapted Mind*. Simon Baron-Cohen, Psychology Press**:** 85-114.

Blakemore, S., J. Winston, and U. Frith (2004). "Social Cognitive Neuroscience: Where Are We Heading?" *TRENDS in Cognitive Science* **8**(5): 216-22.

Blurton Jones, N. G. (1984). "A Selfish Origin for Human Food Sharing: Tolerated Theft." *Ethology and Sociobiology* **5**: 1-3.

Boesch, C. (1994). "Cooperative Hunting in Wild Chimpanzees." *Animal Behaviour* **48**: 1135-48.

Boinski, S. (2000). Social Manipulation and Group Movement. *On the Move: How and Why Animals Travel in Groups*. Sue Boinski and P A Garber**:** 421-69.

Boinski, S., and P. A. Garber, Eds. (2000). *On the Move: How and Why Animals Travel in Groups*.

Bonnie, K. E., and F. B. M. de Waal (2004). Primate Social Reciprocity and the Origin of Gratitude. *The Psychology of Gratitude*. R A Emmons, OUP**:** 213-29.

Bos, R. v. d., B. B. Houx, and B. M. Spruijt (2002). Cognition and Emotion in Concert in Human and Nonhuman Animals. *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. Marc  Bekoff, Colin Allen and Gordon M. Burhgardt. Cambridge, Massachusetts, MIT Press**:** 97-103.

Boyd, R., and P. J. Richerson (1985). *Culture and the Evolutionary Process*. Chicago, University of Chicago Press.

Breland, K., and M. Breland (1961). "The Misbehaviour of Organisms." *American Psychologist* **16**: 681-84.

Brink, D. O., D. Copp, G. Sayre-McCord, and M. Smith (1997). "Symposium on Michael Smith's *the Moral Problem*." *Ethics* **108**: 4-119.

Brooks, R. A. (1999). *Cambrian Intelligence: The Early History of the New Ai*. Cambridge, Massachusetts, MIT Press.

Brown, D. (1991). *Human Universals*. New York, McGraw-Hill.

Browne, K. (2002). *Biology at Work: Rethinking Sexual Equality*. Rutgers, Rutgers University Press.

Bshary, R. (2001). "Cleaner Fish Labroides Dimidiatus Manipulate Client Reef Fish by Providing Tactile Stimulation." *Proceedings of the Royal Society: Biological Sciences* **268**(1475): 1495.

Bshary, R., and A. S. Grutter (2002). "Asymmetric Cheating Opportunities and Partner Control in a Cleaner Fish Mutualism." *Animal Behaviour* **63**: 547-55.

Bshary, R., and D. Schäffer (2002). "Choosy Reef Fish Select Cleaner Fish That Provide High-Quality Service." *Animal Behaviour* **63**: 557-64.

Buchan, J. C., S. C. Alberts, J. B. Silk, and J. Altmann (2003). "True Paternal Care in a Multi-Male Primate Society." *Nature* **425**: 179-81.

Bull, H. (1977/1995). *The Anarchical Society: A Study of Order in World Politics*. Columbia University Press.

Buss, D. M. (2000a). *The Dangerous Passion: Why Jealousy Is as Necessary as Love and Sex*, Free Press.

Buss, D. M. (1994). *The Evolution of Desire: Strategies of Human Mating*. New York, Basic Books/HarperCollins.

Buss, D. M. (2000b). "The Evolution of Happiness." *American Psychologist* **55**(1): 15-23.

Buss, D. M. (1999). *Evolutionary Psychology: The New Science of the Mind*. Boston, Allyn and Bacon.

Buss, D. M. (1989). "Sex Differences in Human Mate Preferences: Evolutionary Hypotheses Tested in 37 Cultures." *Behavioral and Brain Sciences* **12**(4): 1-49.

Butler, J. (1856). *Fifteen Sermons Preached at the Rolls Chapel*. London, Longman.

Byrne, R. W. (2000). How Monkeys Find Their Way: Leadership, Coordination and Cognitive Maps of African Baboons. *On the Move: How and Why Animals Travel in Groups*. Sue Boinski and P A Garber**:** 491-518.

Byrne, R. W., and A. Whiten, Eds. (1988). *Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford, Clarendon Press.

Caporael, L., R. Dawes, J. Orbell, and A. van de Kragt (1989). "Selfishness Examined: Cooperation in the Absence of Egoistic Incentives." *Behavioral and Brain Science* **12**(4): 683-738.

Cavalli-Sforza, L. L., and M. W. Feldman (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, New Jersey, Princeton University Press.

Charnov, E. L. (1984). Behavioural Ecology of Plants. *Behavioural Ecology*. J R Krebs and N B Davies. Sunderland, Massachusetts, Sinauer**:** 362-79.

Charnov, E. L. (1993). *Life History Invariants*. New York, Oxford University Press.

Charvet, J. (1981). *A Critique of Freedom and Equality*, Cambridge University Press.

Cheney, D. L., and R. M. Seyfarth (1990). *How Monkeys See the World: Inside the Mind of Another Species*. Chicago, Chicago University Press.

Cheng, P. W., and K. J. Holyoak (1989). "On the Natural Selection of Reasoning Theories." *Cognition* **33**: 285-313.

Cheng, P. W., and K. J. Holyoak (1985). "Pragmatic Reasoning Schemas." *Cognitive Psychology* **17**: 391-416.

Chodorow, N. (1978). *The Reproduction of Mothering*, University of California Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts, MIT Press.

Chomsky, N. (1988). *Language and Problems of Knowledge: The Managua Lectures*. Cambrdge, MA, MIT Press.

Chwe, M. (2001). *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton, New Jersey, Princeton University Press.

Cicero, M. T. (1991). *On Duties/Officia*, Cambridge University Press.

Cicero, M. T. (1971). *On the Good Life*, Penguin Classics.

Cicero, M. T. (1945). *Tusculan Disputations*, Harvard University Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Back Together Again*. Cambridge, Massachusetts, MIT Press.

Clutton-Brock, T. H. (1991). *The Evolution of Parental Care*. Pinceton, NJ, Princeton University Press.

Clutton-Brock, T. H., and S. D. Albon (1979). "The Roaring of Red Deer and the Evolution of Honest Advertisement." *Behaviour* **69**: 145-70.

Clutton-Brock, T. H., and P. H. Harvey (1984). Comparative Approaches to Investigating Adaptation. *Behavioural Ecology: An Evolutionary Approach*. J. R. Krebs and N. B. Davies. Oxford, Blackwell**:** 7-29.

Clutton-Brock, T. H., and G. A. Parker (1995). "Punishment in Animal Societies." *Nature* **373**: 209-16.

