Human Nature and Morality

An investigation of the evidence for and implications of genetically-based moral traits

by

Bruce Carruthers Martin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2007

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In his recent book, *Moral Minds*, Marc Hauser claims that humans are genetically endowed with a moral faculty operating in much the same way as our linguistic faculty, and that this faculty delimits normative moral systems. Further, he states that this work represents the beginning of what will become a science of morality.

These claims contrast sharply with the conception of human nature presupposed by many of the dominant Western moral theories. For the most part, these conceptions of human nature are not flattering: they suggest that our natural instincts, in large part, or in whole, are not conducive to living a moral life. Given these presuppositions, such theories typically call for setting aside our natural instincts when determining the basis upon which normative moral theory should be established.

This thesis seeks to show that there is a middle ground between these two views. On my account, recent scientific learning about innate traits impacting our behaviour towards others can be employed to construct a conception of human nature that is at odds with that used by a number of the dominant Western moral theories. As the impact of such innate traits is constrained by our analytic intellect, however, I argue that views such as Hauser's overstate the implications for normative moral theory.

Acknowledgements

I would like to thank my supervisor, Dr. Brian Orend, for inspiring me to pursue graduate work in philosophy and for challenging me at every step of the development of this thesis. I would also like to thank my readers, Dr. Patricia Marino and Dr. Dave DeVidi, for agreeing to read my work and for their valuable comments.

To my parents, Forbes and Mary Martin, for always saying “you can do it.”

Table of Contents

Chapter 1 Introduction

In his recent book, *Moral Minds*, Marc Hauser, the head of Harvard's Cognitive Evolution Laboratory, claims that humans are genetically endowed with a moral faculty operating in much the same way as our linguistic faculty, and that this faculty delimits normative moral systems. Further, he states that this work represents the beginning of what will become a science of morality.[1]

Hauser's "scientific moral view" contrasts sharply with what I will refer to as the "dominant moral view" in Western philosophy. Many Western moral philosophers have presupposed a theory of human nature when constructing their normative moral theories. For the most part, these conceptions of human nature are not flattering: they suggest that our natural instincts, in large part, or in whole, are not conducive to living a moral life. Given this, such theories typically call for setting aside our natural instincts when determining the basis upon which normative moral theories should be established. Typically, this results in calling for the suppression of our natural instincts in favour of our analytic intellect, our conscious, free reasoning capacity.

Examples of this approach can be found in most every period of Western philosophy, including the present. In the *Republic*, Plato argues for humans to subdue the passions, which often entice us toward immoral behaviour, and employ reason to develop an

[1] Hauser, 2006

understanding of the universal good.[1] In *Leviathan*, Thomas Hobbes calls for man to give over his right to do whatever he wishes, and his natural instinct for savage behaviour, to the Leviathan, or government, in a type of rational social contract.[2] In the *Groundwork for the Metaphysics of Morals*, Immanuel Kant argues that our animal instincts must be suppressed and reason allowed to rule.[3] Many of these ideas continue to dominate Western moral thinking today. Brian Orend's *War and International Justice: A Kantian Perspective* is an example of how current work in practical philosophy follows this pattern. Orend seeks to provide a coherent view of the ethics of war and peace for the 21st century, using a Kantian perspective. His interpretation of Kant's views on human nature represent a clear depiction of the impact such views have had, and continue to have, on Western moral thinking. Orend writes:

> All of Kant's writings about morality and justice are grounded in a conception of human nature as being split between "free rationality and animal instinctuality," with rationality taking pride of place as our most deep and distinctive sense of identity and interest. Kant stipulates that, regardless of the base inclinations of our animal natures, we are to adhere to the dictates of reason itself. We must, above all, remain true to our most profound identity as rational agents.[4]

Although there are dramatic differences in the normative moral theories they generate, I group the theories identified above, and all others which presuppose a similar

---

[1] Plato, Circ. 360B.C./2005

[2] Hobbes, 1651/1962

[3] Kant, 1785/2002

[4] Orend, 2000, page 16

theory of human nature, under the heading of the "dominant moral view", because of their common view that human nature, in some way, represents an impediment to living a moral life.

To make the contrast clear, the scientific moral view claims that by fully understanding our human nature, our innate behavioural traits, science can tell us what it is to be moral, and how to go about achieving a moral life. The dominant moral view claims that our innate behavioural traits are not helpful to morality and that only an appropriate use of autonomous reasoning can help us decide what values we must follow and how to go about following them.

This thesis seeks to show that there is a middle ground between the scientific moral view and the dominant moral view. I argue that an informed view of human nature, at this point in time, should show humans to possess a number of important instincts, or innate behavioural traits, which are highly conducive to living a moral life. Such traits have only become apparent in the past few decades, through work in the fields of cognitive science, evolutionary psychology and sociobiology. I argue that these traits play a significant enough role in our behaviour towards others to represent a legitimate challenge to the conception of human nature used in the dominant moral view. At the same time, I argue that the impact of these traits on our actions is constrained by our analytic intellect, our conscious, free reasoning capacity. Hence, I suggest that the scientific moral view overstates the normative value of these traits.

1.1 Definitions

It is important to have a clear understanding of how the word “morality” will be used in this thesis, as the term has a number of different connotations within philosophical discourse. This thesis focuses on a foundational issue affecting a variety of moral theories: those which assume the dominant moral view of human nature. Therefore, the definition should incorporate only that which can be generally accepted as fundamental to the term “morality”, avoiding any unnecessary specificity which might align it with a particular normative moral theory. Consistent with this foundational perspective, the definition should also exclude any specific aspects of moral codes derived from culture or religion. Given the above, the term “morality”, within this project, should be considered to refer to a set of values indicating how people should treat each other in order to live peacefully together. This is sufficiently broad for a foundational project of this sort, generally consistent with the underlying definition of morality used for those theories representing the dominant moral view, and still useful for an investigation of innate behavioural traits. When an action, or type of behaviour, is identified as moral in this thesis, therefore, it should be understood as the sort of behaviour that would generally lead to, or maintain, peaceful coexistence with other humans, in the absence of any other moral code established by humans.

One further issue that needs to be addressed in the context of defining morality is: what constitutes a morally relevant act? Some normative theories consider actions as having primary importance in determining morality (most consequentialist approaches, for example), while others view intentions as more important. In a deontological approach, such as that developed by Kant, for instance, an action is deemed moral only if the actor

undertakes it as an exercise of her own free will, and with the intention of following universal moral principles, rather than achieving any specific end result.[1]

This creates difficulties for discussing innate behavioural traits in the context of morality. To illustrate this difficulty, consider that, on Kant's account, if it were possible to show that humans have an innate behavioural trait that plays a significant role in predisposing us to peaceful interactions with others, such traits could not properly be considered morally important, or even morally meaningful, because the actions they promote would not have been undertaken as a decision of the actor using her own free will. In other words, the fact that we might *automatically* or *instinctively* do something morally good makes it an amoral action: one that is neither moral nor immoral.

For the purpose of this thesis, I will set aside these types of constraints on morality. This is appropriate, because the theories in question presuppose a particular conception of human nature and this project seeks to re-evaluate such presuppositions. Constraining a foundational investigation of this sort, by the notions resulting from the theories which the foundations support, requires a circular argument. In other words, an evaluation of the underpinnings of such projects should not be limited by any of the conclusions of the projects, at least for the purposes of the evaluation. It is important to note, however, that I do not suggest that the result of this re-evaluation of human nature will, by itself, undermine any

[1] Kant, 1785/2002

major conclusions of Western moral theory: simply that those conclusions must be set aside for the purposes of the investigation.

Hence, any innate behavioural traits which incline us to, or directly assist us in, behaviour that promotes peaceful human interaction, if such traits exist, will be deemed aspects of our human nature that incline us to act *morally*. It is important to emphasize that whether or not the resulting actions are performed as an end in themselves, or a means to an end, is not a concern in this case. For the purpose of this thesis, it is sufficient to demonstrate that such traits exist, or at least that it is reasonable to assume that they might exist, given our current scientific knowledge. Also, it should be noted that by limiting the set to only those traits which *directly* assist us in moral behaviour, the intention is to eliminate the potentially large set of traits which may facilitate peaceful human interaction, but do so only indirectly, or have only a trivial role: not to set out any sort of strict requirements for direct causal links.

Acts of altruism can be assumed to represent moral actions, as defined above. The term "altruism" can be especially problematic in philosophical discussions, however, so it is necessary to clarify how it will be used in this thesis, as well. Much of the controversy surrounding the definition and proper use of "altruism" can be avoided in this project, however.[1] This is because it stems from the need to identify the intentions motivating acts which may be deemed altruistic. The fact that our interests lie only in identifying innate

[1] See Sober and Wilson, 1998, for a thorough discussion of the issues involved.

behavioural traits related to living a moral life, eliminates the concern for specifying intentions on the part of individual human beings.

An innate behavioural trait will be considered an altruistic trait, in general, if it inclines us to, or directly assists us in, behaviour which in some way benefits other humans. Within the set of such behaviours, we can distinguish those that do so in a manner which seeks reciprocation from those which do not. The former will be labeled traits for "reciprocal altruism" and the latter will be labeled traits for "pure altruism." Although it is necessary to make this distinction for the sake of clarity, it should be noted that, for the purpose of this thesis, both types of altruistic traits will be considered moral traits, because they both facilitate peaceful coexistence.

Lastly, I will use the term "moral phenotypes" to describe all such morally relevant traits, in order to distinguish them from those learned or habituated through human interaction with the environment: a phenotype being the physical or behavioural manifestation of a specific genotype. Two types of moral phenotypes will be identified: motivational traits, those which appear to generate an innate desire to behave in some morally relevant manner; and facilitative traits, or cognitive tools, which appear to assist specifically with evaluating moral issues.

1.2 Outline

This thesis can be viewed as having three major sections: the first section, and the bulk of the thesis, is focused on demonstrating why it is reasonable to posit a conception of human nature that is more conducive to leading a moral life than that assumed in the

dominant moral view; the second section provides an evaluation of the scientific moral view; and the third section provides conclusions and implications for moral philosophy.

The first section includes Chapters 2 through 4. In Chapter 2, I provide a brief outline of our current understanding of evolutionary theory and demonstrate that the selfish gene theory of evolution does not preclude the existence of moral phenotypes. I then outline evidence for the first two of three moral phenotypes that will be discussed in this thesis: the first motivates altruistic behaviour; and the second facilitates resolving certain types of moral dilemmas. In Chapter 3, I outline evidence for the third moral phenotype, which facilitates social exchange reasoning. The case for a mental module for social exchange reasoning is the best documented, most extensively tested, and most discussed of the three posited moral phenotypes. In this chapter, I lay out the general theory of mental modules, an understanding of which is necessary to properly account for the social exchange module. This is followed by an evaluation of the merits and weaknesses of the claim for such a module. In chapter 4, I address the issue of immoral behaviour, the prevalence of which can be raised as a general counter-argument to many of the points raised in this section and in defence of the dominant moral view of human nature.

The second section is covered in Chapter 5 and considers Marc Hauser's more holistic explanation for much of the empirical evidence discussed in the first section: that we have an innate moral faculty operating in much the same way as our linguistic faculty. I find this theory to be lacking in two important respects. First, the analogy to the linguistic faculty is weak. Second, Hauser does not adequately account for the role of our analytic intellect,

our free rational capacity, which I argue is necessary for any comprehensive account of moral decision making.

The third section is covered in Chapter 6 and outlines the overall conclusions of the investigation and the implications for moral philosophy. I argue that it is reasonable to posit that we have at least three moral phenotypes, and that they represent a meaningful part of our human nature. I argue that such learning represents grounds for considering a new conception of human nature: one that is more consistent with leading a moral life than that employed by the dominant moral view. From this, I argue that moral theories within the dominant moral view should be reviewed in order to determine what implications, if any, such a revision might have on the conclusions they draw for morality. Lastly, I argue that there are a number of constraints on the use of this new learning in normative moral theory and show how these undermine current claims for a moral faculty.

Chapter 2
Dispelling the Savage Notion

In 1651, Thomas Hobbes put in writing what many had believed for some time: that humans are naturally selfish and brutish.[1] A great deal has been learned about our human nature in the intervening four centuries, but until recently, very little empirical evidence has been available to either support or refute Hobbes' claim. Our rapidly increasing understanding of evolution, and the impact of genetic variation on human behaviour, over the past few decades, has provided a wealth of new information upon which conclusions about our human nature can be drawn. Some of this information can be interpreted in a way which suggests that our genes should actually 'make' us behave selfishly towards other humans, or at least incline us toward immoral behaviour. In this chapter, I will show that, even though we may be the product of selfish genes, it is reasonable to expect such genes to manifest at least some meaningfully unselfish behavioural phenotypes. Then I will describe two cases where empirical evidence appears to show that we do indeed have specific biological traits which motivate and facilitate moral behaviour.

2.1 Evolutionary Theory

Is it actually possible for our genes to "make" us moral, or at least to promote moral behaviour among humans? We begin to answer this question by confirming some critical

[1] Hobbes, 1651 (1962)

aspects of our current understanding of evolution as it relates to our genes' impact on behaviour in general.

There are two distinct evolutionary processes that matter to our discussion: the random mutation of genes; and the process of natural selection. Random mutation can occur during the reproductive process and result in minor and even quite major variances in the physical and behavioural characteristics of offspring. Some of these variances impact the reproductive success, or fitness, of the offspring. Natural selection, a non-random process, will often drive these random variances to either dominance or extinction over time, depending on whether they increase or decrease the organism's fitness. Over many generations, a series of genetic variations (genotypes) can manifest a collection of physiological and behavioural traits (phenotypes) which increase the chances of the organism thriving over the long term in its particular environment. The collection of such traits is called an adaptation. This general outline of how evolution works has been widely accepted since the early 20th century, when our understanding of genes helped to more fully explain Darwin's original evolutionary theory.

In his 1976 book, *The Selfish Gene*, Richard Dawkins presented some insights on the evolutionary process outlined above which, among other things, had a dramatic impact on our understanding of the causes of innate behavioural traits. Dawkins made the (now quite obvious) observation that, at its core, evolution is all about the long term reproduction of genes; and, as such, the behaviour of the organisms which the genes create can only be fully understood as a function of the purely selfish "mandate" of their genes. In other words,

genes are king and humans are nothing more than very elaborate survival vehicles, "designed by" our selfish genes to give them the best chance of reproducing indefinitely.[1]

One of the implications of this selfish gene view on our understanding of innate behavioural traits is that it appears to undermine the notion of group selection in evolution. Group selection suggests that a genotype can become common within a population because it improves the fitness of the group as a whole, rather than merely an individual. In this way, group selection appeared to provide scientific support for the idea that humans might possess innate instincts, or predispositions, to act altruistically.[2] Acceptance of the alternative selfish gene theory of evolution, then, would appear to support the dominant moral view of human nature.

According to Dawkins, the selfish gene view should not be construed as indicating that humans are incapable of being altruistic, or following a moral code. He does make the point, however, that if we are looking "to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature."[3] The selfish gene theory allows for pure altruism in at least some species, however. In social insects, such as ants, for instance, the worker ants appear to be supreme altruists, risking and often losing their lives by the thousands in defense of the colony, and more especially the colony's queen. A look at the genetic make up of the members of an ant

---

[1] Dawkins, 1976

[2] See Wynne-Edwards, 1986, for a full explanation of the case for group selection.

[3] Dawkins, 1976, page 3

colony reveals that this apparent altruism of the survival vehicle is in fact a behavioural phenotype, or manifestation of the selfish instincts of the survival vehicle's genes. Male worker ants carry genes which are identical to their queen. As she can reproduce males without the need to fertilize her eggs with male sperm, their genes are a complete replication of the queen's, rather than the usual 50% replication found in a typical parent/offspring relationship. Because the males are not able to reproduce and the queen is, saving her life is worth the loss of many males, in evolutionary terms.[1]

Once their genetic relatedness is understood, the behaviour of these insects can be seen as supporting the selfish gene theory. It predicts that the well-being of any individual survival vehicle among two or more vehicles carrying identical DNA should be considered at least equally important to each vehicle. The question is: does this hypothesis hold for humans, and if so, how and to what extent is such kin altruism manifest in our human nature? The exact duplication of DNA occurs in humans only in the case of identical twins. As identical twins are relatively uncommon in our species, it is difficult to conduct large-scale studies to answer questions of this sort with certainty.