Collier, J., and M. Stingl (1993). "Evolutionary Naturalism and the Objectivity of Morality." *Biology and Philosophy* **8**: 47-60.

Confucius (1994). A Single Word. *Ethics*. P Singer. Oxford, Oxford University Press**:** 76.

Cosmides, L. (1989). "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task." *Cognition* **31**: 187-276.

Cosmides, L., and J. Tooby (1992). Cognitive Adaptations for Social Exchange. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. J Barkow, L Cosmides and J Tooby. New York, Oxford University Press**:** 163-228.

Cosmides, L., and J. Tooby (2000). Consider the Source: The Evolution of Adaptations for Decoupling and Metarepresentation. *Metarepresentation: A Multidisciplinary Perspective*. D Sperber. New York, NY, Oxford University Press.

Cosmides, L., and J. Tooby (1981). "Cytoplasmic Inheritance and Intragenomic Conflict." *Journal of Theoretical Biology* **89**: 83-129.

Cosmides, L., and J. Tooby (1989). "Evolutionary Psychology and the Generation of Culture, Part Ii. Case Study: A Computational Theory of Social Exchange." *Ethology and Sociobiology* **10**: 51-97.

Craig, W. (1918). "Appetites and Aversions as Constituents of Instincts." *Biological Bulletin of the Marine Biological laboratory* **34**: 91-107.

Crawford, C., and C. Salmon, Eds. (2004). *Evolutionary Psychology, Public Policy and Personal Decisions*, Lawrence Erlbaum Associates.

Crespi, B. J. (2001). "The Evolution of Social Behaviour in Microorganisms." *Trends in Ecology & Evolution* **16**(4): 178-83.

Cronin, H. (1992). *The Ant and the Peacock*. Cambridge, CUP.

Curry, O. (2003). "Get Real: Evolution as Metaphor and Mechanism." *British Journal of Politics and International Relations* **5**(1): 112-17.

Dall, S. R. X. (2002). "Can Information Sharing Explain Recruitment to Food from Communal Roosts?" *Behavioural Ecology* **13**: 42-51.

Daly, M. (1991). "Natural Selection Doesn't Have Goals, but It's the Reason Organisms Do (Commentary on P. J. H. Shoemaker, "the Quest for Optimality: A Positive Heuristic of Science?")." *Behavioral and Brain Sciences* **14**: 219-20.

Daly, M. (1982). "Some Caveats About Cultural Transmission Models." *Human Ecology* **10**(3): 401-8.

Daly, M., and M. Wilson (2001). "An Assessment of Some Proposed Exceptions to the Phenomenon of Nepotistic Discrimination against Stepchildren." *Annales Zoologici Fennici* **38**: 287-96.

Daly, M., and M. Wilson (1988). *Homicide*. New York, Aldine de Gruyter.

Daly, M., and M. Wilson (1998). *The Truth About Cinderella: A Darwinian View of Parental Love*. London, Weidenfeld & Nicolson.

Daly, M., and M. Wilson (1982). "Whom Are Newborn Babies Said to Resemble?" *Ethology and Sociobiology* **3**: 69-78.

Danielson, P. (1992). *Artificial Morality: Virtuous Robots for Virtual Games*. London, Routledge.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London, John Murray.

Darwin, C. (1872/1998). *The Expression of the Emotions in Man and Animals*. London, John Murray/HarperCollins.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. London, John Murray.

Darwin, C. (1880/1994). The Power of Movement in Plants. *A Darwin Selection*. M Ridley. London, Fontana**:** 253-79.

Das, M. (2000). Conflict Management Via Third Parties: Post-Conflict Affiliation of the Aggressor. *Natural Conflict Resolution*. F Aureli and F B M de Waal. Berkeley, University of California Press**:** 263-80.

Dawkins, R. (1982). *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford, W. H. Freeman.

Dawkins, R. (1980). Good Strategy or Evolutionarily Stable Strategy? *Sociobiology: Beyond Nature/Nurture? Reports, Definitions and Debate*. G. W. Barlow and J. Silverberg. Boulder, Colorado, Westview Press**:** 331-67.

Dawkins, R. "The Information Challenge." *The Skeptic* 1998a.

Dawkins, R. (1976). *The Selfish Gene*. Oxford, Oxford University Press.

Dawkins, R. (1998b). *Unweaving the Rainbow: Science, Delusion and the Appetite for Wonder*. London, Penguin Books.

Dawkins, R., and M. Dawkins (1973). "Decisions and the Uncertainty of Behaviour." *Behaviour* **45**: 83-103.

Dawkins, R., and J. R. Krebs (1978). Animal Signals: Information or Manipulation? *Behavioural Ecology: An Evolutionary Approach*. J. R. Krebs and N. B. Davies. Oxford, Blackwell**:** 282--309.

de Waal, F. (1994). Chimpanzee Justice. *Ethics*. P Singer. Oxford, Oxford University Press**:** 67-9.

de Waal, F. (1982). *Chimpanzee Politics: Power and Sex among Apes*. London, Jonathan Cape.

de Waal, F. (1996). *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, Massachusetts, Harvard University Press.

de Waal, F. (1989). *Peacemaking among Primates*. Cambridge, Massachusetts, Harvard University Press.

Decety, J., and T. Chaminade (2002). "Neural Correlates of Feeling Sympathy." *Neuropsychologia* **41**: 127-38.

Dennett, D. C. (1998). *Brainchildren: Essays on Designing Minds*. London, Penguin Books.

Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London, Allen Lane, The Penguin Press.

Dennett, D. C. (2003). *Freedom Evolves*. London, Allen Lane, The Penguin Press.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, Massachusetts, MIT Press.

Diamond, J. (1999). *Guns, Germs and Steel*, W. W. Norton & Company.

Dittmar, H. (1992). *The Social Psychology of Material Possessions: To Have Is to Be*. Harvester Wheatsheaf, Hemel Hempstead.

Dixit, A., and S. Skeath (1999). *Games of Strategy*. New York, W W Norton & Company.

Doris, J. M., and S. P. Stich ((forthcoming)). As a Matter of Fact: Empirical Perspectives on Ethics. *The Oxford Handbook of Contemporary Analytic Philosophy*. F Jackson and M Smith. Oxford, OUP.

Downie, R. S. (1965). "Forgiveness." *Philosophical Quarterly* **15**(59): 128-34.

Duchaine, B., L. Cosmides, and J. Tooby (2001). "Evolutionary Psychology and the Brain." *Current Opinion in Neurobiology* **11**(2): 225-30.

Dugatkin, L. A. (1997a). *Cooperation among Animals: An Evolutionary Perspective*. New York, Oxford University Press.

Dugatkin, L. A. (1991). "Dynamics of the Tit for Tat Strategy During Predator Inspection in the Guppy (*Poecilia Reticulata*)." *Behavioural Ecology and Sociobiology* **29**: 127-32.