There is some recent directional learning, however, which indicates the presence of greater altruistic tendencies among human identical twins, just as the selfish gene theory predicts for those with matching DNA. Longitudinal studies of twins, who were raised apart and then reunited, show that identical twins tend to have: 1) much higher closeness and

---

[1] Dawkins, 1976, page 176

familiarity ratings than their fraternal counterparts; and 2) higher positivity and lower negativity ratings than fraternal twins.[1] Although neither of the traits identified represent the sort of pure altruism which social insects display, their presence does provide directional support for the claim that gene relatedness may trigger innate behavioural traits, or instincts, which motivate altruistic behaviour among humans.

Looking more broadly at human behaviour, there appears to be ample evidence of kin altruism among those with much less gene relatedness than identical twins. The tendency to be altruistic toward members of one's family is well documented throughout human history. This behavioural trait is also well explained by the selfish gene model of evolution. If we consider that our parents, siblings and children are likely to be carrying 50% of our unique genes, we can see why a selfish gene might actually "want" its particular survival vehicle to act in an altruistic way toward these individuals. Of course we share most of our genes with all of our conspecifics, and as such, might have at least some interest in the fate of other non-kin individuals. However, it is our unique genes, those that make us the individuals that we are, rather than the species to which we belong, which appear to play such a big role in our behaviour.

Dawkins does not suggest that humans are genetically "programmed" with the ability to determine and act on specific levels of genetic commonality. He does show, however, that a gene which promotes suicidal altruism when saving the lives of at least three close relatives

[1] Segal et al, 2006

(parent, child, or sibling) can be successful in evolving over time. This act would deliver something on the order of a 50% increase in the individual's future gene pool, assuming that the three others were indeed in life threatening peril.

Although few of us would consider rationalizing our actions in such a way, general observation of human behaviour suggests that we have at least some inclination toward this type of kin altruism in our genetic make up. (That there are also examples of close kin treating each other poorly is acknowledged and discussed in chapter 4.) Of course, the fact that we tend to treat our family well, generally, might also be accounted for by purely environmental factors, such as the greater familiarity and history of mutual support, which is often in evidence.

Another recent study provides directional confirmation of a kin altruism bias among humans. It shows that altruistic behaviour is stronger among close kin than among others with whom we may have shared the same family environment since early childhood. Where genetic relatedness can be determined, by direct evidence of the same mother, versus only indirect evidence of coexistence in the same family, those with the maternal evidence display stronger dispositional and behavioural tendencies of altruism.[1]

This evidence does not represent proof that humans are strongly predisposed by kin altruism behavioural traits. However, given that selfish gene theory does account for kin altruism, that there is some empirical evidence to support its existence in humans, and that

[1] Lieberman et al 2007 and 2003

there is a great deal of evidence from common experience, it is reasonable to posit that humans have at least some innate kin altruism behavioural traits.

Recall, however, that the purpose of this particular section is not to prove that any specific traits exist, but to show that our current understanding of biological evolution does not *preclude* their existence. Based on the evidence discussed here, it is reasonable to conclude that a purely selfish gene might actually produce a highly altruistic survival vehicle, if that happens to fulfill its self-serving mission. As part of fulfilling that mission, the altruistic trait would be passed on via the gene to the next generation, even though altruism in general appears to reduce the evolutionary fitness of an actor. Over time, if this trait reliably enhanced survival prospects for the gene, natural selection would result in widespread distribution among the population. This shows that it is possible that the current human genome may manifest at least some moral phenotypes.

2.2 Innate Altruism Motivation

In this section, I will show two specific cases of what appear to be phenotypes motivating humans to altruistic behaviour. In the first case, anthropologist James Rilling led two studies which appear to show that our brains are "programmed" to generate feelings of reward when we perform actions of reciprocal altruism. Using functional magnetic resonance imaging (fMRI) scans and personal interviews, this research evaluated test subjects while they were playing an iterated game of the prisoner's dilemma. This particular game is well suited to modeling social relationships, and for examining reciprocal altruism, specifically. The prisoner's dilemma involves two players who have the option of cooperating with each

other for some moderate mutual reward, or defecting and potentially receiving a much greater reward. Rilling's game was structured to involve financial gain based on the total payoff accumulated over forty games, played with the same other player, with whom test subjects had no contact. Incentives were established to provide the greatest financial reward if the player defected while their counterpart cooperated, a more moderate reward if both players cooperated, a significant penalty if the player cooperated and their counterpart defected, and a moderate penalty if both players defected.

Respondents claimed that they felt a greater sense of "personal satisfaction" from the less financially rewarding, mutually cooperative outcome, than the more financially rewarding outcome, whereby they defect and the other player cooperates. More importantly, for our purposes, fMRI results showed that cooperative behaviour "was associated with consistent activation in brain areas that have been linked with reward processing: nucleus accumbens, the caudate nucleus, ventromedial frontal/orbitofrontal cortex, and rostral anterior cingulated cortex."[1]

This learning suggests that it is possible that our brains are "hardwired" in a way which motivates us to seek cooperative interactions, which are examples of reciprocal altruism. Rilling summarizes learning from the study with the following:

> Cooperative social interactions with nonkin are pervasive in all human societies and generally emerge from relationships based on reciprocal altruism. Such relationships arguably lay the foundation for the

[1] Rilling et al 2002

> interdependence upon which societal division of labor is based. We have identified a pattern of neural activation that may be involved in sustaining cooperative social relationships, perhaps by labeling cooperative social interactions as rewarding, and/or by inhibiting the selfish impulse to accept but not reciprocate an act of altruism.[1]

The second case appears to show that our minds may be programmed to actually feel that it is better to give than to receive. More specifically this recent study, led by neurologist Jorge Moll demonstrates that our brains appear to be 'wired' in a way that makes acts of pure altruistic behaviour pleasurable. This study also employed fMRI scanning, but in this case to examine the neural firing associated with donating money to charity. Respondents were given a sum of money which they could choose to donate to certain real-life charities or keep for themselves. As well as the general finding that donating showed more positive stimulation, Moll indicates that there are additional aspects to the positive reward donating created:

> These findings indicate that donating to societal causes recruited two types of reward systems: the VTA-striatum mesolimbic network, which also was involved in pure monetary rewards, and the subgenual area, which was specific for donations . . . [2]

In other words, not only was the area of the brain normally stimulated in response to personal financial gain (the VTA-striatum mesolimbic network) stimulated more so when donating; but donating stimulated additional areas not associated with material gain. This area (the subgenual) is normally associated with social attachment and rewards of affiliation

---

[1] Rilling et al, 2002

[2] Moll et al, 2006

with others. In essence, the altruistic behaviour was providing a multiply positive stimulus to the brain.

The two cases shown here are consistent in demonstrating that there is at least some sort of neural basis for associating altruistic behaviour with feelings of reward. This does not represent conclusive evidence, however, and it does not show directly that such behaviour represents a phenotype, or manifestation of genetic programming. It is possible that such neural constructs are created by environmental forces, such as culture, and present themselves as a function of our brain's plasticity. We will discuss this issue of plasticity in some detail in chapter 3.

2.3 Innate Harm Reduction Reasoning

The second potential moral phenotype to be considered is one that appears to facilitate moral reasoning in certain types of dilemmas. The first indications of such a trait were noted by philosopher Philippa Foot, whose Aristotelian approach to ethics caused her to challenge many of the tenets of the deontological and consequentialist theories of morality.[1] Her work in the 1960's caused many to question the validity of such reason-based approaches to morality and identified what appears to be an innate principle that humans use to solve certain types of moral dilemmas: the principle of double effect. Foot's research showed that humans appear to have pre-established answers, or solutions, to certain types of moral dilemmas: those that follow the parameters of the Trolley test.

---

[1] Foot, 1967

To understand the Trolley test, consider the following two scenarios about which study respondents are asked to make decisions:

A) A trolley car is speeding down the tracks. Upon seeing five workers on the tracks ahead, the driver applies the brakes, which fail and the driver then passes out. The banks are so steep that the five workers will not be able to get off the track in time and so will certainly be killed. Fortunately, Salinee is standing near the tracks, next to a switch that can divert the train onto a side-track. Unfortunately, there is one worker on this track, who will certainly be killed if the train is so diverted. Is it morally permissible for Salinee to throw the switch?

B) A trolley car is speeding down the tracks. Upon seeing five workers on the tracks ahead, the driver applies the brakes, which fail and the driver then passes out. The banks are so steep that the five workers will not be able to get off the track in time and so will certainly be killed. Juan is standing on a bridge above the tracks. There is a large man standing next to Juan; large enough that, if he were to fall onto the tracks the train would stop when it hit him, though the large man would certainly die. Juan has the ability to push the man onto the tracks and save the five workers. Is it permissible for Juan to push the man?

Extensive testing shows that nearly 90% of those taking this type of test respond that in scenarios matching A it is permissible to act, and in those matching B it is not permissible to act. In other words, it is permissible to redirect the train resulting in the death of one individual on the side-track, rather than five on the main track, but it is not permissible to

push one individual onto the tracks to save the five workers. Most respondents answer with little hesitation, yet have a great deal of difficulty justifying there responses.[1]

The problem people have explaining their response appears to stem from the fact that it is not justifiable using either deontological rules or consequentialist reasoning; the type of reasoning that we appear to typically employ when consciously reasoning through such dilemmas. A deontologist might say that one should follow the rule: all acts of killing other humans are immoral. This does not allow for the difference in response to A and B, however. A consequentialist would typically reason that an act is permissible if the consequences are better than any alternative acts. This would also not justify the different responses found.[2]

In fact, the differences can be explained via the principle of double effect. This states that it is permissible to cause harm as a by-product of a greater good, but it is not permissible to cause harm as a means to a greater good.[3] Assuming that the principle does actually come into play in our decisions, it raises the question of how we subconsciously make such a judgement, as few can consciously justify it when tested. One possibility is that there is an innate moral heuristic at work.

Testing on children seems to corroborate this suggestion. Early, small scale testing by philosopher John Mikhail showed that children between the ages of 8 to 12 followed the

---

[1] Hauser, page 128

[2] Mikhail, 2000, page 96

[3] Hauser 2006, page 33

same response patterns as adults.[1] More recent and extensive testing by Hauser supports these results, suggesting that these patterns are not the result of learned rules.[2] Even without this corroborating, early development learning, the presence of an innate behavioural trait, using a moral heuristic, appears to be the best explanation for the adult response pattern. The only apparent alternative - that the responses are learned - is improbable, in part because few have had exposure to such unusual dilemmas, but also, because, if the responses represent a learned solution to such scenarios it is difficult to explain why so few people can justify their response with reasons.

Hauser is expanding the testing to determine if the response patterns hold across a wide array of cultures. He has translated the testing material into Hebrew, Arabic, Indonesian, Chinese and Spanish and begun testing small-scale hunter-gatherer societies. Initial results indicate that the response pattern is the same.

2.4 Chapter Summary

In this chapter I have shown that even the selfish gene view of evolutionary theory supports the possibility of moral phenotypes manifesting in humans. Also, two examples of potential moral phenotypes were outlined. First, a phenotype motivating altruistic behaviour was evidenced by two separate studies indicating the presence of neural constructs for such motivation. Second, a phenotype facilitating harm reduction was shown to be supported by

---

[1] Mikhail, 2000

[2] Hauser, 2006, page 128

extensive testing which indicates that humans possess an innate heuristic reasoning mechanism which follows the principle of the double effect. Although neither of these examples are conclusive, and there is still a great deal to be learned before we can assert their existence as fact, there is enough evidence to warrant positing such traits and considering their implications to the dominant moral view of human nature. The following chapter outlines another posited moral phenotype, the social exchange reasoning module, which has been more extensively tested and critiqued than the two discussed in this chapter.

Chapter 3
Innate Social Exchange Reasoning

In this chapter I will first outline the case for the existence of a specialized neural mechanism for social exchange reasoning broadly, and then provide specifics for each element of the case. The foundation point is that reasoning tests, such as the Wason selection task test (Wason Task), where respondents are asked to determine if a conditional rule has been violated, reveal a marked improvement in scores for social exchange problems, versus other similar problems.[1] The fact that respondents perform much better when solving social exchange versions of the Wason Task test appears to be explained only by the existence of an innate ability to deal with this specific type of problem. There is reason to believe that such a trait could have been adapted for in humans, early in our development, because cooperation in general and social exchange specifically appear to be critical to our success, as I will show.

Further, these traits appear at an early age, well before they could have been learned, and develop consistently across cultures. Two alternative explanations, which have been proposed, do not properly account for the testing differences noted. Specifically, the social exchange advantage cannot be accounted for by the permission schema hypothesis and it is also not a function of social exchange problems being more familiar to test participants.

The chapter is divided into three main sections. In the first section (3.1), I will ensure that there is common ground for the discussion, by defining mental modules and outline the

[1] Barkow at al, 1992; Cosmides and Tooby, 2004 and 2005

case for the existence of mental modules in general. In section 3.2, I discuss the specific evidence for a social exchange reasoning module. Following this, in section 3.3, I lay out three main arguments that call into question the existence of a social exchange reasoning module, and show how each of them fails this task. The first argument questions the evolutionary foundation of modularity in general. It will be discussed in the context presented by Robert Richardson, in his 2007 paper "The Adaptive Program of Evolutionary Psychology." The second argument comes to us via philosopher Stephen Quartz and biologist Terrence Sejnowski, via their 2003 book *Liars, Lovers, and Heroes*. They assert that dramatic climate change over the Pleistocene period would not have provided the stability which evolutionary psychologists claim created our modular minds. In the third case, I will outline philosopher David Buller's argument, from his 2005 book *Adapting Minds*, that the Wason Task experiments used by proponents of the social exchange module do not actually test for social exchange reasoning.

3.1 The Case for Modularity

Our minds are being unlocked, quite literally piece by piece, to a degree and at a speed that only twenty years ago would have seemed impossible. As a result of a variety of new technologies, such as fMRI, which allows us to 'see' the brain in action, ever more sophisticated and precise psychological testing, and wide interest in what many see as the last great frontier of biology, we know a great deal more today than we did even five years ago. For some, this new learning provides compelling evidence to support the assertion that much of our mental functioning is carried out by specialized faculties, or mental modules.

Proponents of this theory believe that these individual modules represent evolutionary adaptations selected for their ability to improve our survival rate in earlier stages of human development. The alternative reading of this new information about the workings of the mind focuses on the plasticity of the brain and asserts that the mind is primarily a learning machine. Many who follow this line of thinking argue that there can be few, if any, fixed modules, because the "hardware" of the brain is in flux and its status at any given point in time is the product of a fundamental interaction between our brain and our environment.

In this section, I seek to show that there is merit to both views, and hence, show that there is good reason to believe that mental modules do exist for at least some of our neural functioning. Leda Cosmides and John Tooby describe their understanding of the mind thusly:

> The human mind is the most complex natural phenomenon humans have yet encountered, and Darwin's gift to those who wish to understand it is a knowledge of the process that created it and gave it its distinctive organization: evolution. Because we know that the human mind is the product of the evolutionary process, we know something vitally illuminating: that, aside from those properties acquired by chance, the mind consists of a set of adaptations, designed to solve the long-standing adaptive problems humans encountered as hunter-gatherers[1]

The term 'mental module' is used in a variety of ways by different authors and, although they are generally similar, there is no commonly accepted definition at this point. I will define mental modules as operating structures in the brain which deal with particular

[1] Barkow et al, 1992, page 163

cognitive capabilities that are functionally specific. In other words, they are systems, or mechanisms in our brain which operate in specialized ways on specific types of mental functioning. It is important to note that mental modules are not necessarily anatomical: modules do not necessarily relate directly to a particular section of the brain. In fact, it appears that they rarely do, often involving neuronal activity in a variety of places in the brain. A module should be thought of as representing a processing construct that incorporates the rough equivalent of hardware, set to provide the basic structures, and software incorporating an innate learning mechanism, which allows the structure to modify itself to varying degrees.

There are two foundational points which need to be understood before it is possible to consider confirming the existence of such adaptations as mental modules. First, it needs to be clear that a significant portion of our cognitive capacity is genetically "programmed", either set structurally at birth, or set to structure at some point later in life. This is important because, if the brain is wholly plastic, then mental modules could not exist as they are defined in this paper. Of course, a type of modularity may still exist in a highly plastic brain development model, but the implications of this are not clear at this point and are also not particularly relevant to this specific project. At the same time, proving that there is a significant genetic impact on brain structure does not, by itself, show that modules exist: only that they are possible. Second, given that there is ample evidence for some level of plasticity, this must be accounted for in the overall model which incorporates mental modules.