Dugatkin, L. A. (1997b). The Evolution of the "Con-Artist". *The Maladapted Mind*. Simon Baron-Cohen, Psychology Press**:** 189-206.

Dumond, D. E. (1987). *The Eskimos and Aleuts*.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language*. London, Faber & Faber.

Dworkin, R. (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA, Harvard University Press.

Edel, M., and A. Edel (1959/1968). *Anthropology and Ethics: The Quest for Moral Understanding*. Cleveland, Ohio, Case Western Reserve University Press.

Ekman, P. (1985). *Telling Lies*. New York, WW Norton.

Ekman, P., Ed. (1973). *Darwin and Facial Expression: A Century of Research in Review*. New York, Academic Press.

Ekman, P., W. Friesen, and K. Scherer (1976). "Body Movements and Voice Pitch in Deceptive Interaction." *Semiotica* **16**: 23-7.

Ellis, B. J. (1992). The Evolution of Sexual Attraction: Evaluative Mechanisms in Women. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Jerome H. Barkow, Leda Cosmides and John Tooby. New York, Oxford University Press**:** 267-88.

Elster, J. (1999). *Alchemies of the Mind: Rationality and the Emotions*. London, Cambridge University Press.

Euler, H. A., and B. Weitzel (1999). Grandparental Caregiving and Intergenerational Relations Reflect Reproductive Strategies. *The Darwinian Heritage and Sociobiology*. Johan M G van der Dennen, David Smillie and Daniel R Wilson. Westport, Conneticut & London, Praeger.

Evans, D., and O. Zarate (1999). *Introducing: Evolutionary Psychology*, Icon Books.

Farber, P. L. (1994). *The Temptations of Evolutionary Ethics*. Berkeley, University of California Press.

Fehr, E., and S. Gachter (2002). "Altruistic Punishment in Humans." *Nature* **415**: 137-40.

Fiddick, L. (forthcoming). "Domains of Deontic Reasoning: Resolving the Discrepancy between the Congitive and Moral Reasoning Literatures." *The Quarterly Journal of Experimental Psychology*.

Fiddick, L. (2003). Is There a Faculty of Deontic Reasoning? A Critical Reevaluation of Abstract Deontic Versions of the Wason Selection Task. *Evolution and the Psychology of Thinking: The Debate*. D Over, Psychology Press.

Flack, J. C., and F. B. M. de Waal (2000). 'Any Animal Whatever': Darwinian Building Blocks of Morality in Monkeys and Apes. *Evolutionary Origins of Morality*. L D Katz, Imprint Academic: 71-99.

Flew, A. (1978). From Is to Ought. *The Sociobiology Debate: Readings on Ethical and Scientific Issues*. A L Caplan. New York, Harper & Row: 142-62.

Foss, N. J. (2000). "Leadership, Beliefs and Coordination: An Explorative Discussion." *Industrial and Corporate Change* **10**: 357-88.

Frank, R. (1992). The Role of Moral Sentiments in the Theory of Intertemporal Choice. *Choice over Time*. G Loewenstein. New York, Russell Sage Foundation.

Frank, R. (1990). A Theory of Moral Sentiments. *Beyond Self-Interest*. Jane J Mansbridge. Chicago, London, University of Chicago Press: 71-96.

Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. New York, W. W. Norton and Company.

Frankena, W. K. (1957). "Ethical Naturalism Renovated." *The Review of Metaphysics* **10**: 457-73.

Frankena, W. K. (1939). "The Naturalistic Fallacy." *Mind* **48**: 464-77.

Fukuyama, F. (1996). *Trust: The Social Virtues and the Creation of Prosperity*. London, Penguin Books.

Gallistel, C. R. (1999). The Replacement of General-Purpose Learning Models with Adaptively Specialized Learning Modules. *The Cognitive Neurosciences*. Gazzaniga. Cambridge, MA, MIT Press: 1179-91.

Garcia, J., and R. A. Koelling (1966). "Relation of Cue to Consequence in Avoidance Learning." *Psychonomic Science* **4**: 123-24.

Gaulin, S. J., and A. Schlegel (1980). "Paternal Confidence and Paternal Investment: A Cross-Cultural Test of a Sociobiological Hypothesis." *Ethology and Sociobiology* **1**: 301-9.

Geary, D. C. (2000). "Evolution and Proximate Expression of Human Paternal Investment." *Psychological Bulletin* **126**: 55-77.

Geary, D. C. (1998). *Male, Female: The Evolution of Human Sex Differences*. Washington, DC, American Psychological Association.

Geertz, C. (1973). *The Interpretation of Cultures*. New York, Basic Books.

Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Oxford, Clarendon Press.

Gibbard, A. (1997). Wise Choices, Apt Feelings. *Moral Discourse and Practice: Some Philosophical Approaches*. S L Darwall, A Gibbard and R A Railton. New York, Oxford University Press**:** pp179-98.

Gigerenzer, G., and K. Hug (1992). "Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change." *Cognition* **43**: 127-71.

Gigerenzer, G., and R. McElreath (2003). "Social Intelligence in Games." *Journal of Institutional and Theoretical Economics* **159**: 188-94.

Gigerenzer, G., P. M. Todd, and A. R. Group, Eds. (1999). *Simple Heuristics That Make Us Smart*. Evolution and Cognition. New York and Oxford, Oxford University Press.

Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development*, Harvard University Press.

Girotto, V., A. Mazzacco, and P. Cherubini (1992). "Pragmatic Judgements of Relevance in Reasoning: A Reply to Jackson and Griggs." *Quarterly Review of Experimental Psychology* **45a**: 547-74.

Godfrey-Smith, P. (1996). Complexity and the Function of Mind in Nature, CUP.

Goldman, A. (1993). "Ethics and Cognitive Science." *Ethics* **103**(2): 337-60.

Goodall, J. (1994). Helping Kin in Chimpanzees. *Ethics*. P Singer. Oxford, Oxford University Press**:** 60-1.

Goode, E. (1978/97). *Deviant Behavior*. New Jersey, Prentice Hall.

Gray, J. (1995). *Berlin*, Fontana.

Greene, J. (2003). "From Neural 'Is' to Moral 'Ought': What Are the Moral Implications of Neuroscientific Moral Psychology?" *Nature Reviews: Neuroscience* **4**(October): 847-50.

Greene, J., and J. Haidt (2002). "How (and Where) Does Moral Judgment Work?" *Trends in Cognitive Sciences* **6**(12): 517-23.

Gregersen, E. (1982). *Sexual Practices: The Story of Human Sexuality*. New York, F Watts.

Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA, Harvard University Press.