3.1.1 Evidence of Modularity

According to Cosmides and Tooby, there is ample evidence to show that a great deal of our mental architecture is pre-programmed, or content-independent. Before detailing this evidence, I will outline the four general arguments they make for such a structure.[1]

Firstly, given that there are innumerable ways of thinking and acting, and that the vast majority of these ways will lead to a short life, an architecture which specifies a few successful ways of thinking and acting would very likely evolve.

The second argument, which is essentially a different aspect of the first, relates to the potential for overloading a system that is not innately structured to avoid the combinatorial explosion problem. The combinatorial explosion refers to the exponential increase in options that must be considered in a system which seeks to identify and evaluate all possible options in a given situation, without some sort of limiting or restraining factors. This further supports the need for at least some level of pre-programming in any intelligent system.

The third argument is that the need for a high level of pre-programming in the mind is not just related to higher level decision making: even very basic human acts often require an immense amount of neural work. The simple act of lifting the hand, or straightening the back, involves so many individual signals, commands and feedback messages, that it is not

[1] Barkow et al, 1992

possible to imagine such complex activity taking place without meaningful levels of pre-programming.[1]

Lastly, Cosmides and Tooby ask us to consider the framing effect, which highlights the need for a decision maker to have some predefined set of parameters. These parameters need to establish what goals or desires are being sought and how, in general, to go about making decisions about the best way to achieve them. We can understand this need very well when we look at work in the area of artificial intelligence. Learning in this area shows that robots must have these framing aspects integrated into their decision making system in order to complete even the simplest tasks.[2]

Overall, Cosmides and Tooby argue that human minds must have a "universal design", which is equipped with 'specialized mechanisms that "know" many things about humans, social relations, emotions and facial expressions, the meaning of situations to others, the underlying organization of contingent social actions such as threats and exchanges, language, motivation, and so on.'[3] They believe that a system which integrates a variety of genetically determined, domain-specific sub-systems will do a much better job of optimizing belief generation and action taking and, hence, would be selected for over time.

Steven Pinker provides a number of specific examples which he claims show, not only that genetic pre-programming exists, but that it represents an important part of our

---

[1] Barkow et al, 1992, page 5

[2] Barkow et al, 1992, page 106

[3] Barkow et al, 1992, page 89

mental capacity. He points to evidence for the existence of different domains of knowledge utilizing different types of processes, or modes of representation, for each domain. The concept of words and rules for language and the concept of an enduring object for understanding the physical world are two examples of this. "Developmental psychology has shown that these distinct modes of interpreting experience come on early in life: infants have a basic grasp of objects, faces, tools, language and other domains of human cognition."[1] This suggests that such mental processing is genetically programmed at birth, rather than through any significant environmental interaction.

Other evidence supporting the same point is that a great deal of the variance in psychological traits appears to stem from genetic factors, rather than environmental factors. Pinker draws these conclusions from longitudinal studies of siblings, which show that "identical twins are more similar than fraternal twins, and biological siblings are more similar than adoptive siblings, whether reared together or apart." Further, "both personality and intelligence show few or no effects of children's particular home environments within their culture: children reared in the same family are similar mainly because of their shared genes."[2] Other implications of this learning will be discussed in the following section.

---

[1] Pinker, 2002, page 102

[2] Pinker, 2002, page 102

3.1.2 Allowing for Plasticity

The second point that was raised at the beginning of this section, on the case for modularity in general, is that we need to confirm that there is a place for the plasticity concept within the concept of a genetically pre-programmed mind. Although it has been shown, so far, that there is good reason to believe genetic programming plays a meaningful role in our brain's architecture, there is ample evidence that a high degree of plasticity is at work, as well. Pinker tells us that "neural plasticity is not a magical protean power of the brain but a set of tools that help turn megabytes of genome into terabytes of brain, that make sensory cortex dovetail with its input, and that implement the process called learning."[1] This plasticity does not undermine, in any way, the fact that we are genetically programmed for a great deal of the behaviours we display. "Humans behave flexibly because they are programmed: their minds are packed with combinatorial software that can generate an unlimited set of thoughts and behaviour."[2]

3.1.2.1 The Nature / Nurture Debate Quantified

Some of the strongest confirming evidence for Pinker's assertion comes from longitudinal studies of siblings. There is now a sizable body of such research, which allows psychologists and cognitive scientists to draw much clearer pictures of the effects of nature

[1] Pinker, 2002, page 100

[2] Pinker, 2002, page 40

and nurture on our behaviour. Psychologist Eric Turkheimer's studies of this research led him to conclusions summarized in what he calls the three laws of behavioural genetics:[1]

1. All human behaviour traits are heritable.
2. The effect of being raised in the same family is smaller than the effect of the genes.
3. A substantial portion of the variation in complex human behavioural traits is not accounted for by the effects of genes.

With these conclusions established, Turkheimer suggests that the nature/nurture debate may now be over. This appears to be an overstatement, as there are still many unanswered questions within the general debate. Pinker argues, however, that, as it relates to understanding "what makes people within the mainstream of society different from one another, whether they are smarter, or duller, nicer or nastier, bolder or shyer – the nature-nurture debate, as it has been played out for millennia, really is over, or ought to be."[2]

As referenced earlier, Pinker goes further to quantify Turkheimer's three laws by specifying the degree to which each aspect impacts our behaviour and intelligence: genes 40-50 percent; shared environment 0 – 10 percent; unique environment 50 percent.[3]

How can Pinker be so bold as to put figures to these affects? The answer is by working through the same large-scale studies that Turkheimer used to develop his laws, but attacking it with the goal of quantification. These studies compare IQ and personality test

---

[1] Turkheimer, 2000

[2] Pinker, 2002, page 372

[3] Pinker, 2002, page 381

results among a variety of sample populations. Describing this work in any detail is not necessary for this project, but it is worth noting several points. First, it can be concluded that the effects of genes are in the order of 50 percent because that is how similar identical twins are regardless of shared or separate upbringing. Second, it is estimated that parenting/shared environment affects less than 10% because; 1) the similarities among adult siblings does not change with shared or separate upbringing; 2) the similarities among adopted siblings with shared upbringing is no different than any two people chosen randomly; and 3) the similarities among identical twins are no different than the degree to which their genes are similar.[1] Though the precise numbers are not critical to the argument for the existence of mental modules, understanding that both nature and nurture have significant roles to play in shaping our mental processes, is important, and Pinker and Turkheimer's studies offer important corroborating evidence.

3.1.2.2 Dual Processing Theory

Pinker's general hypothesis for a mix of pre-programmed and more plastic mechanisms in the brain supports dual processing theories which have gained some acceptance recently. Consistent patterns of strengths and weaknesses in testing for problem-solving skills suggest that there are two very different processing systems at play. According to psychologist Keith Stanovich, the one system should be considered our "heuristic intelligence". It is an "associative system with computational mechanisms that reflect

[1] Plomin et al, 2001

similarity and temporal contiguity." The other, which he refers to as the "analytic intelligence", is a "rule-based system that operates on symbolic structures having logical content."[1] It appears that, while both systems are always in play, one will tend to dominate in situations where its specialty is most valued. Although the mechanism which is dominant at any given point seems to prevail, the subservient system still operates in its own way on the same problem. One example of this manifested in our behaviour is when people determine that the right answer to a question must be X, and yet something inside seems to be telling them it is really Y. Stanovich relays how this is revealed in many reasoning test studies. One example of the type of problems being studied is known as the Linda Problem. In this study, test subjects are told the following:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
> a. Linda is a teacher in an elementary school
> b. Linda works in a bookstore and takes yoga classes
> c. Linda is active in the feminist movement
> d. Linda is a psychiatric social worker
> e. Linda is a member of the League of Women Voters
> f. Linda is a bank teller
> g. Linda is an insurance salesperson
> h. Linda is a bank teller and is active in the feminist movement [2]

Stanovich reports that 85% of respondents assign higher probabilities to Linda being a bank teller who is in the feminist movement than they do to Linda being a bank teller.[1]

---

[1] Stanovich, 1999, page 102

[2] Stanovich, 2004, page 69

Yet, as the first case is a subset of the second, it must have a lower probability. Stanovich has found that many people tend to give this non-Bayesian answer to the problem (the one that is probabilistically incorrect) and yet report that they somehow know it is not right, but can't quite explain why. In his view, this results from the two different systems working on the same problem and, because of their different reasoning approaches, coming up with different answers.