Griffin, R., and S. Baron-Cohen (2002). The Intentional Stance: Developmental and Neurocognitive Perspectives. *Daniel Dennett*. A Brook and D Ross. Cambridge, Cambridge University Press**:** 83-116.

Grimes, K. "To Trust Is Human." *New Scientist*, 10 May 2003 2003, 32.

Grimshaw, J. (1991). The Idea of a Female Ethic. *A Companion to Ethics*. Peter Singer, Blackwell**:** 491-99.

Gross, M. R., and J. Repka (1998). Game Theory and Inheritance in the Conditional Strategy. *Game Theory and Animal Behavior*. Lee Alan Dugatkin and Hudson Kern Reeve. Oxford, Oxford University Press**:** 168-87.

Haig, D. (2003). "On Intrapersonal Reciprocity." *Evolution and Human Behavior* **24**: 418-25.

Haig, D. (1997). The Social Gene. *Behavioural Ecology: An Evolutionary Approach*. J R Krebs and N B Davies. Oxford, Blackwell Science**:** 284-304.

Hamilton, W. D. (1964). "The Genetical Evolution of Social Behaviour." *Journal of Theoretical Biology* **7**: 1-16, 17-52.

Hamilton, W. D. (1971). "Geometry for the Selfish Herd." *Journal of Theoretical Biology* **31**: 295-311.

Harcourt, A., and F. B. M. de Waal, Eds. (1992). *Coalitions and Alliances in Humans and Other Animals*. Oxford, Oxford University Press.

Hare, R. M. (1991). Universal Prescriptivism. *A Companion to Ethics*. Peter Singer, Blackwell.

Harman, G. (2000). *Explaining Value and Other Essays in Moral Philosophy*. Oxford, OUP.

Harman, G. *Three Trends in Moral and Political Philosophy* 2003 [cited. Available from http://www.princeton.edu/~harman/Papers/Trends.pdf.

Harris, J. R. (1998). *The Nurture Assumption: Why Kids Turn out the Way They Do*. London, Bloomsbury.

Harvey, I. (2000). Robotics: Philosophy of Mind Using a Screwdriver. *Evolutionary Robotics: From Intelligent Robots to Artificial Life*. T Gomi. Ontario, Canada, AAI Books. **III:** 207-23.

Hauser, M. (2001). *Wild Minds: What Animals Really Think*. London, Penguin.

Hausman, D. M. *Trust in Game Theory* 1997 [cited. Available from [http://philosophy.wisc.edu/hausman/papers/trust.htm](http://philosophy.wisc.edu/hausman/papers/trust.htm).

Hausman, D. M., and M. S. McPherson (1996). *Economic Analysis and Moral Philosophy*. Cambridge, Cambridge University Press.

Hawkes, K., O. C. J. F, and e. al (2001). "Hadza Meat Sharing." *Evolution and Human Behavior* **22**(2): 113-42.

Hazlitt, H. (1964). *The Foundations of Morality*. Princeton, New Jersey, D. Van Nostrand Company Inc.

Held, V. (2002). "Moral Subjects: The Natural and the Normative." *Proceedings and Addresses of the American Philosophical Association* **76**(2): 7-24.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, R. McElreath, A. Alvard, A. Barr, J. Ensminger, K. Hill, F. Gil-White, M. Gurven, F. Marlowe, J. Q. Patton, N. Smith, and D. Tracer ((forthcoming)). 'Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. *Cooperation, Reciprocity and Punishment: Experiments from 15 Small Scale Societies*. J Henrich, R Boyd, S Bowles, C Camerer, E Fehr, H Gintis and R McElreath.

Hobbes, T. (1651/1958). *Leviathan*, Macmillan.

Hoffman, M. L. (2001). *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge, Cambridge University Press.

Holekamp, K. E., E. E. Boydston, and L. Smale (2000). Group Travel in Social Carnivores. *On the Move: How and Why Animals Travel in Groups*. Sue Boinski and P A Garber**:** 587-627.

Holekamp, K. E., and A. L. Engh (2002). Field Studies of Social Cognition in Spotted Hyenas. *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. Marc Bekoff, Colin Allen and Gordon M Burhgardt. Cambridge, Massachusetts, MIT Press**:** 371-77.

Hollis, M. (1998). *Trust within Reason*. Cambridge, Cambridge University Press.

Hrdy, S. B. (1999). *Mother Nature: A History of Mothers, Infants, and Natural Selection*. New York, Pantheon Books.

Hume, D. (1777). *An Enquiry Concerning the Principle of Morals*.

Hume, D. (1739/1985). *A Treatise of Human Nature*, Penguin Classics.

Humphrey, N. (1976). The Social Function of Intellect. *Growing Points in Ethology*. P. P. G. Bateson and R. A. Hinde. Cambridge, Cambridge University Press**:** 303-25.

Jackson, F., P. Pettit, and M. Smith (2004). *Mind, Morality, and Explanation: Selected Collaborations*. Oxford, OUP.

Johnson, R. C. (1996). "Attributes of Carnegie Medalists Performing Acts of Heroism and of the Recipients of Those Acts." *Ethology and Sociobiology* **17**: 355.

Johnson-Laird, P. N., P. Legrenzi, and M. S. Legrenzi (1972). "Reasoning and a Sense of Reality." *Brtish Journal of Psychology* **63**: 395-400.

Johnstone, R. A. (1998). Game Theory and Communication. *Game Theory and Animal Behavior*. Lee Alan Dugatkin and Hudson Kern Reeve. Oxford, Oxford University Press**:** 94-117.

Jones, M. C. (1957). "The Later Careers of Boys Who Were Early or Late Maturing." *Child Development* **28**: 113-28.

Kahneman, D., P. Slovic, and A. Tversky, Eds. (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, Cambridge University Press.

Kahneman, D., and A. Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica*: 263-91.

Kamil, A. C. (1994). A Synthetic Approach to the Study of Animal Intelligence. *Behavioral Mechanisms in Evolutionary Ecology*. L A Real. Chicago, Chicago University Press**:** 11-45.

Keller, L., Ed. (1999). *Levels of Selection in Evolution*, Princeton University Press.

Ketelaar, T., and W. T. Au (2003). "The Effects of Feeling Guilt on the Behaviour of Uncooperative Individuals in Repeated Social Bargaining Games: An Affect-as-Information Interpretation of the Role of Emotion in Social Interaction." *Cognition and Emotion* **17**(3): 429-53.

Kohlberg, L. (1981). *The Philosophy of Moral Development: Moral Stages and the Idea of Justice*, HarperCollins.

Korsgaard, C. M. (1996). *The Sources of Normativity*. Cambridge, Cambridge University Press.

Krebs, D. L. (1998). The Evolution of Moral Behaviors. *Handbook of Evolutionary Psychology: Ideas, Issues, and Applications*. C Crawford and D L Krebs, Lawrence Erlbaum Associates**:** 337-68.