Further understanding of these systems provides more support for a combined genetic/plastic mental architecture. There are indications that the heuristic intelligence is an older system in evolutionary terms and works subconsciously. The analytic intelligence appears to be much newer and takes more conscious effort.[2] Also, the heuristic intelligence is believed to be "largely innate", and "shaped by natural selection to do a good job on problems like those that would have been important to our hominid forebears." The analytic intelligence appears to be primarily a "maximizer of personal utility", is "more influenced by culture and formal education" and is "often more adept at dealing with many of the problems posed by a modern, technologically advanced, and highly bureaucratized society."[3]

I will return to this issue and elaborate on a number of important aspects in Chapter 5, but for now it should be noted that the two different processing systems should not be linked to the modular and non-modular aspects of our mind in any direct sense. The value of

---

[1] Stanovich, 2003, page 72

[2] Samuels and Stich, 2004

[3] Samuels and Stich, 2004, page 297

recognizing these systems at this point is that their existence supports the assertion that there is a meaningful amount of pre-programming at work in our minds. Also, it adds to the evidence for Pinker's assertion that a variety of systems are working in an integrated fashion.

3.2 The Case for a Social Exchange Reasoning Module

In the previous section, I demonstrated that it is reasonable to conclude that humans have at least some degree of modularity at work in their mental functioning. In this section, I will show why there is good reason to believe that modularity is at work in at least one specific area of our social interactions: that for social exchange reasoning. Leda Cosmides, one of the early proponents of the modularity hypothesis, writes that "(t)he human brain is not just better than that of other animals, it is different. And it is different in a fascinating way: it is equipped with special faculties to enable it to exploit reciprocity, to trade favours and to reap the benefits of social living."[1]

Psychologist Gerd Gigerenzer has conducted and analyzed a great deal of psychological research using game theory and logic puzzles, such as the Wason Task test, in order to better understand how we reason. From his studies, he concludes that mental modules do exist for a host of special functions. In the case of the social exchange reasoning module, which Gigerenzer believes operates in all normal functioning human beings, the following mechanisms appear to be working in tandem: "perceptual machinery to recognize different individuals; a long-term memory that stores the history of past exchanges with other

[1] Barkow et al, 1992

individuals in order to know when to cooperate, when to defect and when to punish for defection; knowledge about what constitutes a benefit and what a cost for oneself; and emotional reactions such as anger that signal to others that one will go ruthlessly after cheaters."[1]

Further studies of social exchange reasoning show that the module may be more precisely specified as a cheater detection module. This more refined testing indicates that it is not enough to just perceive that a situation involves a rule related to social exchange: specific cueing for the possibility of being cheated was necessary to engage the cognitive mechanism.[2]

3.2.1 A Theory of Mind

A related modular example can be found working on our theory of mind. This module develops as early as two years old and allows children to make inferences about mental entities that are different than physical entities. This is a foundational point for the existence of a social exchange reasoning module, in that one must first be able to understand others as having minds of their own, in order to be able to interact with them in a meaningful social context. "Between the ages of three to five, this domain-specific inferential system develops in a characteristic pattern that has been replicated cross-culturally in North

---

[1] Gigerenzer, 2000, page 234

[2] Gigerenzer and Hug, 1992

America, Europe, China, Japan and a hunter-gatherer group in Cameroon."[1] The command of this system early in a child's life and the invariability of it across a wide variety of environments suggest that it is a pre-programmed module, rather than a learned or adopted mechanism.

3.2.2 Mirror Neurons

Knowing that others have minds of their own provides a foundation upon which another element of our neuronal capabilities for socialization depend: mirror neurons. Initially discovered by Giacomo Rizzolatti, apparently quite by accident, mirror neurons activate in imitation of the actions of others.[2] In Rizzolatti's work with macaque monkeys, mirror neurons associated with certain physical actions were found to be firing in a monkey who was not performing the actions, but only observing another monkey who was performing the actions.

Subsequent work has shown that mirror neurons are at work in the human brain and play an important role in empathy. They appear to be the root of what makes us tend to smile when we see others smiling, or wince when we see others in pain, even though we experience no physical pain ourselves. More generally, mirror neurons appear to be triggered in response to one's own goal directed action and such action in others.

---

[1] Barkow et al, 1992, page 90

[2] Rizzolatti et al, 1996

Studies with autistic children, led by Mirella Dapretto in 2006, provide a dramatic demonstration of just how important these neurons are to our ability to socialize.[1] Autistic children have virtually no activity in the areas of the brain where mirror neuron activity takes place. This may be the underlying cause of the social deficits of those with autism. Dapretto's work indicates that the level of mirror neuron activity is directly correlated with the deficiencies in social performance. This learning suggests that mirror neurons are necessary, though not sufficient, for positive social interaction.

The latest studies from researchers at University of California's Center for Brain and Cognition conclude that "internal simulation mechanisms, such as the mirror neuron system, are necessary for normal development of recognition, imitation, theory of mind, empathy and language. When these mirror-neuron-based simulation mechanisms are deficient "one is left with an individual with qualitative impairments in social interaction," yet those same individuals "have no difficulty understanding rule-based accounts of the environment."[2]

This learning is supported by other findings, such as Wason Task testing results, which show a marked performance difference between social and non-social cognitive skills that otherwise appear quite similar. Of course, the specific causes of the Wason Task test differences in favour of social exchange scenarios can not be attributed to mirror neurons, because the testing does not typically involve interaction with other people: only written or verbal scenarios involving people. Mirror neurons appear to play an important role in our

[1] Dapretto et al, 2006

[2] Oberman and Ramachandran, 2007

ability to "read other people's minds", however, at least in a superficial sense of interpreting what they are doing and how they are feeling. They seem to be part of what is necessary to utilize a rule-based approach to interpreting other people, but not for using a rule-based approach to interpreting other aspects of our environment.

The above two areas of developmental learning theory are consistent with a variety of studies which support the conclusion that, by the age of three, most of us already demonstrate the same social exchange reasoning advantage that appears in testing on adults. These studies use pictures to get children to identify which character is violating a particular rule, rather than the written problem scenarios used on adults. The results show that the presence of a benefit and the potential for cheating generates markedly higher test scores; just as it does in adult testing across a wide variety of cultures.[1]

3.2.3 Familiarity and Permission Schemas

Some of Cosmides' and Tooby's latest work confirms that the difference in Wason Task results, in favour of social exchange scenarios, is not due to the greater familiarity most people have with social exchange situations: a common criticism. By adding scenarios unfamiliar to the culture of their test subjects to their study database, Cosmides and Tooby were able to show that familiarity does not result in higher performance: unfamiliar scenarios elicited high test scores. "(P)eople are just as good at detecting cheaters on culturally

[1] Cosmides and Tooby, 2004, page 1305

unfamiliar or imaginary social contracts as they are at detecting cheaters on completely familiar social contracts."[1]

The permission schema hypothesis has also been included in such testing, and results show that it does not explain Wason Task result differences for social exchange. Permission schemas represent a general set of rules incorporating a deontic conditional. This set includes the subsets of social contract rules and precaution rules and also includes a much wider variety of rules, such as rules for civil society, bureaucratic rules, etc.[2] Comparisons were made via Wason Task experiments with either a benefit involved (social contract permission rule), or an unpleasant task involved (non-social contract permission rule). Similar results from four separate studies showed that the difference in reasoning was dramatically in favour of the social contract scenarios.[3]

3.3 Arguments Against the Social Exchange Reasoning Module

In this section I will outline three major attacks on modularity and demonstrate that they fail to undermine the hypothesis that posits a mental module for social exchange reasoning. In the process of demonstrating this, I hope to convey some additional valuable insights into the general case for modularity and the social exchange reasoning module specifically.

---

[1] Cosmides and Tooby, 2004, page 1300

[2] Cheng and Holyoak, 1985

[3] Cosmides and Tooby, 2004, page 1303

3.3.1 Richardson's Adaptation Concerns

The first argument, put forward by Robert Richardson,[1] claims that the case for any type of modularity is undermined, or at least called into question, by the fact that there is insufficient evidence to conclude that any such modules are adaptations. Richardson's argument is best shown by comparing how well the modularity thesis meets Robert Brandon's five conditions for adaptive explanations, which are paraphrased as follows:[2]

1. There must be evidence that selection has occurred.
2. There must be an ecological basis for selection.
3. The differences between individuals must be heritable.
4. There must be information on the environment, the population, and the gene flow.
5. It must be possible to distinguish between primitive and derived traits.

Richardson uses language traits as the basis for making his comparison. As mentioned earlier, the case for modularity in the area of language is among the strongest of the higher level cognitive functions. Hence, this is a not uncharitable example to test. I will review Richardson's case, putting it in the context of social exchange, to the extent possible, based on the information available. This is not particularly difficult, as his evaluation of

[1] Richardson, 2007

[2] Richardson, 2007

evolutionary psychology's claim, that neural modularity represents an adaptation, is negative on just about all counts.

To meet Brandon's first condition requires, among other things, evidence that those who did not have the specific module in question died out, while those possessing it survived. There is no such evidence available. It is known, of course, that a variety of our hominid ancestors have become extinct. There is insufficient information, however, about the social factors which may have impacted their fate. As for condition two, while there is a good theoretical basis for arguing that the social exchange reasoning module may have had an ecological basis for selection, there is no direct evidence for this. Any claim for heritability, required to meet condition three, is also weakened by the fact that there is no specific evidence of variance, which results in heritability being classified as "undefined," according to Richardson. Making a case for condition four requires showing specific evidence of particular social structures, the key determining factor in a social exchange reasoning module. Although, it is possible to make certain estimations using extant cases of primitive cultures, this does not provide the sort of evidence required for an adaptation claim. Even if it were possible to know with certainty that a particular culture had not been impacted by the developed world (highly unlikely at this point), there is too little information about that particular trait's impact in the social environment during the key evolutionary period in question. As for the fifth condition, Mathew Ridley shows that social exchange, in the form of reciprocal altruism, is a trait found in closely related primates, such as

chimpanzees.[1] He also shows that goods were traded among early humans. However, this does not represent the sort of direct evidence that is required to make a clear distinction between what is primitive and what is derived.

Based on this evaluation, it is clear that the evolutionary underpinnings of the social exchange reasoning module cannot be substantiated sufficiently to label it an adaptation. Thus it should not be asserted definitively; only hypothesized. Although it is valuable to know this, and it is certainly relevant to the case, I would argue that it is not particularly damaging to the hypothesis of a social exchange reasoning module. This is because the strongest evidence for the existence of such a module stems from the marked differences in our ability to reason in social exchange situations, as shown in Wason Task experiments. The best way of explaining such a capability is that it constitutes a behavioural phenotype that facilitates cooperative interaction. Showing that the available evidence is insufficient to prove an evolutionary adaptation claim is valuable, but it is not the same as showing that it is not the case, or even just unlikely to be the case, that this capability is an evolutionary adaptation. It is also not the same as showing that there is some other, possibly more compelling explanation for the capability. As such, given the available information and considering the possible alternatives, the evolutionary adaptation hypothesis of a mental module affecting superior social exchange reasoning continues to be the best explanation.

---

[1] Ridley, 1997, page 91

3.3.2 The Strong Plasticity Claim

The second argument I will examine is made by Steven Quartz and Terence Sejnowski and represents a more direct attack on the modularity hypothesis. They claim that dramatic climate changes over much of the Pleistocene era would not have favoured the sort of rigidity that modularity entails: it would have favoured a highly plastic neural structure.[1] If sound, this argument might have a more significant impact on the modularity hypothesis than Richardson's. I will show that it is not sound, however. In fact, it may actually represent a supporting argument for the modularity hypothesis. The argument is as follows:

1. There were extreme variations in climate during the period when much of the evolutionary development which created our current behavioural traits took place.

2. Extreme variations in climate would have made it difficult for humans to survive.

3. Those who could adapt quickest and most effectively to each successive change would be most likely to have their genes carried forward.

4. The best neural architecture for achieving point three is one that is highly plastic, not one that is highly modular.

5. Therefore, extreme climate change would not have led to a modular neural structure.

[1] Quartz and Sejnowski, 2004, page 77

The mistake is in point four. What is clear is that two general types of cognitive skills would have been critical to success in a harsh environment: 1) those enabling good cooperation among conspecifics; and 2) those enabling the best exploitation of the environment for shelter and sustenance. Success in the first area would be aided by consistency, as the nature of what makes for good cooperation would not change in line with environment. Success in the second area would require a great deal more variation and ingenuity, however, as ways of providing sustenance and shelter might change dramatically with environmental change. Hence, point four should read:

> 4a. The best neural architecture for achieving point three is one that ensures the presence of a consistent set of social skills required to cooperate effectively and a more flexible set of problem solving skills required to identify and select appropriate means for responding to physical changes in the environment.

Point five might then change to read:

> 5a. Therefore, extreme climate change would have led to a neural structure combining modular elements for some skills with a more plastic, general learning mechanism.

If it is reasonable to conclude, as I contend here, that good social exchange reasoning would have been important in most environments, and especially important in harsher environments, then the fact that climate change was dramatic during much of the Pleistocene era is not inconsistent with the modularity hypothesis generally, and the social exchange module specifically. In fact, I suggest that Quartz and Sejnowski's argument provides

theoretical support for a dual processing theory of mind, such as that outlined in Section 1 of this chapter.

3.3.3 Buller on Wason Task Testing

The third argument that I will address calls into question the validity of the Wason Task results, which purport to show dissociation between social exchange reasoning and other similar types of reasoning. If Buller's argument truly does undermine the credibility of this information, it would certainly be a critical blow to the case for a social exchange reasoning module. It was in response to this dissociation in our reasoning that the notion of such a module was first posited. Hence, if there is no clear indication that we do indeed reason differently in this area, building a strong case for the existence of a specialized module is unlikely.

Buller makes two claims which might weaken the case for a social exchange reasoning module.[1] First, he identifies what he views as a meaningful difference between social exchange and social contract. He says that the theoretical support for cheater-detection is inconsistent with the stimulus that is being used in the Wason Task experiments. The need for a cheater detection module stems from the fact that game theory predicts that reciprocal altruism is an evolutionarily stable strategy only if the participants are able to detect cheating. Buller claims that Cosmides relates reciprocal altruism to what she calls social exchange, which involves two individuals exchanging for mutual benefit. The Wason Task tests are

[1] Buller, 2005, page 171

conducted using what Cosmides calls social contracts and these are defined to allow for exchange between groups, not just individuals. "Virtually all of the experimental results that purportedly provide evidence of a cheater-detection module, derive from selection tasks involving what Cosmides and Tooby call social contracts, which is a much broader class of phenomena than the class of social exchanges," states Buller.[1]

There are two problems with Buller's claim. First, he makes a major distinction out of what is actually a minor difference. The claim that group cheater-detection, versus individual cheater-detection is a critical difference is not at all clear, and Buller provides little substantiation for this assertion. The skills required for effective cheater detection in one-to-one, versus one-to-group versus group-to-group situations would, of course, vary somewhat, because of the varying number of participants. Two things can be said about this difference, if it is indeed meaningful: first, possessing the skills for each type of situation would likely be valuable for evolutionary success; second, the Wason Task experiments do, in fact, include both of the two types that Buller identifies, and the results are similar. Cosmides and Tooby report experiments using both one-to-group (e.g. "If a person is drinking beer, then he must be over 20 years old"[2]) and one-to-one (e.g. "If you borrow my car, then you have to fill up the tank with gas"[3]) scenarios. I was unable to find any examples of group-to-group scenarios in the various studies cited, but this is not of particular concern, as Buller has not

---

[1] Buller, 2005, page 171

[2] Barkow et al, 1992, page 182

[3] Cosmides and Tooby, 2005, page 597

raised this particular variation as an issue. It may be an area for future consideration, however, if there is evidence that the different scenarios require meaningfully different problem solving skills sets. At this point there appears to be no such evidence.

The second claim Buller makes against the Wason Task experiments for social exchange scenarios is that they represent a comparison of indicative versus deontic conditionals, rather than a comparison of different types of deontic conditionals. If this is correct, then the results indicate superior reasoning for deontic conditionals and say nothing about social exchange specifically.

I believe that Buller's claim is mistaken. Before explaining why, it is important to first clarify the different types of conditionals being discussed here. Indicative conditionals are what one might consider to be "ordinary" conditionals, where the conditional assertion, "if P, then Q" indicates an assertion of fact, in the sense that if P is true, then it is a fact that Q. Deontic conditionals, as the name implies, are assertions of duty, or obligation. They should be interpreted as asserting that if P is true, then you are obligated to Q.