Kropotkin, P. (1902). *Mutual Aid: A Factor of Evolution*, Penguin, Middlesex.

Kummer, H. (1991). "Evolutionary Transformations of Possessive Behavior." *Journal of Social Behavior and Personality (To Have Possessions: A handbook on ownership and property)* **6**(6): 75-83.

Kummer, H., and M. Cords (1991). "Cues of Ownership in Long-Tailed Macaques, *Macaca Fascicularis*." *Animal Behaviour* **42**: 529-49.

Kurzban, R., J. Tooby, and L. Cosmides (2001). "Can Race Be Erased? Coalitional Computation and Social Categorization." *Proceedings of the National Academy of Sciences* **98**(26): 15387-92.

Lawrence, A. D., and A. J. Calder (2004). Homologizing Human Emotions. *Emotions, Evolution and Rationality*. D Evans and P Cruse. Oxford, Oxford University Press**:** 15-47.

Leimar, O., and A. H. Axén (1993). "Strategic Behavior in an Interspecific Mutualism: Interactions between Lycaenid Larvae and Ants." *Animal Behaviour* **46**: 1177-82.

Lenski, R. E., and G. J. Velicer (2000). "Games Microbes Play." *Selection* **1**: 51-57.

Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, Massachusetts, Harvard University Press.

Lieberman, D., J. Tooby, and L. Cosmides (2003). "Does Morality Have a Biological Basis? An Empirical Test of the Factors Governing Moral Sentiments Relating to Incest." *Proceedings of the Royal Society* **February 28**.

Ligon, J. D. (1991). Cooperation and Reciprocity in Birds and Mammals. *Kin Recognition*. PG Hepper. Cambridge, Cambridge University Press**:** 30-59.

Litwinski, L. (1942). "Is There an Instinct of Possession?" *British Journal of Psychology* **33**: 28-39.

Locke, J. (1992). Second Treatise of Government. *Classics of Moral and Political Theory*. M L Morgan. Indianapolis/Cambridge, Hackett**:** 736-817.

Lorenz, K. (1966a). *Evolution and Modification of Behavior*. London, Methuen.

Lorenz, K. (1966b). *On Aggression*. London, Methuen.

Lormand, E. (1999). Frame Problem. *The Mit Encyclopedia of the Cognitive Sciences*. R A Wilson and F C Kiel, MIT Press**:** 326-27.

Low, B. S. (1989). "Cross-Cultural Patterns in the Training of Children: An Evolutionary Perspective." *Journal of Comparative Psychology* **103**(4): 311-19.

Lumsden, C. J., and E. O. Wilson (1981). *Genes, Mind, and Culture: The Coevolutionary Process*. Cambridge, Mass, Harvard University Press.

Machiavelli, N. (1988). *The Prince*, Cambridge University Press.

MacIntyre, A. C. (1981). *After Virtue*. London, Gerald Duckworth & Co. Ltd.

MacIntyre, A. C. (1959). "Hume on "Is and "Ought"." *The Philosophical Review* **68**(4): 451-68.

MacIntyre, A. C. (1966). *A Short History of Ethics*. New York, MacMillan.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*, Penguin.

Mackie, J. L. (1980). *Hume's Moral Theory*. London, Routledge & Kegan Paul Ltd.

MacPherson, C. B. (1962). *The Political Theory of Possessive Individualism*. Oxford, Oxford University Press.

Mansbridge, J. J., Ed. (1990). *Beyond Self-Interest*. Chicago/London, University of Chicago Press.

Masters, R. D. (1991). Political Science. *The Sociobiological Imagination*. M Maxwell, State University of New York Press.

May, L., M. Friedman, and A. Clark, Eds. (1996). *Mind and Morals: Essays on Cognitive Science and Ethics*. Cambridge, Massachusetts, MIT Press.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge, Cambridge University Press.

Maynard Smith, J., and G. R. Price (1973). "The Logic of Animal Conflict." *Nature* **246**: 15-18.

Maynard Smith, J., and E. Szathmáry (1995). *The Major Transitions in Evolution*. Oxford, Oxford University Press.

Mazur, A., and A. Booth (1998). "Testosterone and Dominance in Men." *Behavioral and Brain Sciences* **21**: 353-97.

McGrew, W. C., and A. T. C. Feistner (1992). Two Nonhuman Primate Models for the Evolution of Human Food Sharing: Chimpanzee and Callitrichids. *The Adapted Mind: Evolutionary Psychology and the*

*Generation of Culture*. J Barkow, L Cosmides and J Tooby. New York, Oxford University Press**:** 229-43.

Mealey, L. (1997). The Sociobiology of Sociopathy: An Integrated Evolutionary Model. *The Maladapted Mind*. Simon Baron-Cohen, Psychology Press.

Mealey, L., C. Daood, and M. Krage (1996). "Enhanced Memory for Faces of Cheaters." *Evolution and Human Behavior* **17**(2): 119-28.

Mech, L. D. (1970). *The Wolf: The Ecology and Behavior of an Endangered Species*. Garden City, NY, Natural History Press.

Milinski, M., D. Ḳlling, and R. Kettler (1990). "Tit for Tat: Sticklebacks (*Gasterosteus Aculeatus*) 'Trusting' a Cooperating Partner." *Behavioural Ecology* **1**: 7-11.

Mill, J. S. (1868/1999). In Favour of Capital Punishment. *Political Thought*. M Rosen and J Wolff. Oxford, Oxford University Press**:** 181-82.

Mill, J. S., and J. Bentham (1987). *Utilitarianism and Other Essays*, Penguin.

Miller, G. F. (1998). How Mate Choice Shaped Human Nature: A Review of Sexual Selection and Human Evolution. *Handbook of Evolutionary Psychology: Ideas, Issues, and Applications*. C Crawford and D Krebs, Lawrence Erlbaum**:** 87-129.

Miller, G. F. (2000a). *The Mating Mind*. London, William Heinemann.

Miller, G. F. (2000b). "Memetic Evolution and Human Culture." *Quarterly Review of Biology* **75**(4): 434-36.

Miller, G. F. (2004). *Sexual Selection and Moral Belief*. The Psychology and Biology of Morality, Dartmouth College, Hanover, NH.

Miller, J. G., and D. M. Bersoff (1988). "When Do American Children and Adults Reason in Social Conventional Terms?" *Developmental Psychology* **24**(3): 366-75.

Moore, G. E. (1903). *Principia Ethica*, Cambridge University Press.

Murphy, D. P., and S. Stich (2000). Darwin in the Madhouse: Evolutionary Psychology and the Classification of Mental Disorders. *Evolution and the Human Mind*. P Carruthers and A Chamberlain, Cambridge University Press**:** 62-92.

Murphy, J. G. (1982). *Evolution, Morality, and the Meaning of Life*. Totowa, NJ, Rowman and Littlefield.

Nagel, T. (1970). *The Possibility of Altruism*, Clarendon Press.

Nesse, R. M. (2001a). Natural Selection and the Capacity for Subjective Commitment. *Evolution and the Capacity for Commitment*. R Nesse. New York, Russell Sage Press**:** 1-45.

Nesse, R. M., Ed. (2001b). *Evolution and the Capacity for Commitment*. New York, Russell Sage Press.

Nesse, R. M., and G. C. Williams (1994). *Evolution and Healing: The New Science of Darwinian Medicine*. London, Pheonix.

Noddings, N. (1978). *Caring: A Feminine Approach to Ethics and Education*. Berkeley, University of California Press.

Noë, R. (1992). Alliance Formation among Male Baboons: Shopping for Profitable Partners. *Coalitions and Alliances in Humans and Other Animals*. A Harcourt and F B M de Waal. Oxford, Oxford University Press.

Noë, R., J. A. R. A. M. van Hooff, and P. Hammerstein, Eds. (2001). *Economics in Nature: Social Dilemmas, Mate Choice and Biological Markets*. Cambridge, CUP.

Norris, K. S., and C. R. Schilt (1988). "Cooperative Societies in Three-Dimensional Space: On the Origins of Aggregations, Flocks, and Schools with Special Reference to Dolphins and Fish." *Ethology and Sociobiology* **9**(2-4): 149-79.

Nowak, M. A., and K. Sigmund (1998). "Evolution of Indirect Reciprocity by Image Scoring." *Nature* **393**: 573-77.

Nunn, C. L., and R. J. Lewis (2001). Cooperation and Collective Action in Animal Behaviour. *Economics in Nature: Social Dilemmas, Mate Choice and Biological Markets*. Cambridge, CUP**:** 42-66.

O'Hear, A. (1985). *What Philosophy Is*, Penguin.

O'Neill, O. (2002). *A Question of Trust*. Cambridge, Cambridge University Press.

O'Neill, P., and L. Petrinovich (1998). "A Preliminary Cross-Cultural Study of Moral Intuitions." *Evolution and Human Behaviour* **19**(?3): 49-167?

Orbell, J., T. Morikawa, and N. Allen (2000). "The Evolution of Political Intelligence: Simulation Results." *(manuscript)*.

Ostrom, E. (1998). "A Behavioral Approach to the Rational Choice Theory of Collective Action." *American Political Science Review* **92**(1).

Packer, C. (1977). "Reciprocal Altruism in *Papio Anubis*." *Nature* **201**: 441-43.

Palameta, B., and W. M. Brown (1999). "Human Cooperation Is More Than by-Product Mutualism." *Animal Behaviour* **57**: F1-F3.

Panksepp, J. (2000). Emotions as Natural Kinds within the Mammalian Brain. *The Handbook of Emotions*. M Lewis and J Haviland. New York, Guilford**:** 137-56.

Persico, N., A. Postlewaite, and D. Silverman (2001). "The Effect of Adolescent Experience on Labor Market Outcomes: The Case of Height." *Penn Institute for Economic Research Working Paper* **01**(050).

Piaget, J. (1997). *The Moral Judgement of the Child*, Free Press.

Pigden, C. R. (1991). Naturalism. *A Companion to Ethics*. Peter Singer, Blackwell**:** 421-31.

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. London, Allen Lane.

Pinker, S. (1997). *How the Mind Works*. New York, W W Norton.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York, William Morrow and Company.

Pitchford, I. *Theory of Mind Bibliography* [Webpage]. 2001 [cited. Available from http://groups.yahoo.com/group/evolutionary-psychology/message/11499.

Platek, S. M. (2003). *Paternal Uncertainty, the Brain, and Children's Faces: Neural Correlates of Child Facial Resemblance*. HBES, Lincoln, Nebraska, USA.

Platek, S. M., R. L. Burch, I. S. Panyavin, B. H. Wasserman, and G. G. Gallup Jr (2002). "Reactions to Children's Faces: Resemblance Affects Males More Than Females." *Evolution and Human Behavior* **23**: 159-66.

Platek, S. M., J. A. Davis, F. B. Mohamed, and G. G. J. Gallup (under review). "Children on the Mind: Sex Differences in Neural Correlates of Facial Resemblance." *Evolution and Human Behavior*.

Plato (1974). *The Republic*, Penguin Books.

Plomin, R. (1990). *Nature and Nurture: An Introduction to Human Behavioral Genetics*. Belmont, California, Brooks/Cole, A Division of Wadsworth.

Popper, K. R. (1999). *All Life Is Problem Solving*. Bristol, Routledge.

Popper, K. R. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London and New York, Routledge.

Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach.* Oxford, Oxford University Press.

Popper, K. R. (1945). *The Open Society and Its Enemies.* London, Routledge.

Popper, K. R. (1990). *A World of Propensities.* Bristol, Thoemmes.

Posner, E. A. (2000). *Law and Social Norms.* Cambridge, Massachusetts, Harvard University Press.

Posner, R. A. (1973). *Economic Analysis of Law.* Boston & Toronto, Little, Brown and Company.

Posner, R. A. (1981). *Economics of Justice.* Cambridge, Massachusetts, Harvard University Press.

Posner, R. A. (1990). *The Problems of Jurisprudence,* Harvard University Press.

Posner, R. A. (1998). "Rational Choice, Behavioral Economics, and the Law." *Stanford Law Review* **50**: 1551-75.

Povinelli, D., and L. R. Godfrey (1993). So Noble in Reason? Chimpanzees' Theory of Mind? *Evolutionary Ethics.* M H Nitecki and D V Nitecki. Albany, NY, State University of New York Press.

Preston, S. D., and F. B. M. de Waal (2002). "Empathy: Its Ultimate and Proximate Bases." *Behavioral and Brain Sciences* **25**: 1-72.

Preuschoft, S., and C. P. van Schaik (2000). Dominance and Communication: Conflict Management in Various Social Settings. *Natural Conflict Resolution.* F Aureli and F B M de Waal. Berkeley, University of California Press**:** 77-105.

Price, M. E., L. Cosmides, and J. Tooby (2002). "Punitive Sentiment as an Anti-Free Rider Psychological Device." *Evolution and Human Behavior* **23**(203-231).

Putnam, H. (1981). *Reason, Truth and History.* Cambridge, Cambridge University Press.

Radcliffe Richards, J. (2001). *Human Nature after Darwin: A Philosophical Introduction.* London, Routledge.

Railton, P. (2000). Darwinian Building Blocks. *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives.* L D Katz. Devon, Imprint Academic**:** 126-30.

Rawls, J. (2000). *Lectures on the History of Moral Philosophy,* Harvard University Press.

Rawls, J. (1971). *A Theory of Justice,* Harvard University Press.

Real, L. A. (1994). A Synthetic Approach to the Study of Animal Intelligence. *Behavioral Mechanisms in Evolutionary Ecology*. L A Real. Chicago, Chicago University Press: 1-8.

Richards, N. (1988). "Forgiveness." *Ethics* **99**(1): 77-97.

Ridley, M. (2000). *Mendel's Demon: Gene Justice and the Complexity of Life*. London, Weidenfeld & Nicolson.

Ridley, M. (1996). *The Origins of Virtue*, Penguin.

Riechert, S. E. (1998). Game Theory and Animal Contests. *Game Theory and Animal Behavior*. Lee Alan Dugatkin and Hudson Kern Reeve. Oxford, Oxford University Press: pp64-93.

Rilling, J. K., D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kilts (2002). "A Neural Basis for Social Cooperation." *Neuron* **35**: 395-405.

Rilling, J. K., and C. D. Kilts (2003). "Imaging the Neural Correlates of Mate Competition in Dominant Male Rhesus Monkeys." *American Journal of Physical Anthropology Supplement* **36**(179).

Rosenblueth, A., N. Weiner, and J. Bigelow (1943). "Behaviour, Purpose and Teleology." *Philosophy of Science* **10**(1): 18-24.

Rottschaefer, W. A. "The Scientific Naturalization of Ethics." 1997.

Ruddick, S. (1980). "Maternal Thinking." *Feminist Studies* **6**.

Ruse, M. (1987). "Darwinism and Determinism." *Zygon* **22**(4): 419-42.

Ruse, M. (1995). *Evolutionary Naturalism: Selected Essays*. London, Routledge.

Ruse, M. (1986). *Taking Darwin Seriously: A Naturalistic Approach to Philosophy*. Oxford, Blackwell.

Russel, B. (1927). *Philosophy*. New York, Norton.

Sahlins, M. (1972). *Stone Age Economics*. Chicago, Aldine.

Samuels, R., S. Stich, and M. Bishop (2002). Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear. *Common Sense, Reasoning and Rationality*. Renee Elio. New York, Oxford University Press.

Samuels, R., S. Stich, and P. D. Tremoulet (1999). Rethinking Rationality: From Bleak Implications to Darwinian Modules. *What Is Cognitive Science?* E LePore and Z Pylyshyn. Oxford, Blackwell.

Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2003). "The Neural Basis of Economic Decision-Making in the Ultimatum Game." *Science* **300**(5626): 1755-58.

Schelling, T. C. (1978). "Altruism, Meanness, and Other Potentially Strategic Behaviours." *American Economic Review* **68**(1): 229-30.

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, Massachusetts, Harvard University Press.

Schneewind, J. B. (1991). Modern Moral Philosophy. *A Companion to Ethics*. Peter Singer, Blackwell**:** 147-57.

Seyfarth, R. M., and D. L. Cheney (2002). Dennett's Contribution to Research on the Animal Mind. *Daniel Dennett*. A Brook and D Ross. Cambridge, Cambridge University Press**:** 117-39.

Seyfarth, R. M., and D. L. Cheney (1984). "Grooming, Alliances, and Reciprocal Altruism in Vervet Monkeys." *Nature* **308**: 541-43.

Sidgwick, H. (1962). *The Methods of Ethics*, MacMillan.

Sifferd, K. "Dis-Solving (Not Solving) the Frame Problem: What Evolution Can Tell Us About Organic Intelligence." MA, King's College, 2002.

Sigg, H., and J. Falett (1985). "Experiments on Respect of Possession in Hamadryas Baboons (*Papio Hamadryas*)." *Animal behaviour* **33**: 181-88.

Silk, J. B. (1998). "Making Amends: Adaptive Perspectives on Conflict Remediation in Monkeys, Apes, and Humans." *Human Nature* **9**: 341-68.

Silk, J. B. (2002). "Using the F-Word in Primatology." *Behaviour* **139**(Special Issue: What are Friends For? The adaptive value of social bonds in primate groups (February/March)): 421-46.

Singer, P. (1999). *A Darwinian Left?: Politics, Evolution and Cooperation*. London, Weidenfeld & Nicolson.

Singer, P. (1981). *The Expanding Circle: Ethics and Sociobiology*, Straus and Giroux.

Singer, P., Ed. (1994). *Ethics*. Oxford Readers. Oxford, Oxford University Press.

Singh, D. (1993). "Body Shape and Women's Attractiveness: The Critical Role of Waist-to-Hip Ratio." *Human Nature* **4**(3): 297--321.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge, Cambridge University Press.

Skyrms, B. (2004). *The Stag Hunt and the Evolution of Social Structure*. Cambridge, Cambridge University Press.

Slobodchikoff, C. N. (2002). Cognition and Communication in Prairie Dogs. *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. Marc Bekoff, Colin Allen and Gordon M Burhgardt. Cambridge, Massachusetts, MIT Press**:** 257-64.

Smith, A. (1759/1976). *The Theory of Moral Sentiments*. Oxford, Oxford Univerity Press.

Smith, E. A. (1999). "Costly Signaling and Collective Action Problems." *(Paper on file)*.

Smith, E. A., and R. L. Bleige Bird (2000). "Turtle Hunting and Tombstone Opening: Public Generosity as Costly Signaling." *Evolution and Human Behavior* **21**: 245-61.

Smith, M. (1994). *The Moral Problem*.

Smith, M. (1991). Realism. *A Companion to Ethics*. Peter Singer, Blackwell**:** 399-410.

Smolker, R. (2000). Group Movement in Dolphins. *On the Move: How and Why Animals Travel in Groups*. Sue Boinski and P A Garber**:** 559-86.

Smuts, B. B. (1985). *Sex and Friendship in Baboons*. New York, Aldine Publishing Company.

Sober, E., and D. S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge, Massachusetts, Harvard University Press.

Sperber, D. (2001). "An Evolutionary Perspective on Testimony and Argumentation." *Philosophical Topics* **29**: 401-13.

Sperber, D., and D. Wilson (1986). *Relevance: Communication and Cognition*. Oxford, Blackwell.

Stich, S. P. (1993). Moral Philosophy and Mental Representation. *The Origin of Values*. M Hechter, L Nadel and R E Michod. New York, Aldine de Gruyter**:** 215-28.

Stone, V. E., S. Baron-Cohen, L. Cosmides, J. Tooby, and R. T. Knight (1997). *Selective Impairment of Social Inference Abilities Following Orbitofrontal Cortex Damage*. Nineteenth Annual Conference of the Cognitive Science Society, London, Larwrence Erlbaum.

Stone, V. E., L. Cosmides, J. Tooby, N. Kroll, and R. T. Knight (2002). "Selective Impairment of Reasoning About Social Exchange in a

Patient with Bilateral Limbic System Damage." *Proceedings of the National Academy of Sciences* **99**(17): 11531-36.

Sturgeon, N. L. (2003). "Moore on Ethical Naturalism." *Ethics* **113**: 528-56.

Sugiyamaa, L. S., J. Tooby, and L. Cosmides (2002). "Cross-Cultural Evidence of Cognitive Adaptations for Social Exchange among the Shiwiar of Ecuadorian Amazonia." *Proceedings of the the National Academy of Sciences* **99**(17): 11537-42.

Symons, D. (1995). Beauty Is in the Adaptations of the Beholder: The Evolutionary Psychology of Human Female Sexual Attractiveness. *Sexual Nature, Sexual Culture*. Paul R. Abramson and Steven D. Pinkerton. Chicago, University of Chicago Press**:** 80-118.

Symons, D. (1979). *The Evolution of Human Sexuality*. New York, Oxford University Press.

Thornhill, R., and C. T. Palmer (2000). *A Natural History of Rape: Biological Bases of Sexual Coercion*. Cambridge, Massachusetts, MIT Press.

Tiger, L., and J. Shepher (1975). *Women in the Kibbutz*. New York, Harcourt Brace Jovanovich.

Tooby, J., and L. Cosmides (1988). "The Evolution of War and Its Cognitive Foundations." *Institute for Evolutionary Studies Technical report* **88**(1).

Tooby, J., and L. Cosmides (1996). Friendship and the Banker's Paradox: Other Pathways to the Evolution of Adaptation for Altruism. *Evolution of Social Behaviour Patterns in Primates and Man: A Joint Discussion Meeting of the Royal Society and the British Academy (Proceedings of the British Academy 88)*. W G Runciman, John Maynard Smith and Robin I M Dunbar, British Academy/Oxford University Press.

Tooby, J., and L. Cosmides (1990). "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments." *Ethology and Sociobiology* **11**: 375-424.

Tooby, J., and L. Cosmides (1992). The Psychological Foundations of Culture. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Jerome H. Barkow, Leda Cosmides and John Tooby. New York, Oxford University Press**:** 19-136.

Tooby, J., and I. DeVore, Eds. (1987). *The Reconstruction of Hominid Behavioral Evolution through Strategic Modeling*. The Evolution of Human Behavior: Primate Models. Albany, N. Y., SUNY Press.

Townsend, J. M. (1998). "Dominance, Sexual Activity, and Sexual Emotions." *Behavioral and Brain Sciences* **21**: 386.

Trivers, R. L. (1983). *The Evolution of a Sense of Fairness*. Absolute Values and the Creation of the New World: Proceedings of the Eleventh International Conference on the Unity of the Sciences, New York, International Cultural Foundation Press.

Trivers, R. L. (1971). "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* **46**: 35-57.

Trivers, R. L. (1974). "Parent-Offspring Conflict." *American Zoologist* **14**: 249-64.

Trivers, R. L. (1972). Parental Investment and Sexual Selection. *Sexual Selection and the Descent of Man 1871--1971*. B. Campbell. London, Heinemann**:** 136-79.

Trivers, R. L. (1985). *Social Evolution*. Menlo Park, California, Benjamin/Cummings.

Trivers, R. L. (1981). Sociobiology and Politics. *Sociobiology and Human Politics*. Elliott White. Lexington, Mass., D. C. Heath and Company**:** 1-43.

Uexküll, J. v. (1934/1957). A Stroll through the Worlds of Animals and Men: A Picture Book of Invisible Worlds. *Instinctive Behavior*. C H Schiller, International Universities Press**:** 5-80.

Ulen, T. S. (2001). "Evolution, Human Behaviour and Law: A Response to Owen Jones's Dunwoody Lecture." *Florida Law Review* **53**(5): 931-46.

Walzer, M. (1983). *Spheres of Justice: A Defense of Pluralism and Equality*, Basic Books.

Westermarck, E. A. (1906). *The Origin and Development of Moral Ideas*. London, Macmillan.

Westermarck, E. A. (1994). A Universal Duty. *Ethics*. P Singer. Oxford, Oxford University Press**:** 61-63.

Whiten, A. ((forthcoming)). "Cutural Panthropology." *Evolutionary Anthropology*.

Whiten, A. (1996). When Does Smart Behaviour-Reading Become Mind-Reading? *Theories of Theories of Mind*. P Carruthers and P K Smith. Cambridge, CUP.

Whiten, A., and R. W. Byrne, Eds. (1997). *Machiavellian Intelligence Ii*. Cambridge, Cambridge University Press.

Wilkinson, G. S. (1988). "Reciprocal Altruism in Bats and Other Mammals." *Ethology and Sociobiology* **9**: 85-100.

Wilkinson, G. S. (1984). "Reciprocal Food Sharing in the Vampire Bat." *Nature* **308**: 181-84.

Wilkinson, R. G. (2001). *Mind the Gap: Hierarchies, Health and Human Evolution*. London, Weidenfeld & Nicolson.

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton, New Jersey, Princeton University Press.

Wilson, E. O. (1978). *On Human Nature*. Cambridge, Mass, Harvard University Press.

Wilson, E. O. (1975). *Sociobiology: The New Synthesis*. Cambridge, Mass, Harvard University Press.

Wilson, M., and M. Daly (1985). "Competitiveness, Risk Taking, and Violence: The Young Male Syndrome." *Ethology and Sociobiology* **6**: 59-73.

Wilson, M., and M. Daly (1997). "Life Expectancy, Economic Inequality, Homicide, and Reproductive Timing in Chicago Neighbourhoods." *British Medical Journal* **314**(26 April): 1271--4.

Wilson, M., and M. Daly (1992). The Man Who Mistook His Wife for a Chattel. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. J H Barkow, L Cosmides and J Tooby. New York, Oxford University Press**:** 289-322.

Wong, D. (1984). *Moral Relativity*, UC California Press.

Wong, D. (1991). Relativism. *A Companion to Ethics*. Peter Singer, Blackwell**:** 442-50.

Wrangham, R., and D. Peterson (1996). *Demonic Males: Apes and the Origins of Human Violence*. New York, Houghton Mifflin.

Wright, R. (1994). *The Moral Animal: The New Science of Evolutionary Psychology*. New York, Pantheon Books.

Wright, R. (2000). *Non-Zero: The Logic of Human Destiny*. New York, Pantheon Books.

Zahavi, A., and A. Zahavi (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford, Oxford Univerity Press.