Buller claims that all of the social exchange examples of Wason Task scenarios are deontic and that all of the control examples are indicative.[1] This is not the case, however, as psychologist Laurence Fiddick has demonstrated. Fiddick conducted Wason Task research comparing deontic conditionals with and without social contract aspects. His results confirm

[1] Buller, 2005, page 173

the social exchange module predictions.[1] An example of a social contract deontic conditional is "If you drive my car, then you must fill up the gas tank." An example of a non-social contract deontic conditional is "If you are working with dangerous gases, then you must wear a gas mask."

The results of Fiddick's testing supports those of Cosmides' and Tooby's, which do indeed make comparisons among different types of deontic rules, contrary to Buller's claim. They specified comparisons involving deontic conditionals incorporating cheater-detection scenarios with those lacking a cheater-detection element. Importantly, these comparisons show the same differential performance in favour of scenarios with cheater-detection.[2]

The research is consistent in making the case for a cheater-detection-focused, social contract module. People responding to social exchange deontic conditionals performed better than those responding to non-social exchange deontic conditionals and the same was true when cheater-detection specified deontic conditionals were compared with those without this element.

3.4 Chapter Summary

In this chapter, I have shown that there is a compelling case to be made for the existence of a social exchange reasoning module, which gives humans a marked advantage in solving problems of a social exchange nature. There is reason to believe that such a trait

---

[1] Fiddick, 2004 and 1998

[2] Cosmides and Tooby, 2005

could have been adapted for in humans, early in our development, because cooperation in general, and social exchange specifically, appear to have been critical to our success. It is not possible to prove an adaptation claim, however, as the evidence is insufficient. This is not a critical set back to the case for a social exchange reasoning module, however, as there is ample evidence demonstrating a marked advantage in reasoning ability for social exchange scenarios specifically: an ability which appears at an early age, well before it could have been learned, and one which develops consistently across cultures. No viable alternative explanations for this ability are available, and the major arguments against the social exchange reasoning module have been shown to be weak.

Chapter 4
Accounting for Immoral Behaviour

In the previous two chapters, I have identified a number of possible innate behavioural traits that support a view of human nature which is more conducive to living a moral life than that posited in the dominant moral view. Two questions arise, however, when considering the importance of such traits. First, are there also indications of innate traits for *immoral* behaviour, which were previously unknown? Second, how can the assertion, that human nature is conducive to leading a moral life, be squared with the abundance of immoral acts, which have been perpetrated over the history of human existence?

In answer to the first question: there is no new evidence of innate behavioural traits predisposing humans to immoral behaviour, within the general population, based on a search of all relevant research databases. This does not mean that no such evidence exists, however, as there are a variety of ways to search for and interpret such information. It does provide a measure of assurance that the new evidence presented here is not simply a choice selection taken from a sample which also includes contradictory evidence. The same search methods were used and the same databases were employed when seeking evidence of traits for moral and immoral behaviour.

Answering the second question is considerably more involved. Before proceeding, it will be valuable to remember that, given the definition of morality used in this thesis, immoral behaviour should be understood as only that behaviour which would generally work against creating, or maintaining peaceful coexistence with other humans, in the absence of

any other moral code established by humans. As such, many examples of behaviour commonly considered immoral by a particular culture or religion (certain sexual behaviour among consenting adults, for instance) do not constitute immoral behaviour in this discussion.

Also, it is important to acknowledge at the outset that there are many examples of immoral behaviour among humans and that there has been throughout much of our history. Accounting for all, or even most, of this behaviour, with any degree of specificity and certainty, is well beyond the scope of this thesis and, I would argue, beyond the ability of either philosophy or science, at this time. Nevertheless, three points can be made to show that the presence of immoral behaviour does not undermine the major focus of this thesis.

First, the thesis does not argue that human nature is such that immoral behaviour should not occur: only that a properly updated view of human nature will show it to be more conducive to living a moral life than that which is assumed in the dominant moral view. As such, it is only necessary to show that the new learning presented in support of the thesis is not undermined by evidence of immoral traits, or by more general indications that our human nature is so strongly inclined to immoral behaviour that the new traits identified here could play only a minor role in our overall nature.

Second, learning from the fields of anthropology and sociology can help to explain the basis for a significant amount of immoral behaviour, and do so in a manner which is generally supportive of this thesis. There is evidence that humans tend to interact cooperatively, with little conflict, when living in relatively small groups in traditional hunter-

gatherer societies.[1] Observation of extant examples of hunter-gatherer societies, such as the Waoranis of the Amazon, the Bushmen of the Kalahari in Africa and the Aborigines of Australia support this view. This appears to hold true throughout our history, as well. According to anthropologist George Pugh, "within primitive human societies, sharing is a way of life. The sharing is not limited to food, but extends to all types of resources. The practical result is that scarce resources are shared within the societies approximately in proportion to need."[2]

It appears that conflict increases when groups expand to the level that they develop sub-groups within them, or when groups begin to have regular interaction with other groups. At this point, a phenomenon known as ingroup bias becomes apparent and is thought to play a role in increasing the likelihood of conflict. Extreme examples of this type of cognitive bias can be seen in cases of genocide, such as that which took place in Nazi Germany and, more recently, in Rwanda. But this tendency to differentiate ingroup from outgroup is not necessarily a trait which promotes immoral behaviour.

Social Psychologist Jacques-Philippe Leyens and his colleagues, building on the large body of literature on ingroup bias with their own multi-cultural research, show that humans have an instinctive need to have close relations with at least a few significant others. "(I)t is evident that the existence of an ingroup is a primary necessity" among all normal functioning

---

[1] Ridley, 1997, page 198

[2] Pugh, 1978, page 64

humans.[1] Importantly, the presence of others who are not in this group (the outgroup) does not necessarily trigger negative emotions, or behaviour; simply the absence of these special emotions for the ingroup. This suggests that the motivation which generates ingroup bias is "intended" to generate cooperation and cohesiveness, rather than to generate any sort of negative bias to those outside the group.

Leyens and his colleagues designed studies to examine these differences in more detail. This work shows that people tend to associate what Leyens defines as "secondary emotions", such as "love, hope, contempt, resentment" to ingroup members, but only "primary emotions," such as "joy, surprise, fear, anger," to outgroup members.[2] These primary emotions are understood to be present in humans as well as many other animals, whereas the secondary emotions are considered to be exclusively human. Further, their work has shown that the lack of a positive bias to the outgroup stems from reservation of positive secondary emotions to the ingroup, rather than specifically negative emotions towards the outgroup.

Leyens believes that these secondary emotions may be what makes socialization a positive experience. When secondary emotions are not involved, social interaction is not necessarily a negative experience, but simply an experience which is to be judged on its outcome, rather than one which has inherent value. This view is supported by learning that infra-humanization (the attribution of less than human characteristics to other people) does

[1] Leyens et al, 2003, page 704

[2] Leyens et al, 2003, page 707

not tend to occur when members of the outgroup can be individualized as specific human beings.[1]

Viewed in this way, the phenomenon of ingroup bias can be seen as stemming from an innate behavioural trait which motivates humans to seek close, cooperative relations. Although this trait does not manifest a corresponding negative bias to others, the a priori constraint on the number of people with whom we can have close relations, combined with an expanded population, may contribute to developing an outgroup negative bias and hence to immoral behaviour.

The third point in this discussion involves learning which suggests that impairment of certain innate behavioural traits, those common in the general population, may account for at least some immoral behaviour. Recent learning about psychopaths, for instance, has been especially enlightening. Even though the incidence of this condition is very low - less than 3% of the population[2]- devoting some attention to understanding psychopathy is useful, for two reasons. First, it represents an example of how the absence of certain 'normal' genetic traits may account for some immoral actions. Second, recent learning about psychopathy is informative to more general aspects of human nature, as they relate to morality.

It will be useful to clarify what constitutes a psychopath, as there are varying interpretations within the field of psychology. According to psychologist James Blair, the

---

[1] Leyens et al, 2003

[2] Myers, 2001, page 565

most accurate and useful classification of psychopathy is that of a developmental disorder involving both affective-interpersonal components, such as a lack of empathy and guilt, and behavioural components, such as criminal activity and poor behaviour controls; with the following defining characteristics:[1]

- Excessive displays of instrumental aggression. Instrumental aggression is purposeful and goal directed, such as using aggression to heighten one's standing in a social hierarchy through bullying, for instance. Reactive aggression is the sort of aggression triggered by frustrating or threatening events, which also generate feelings of anger. High levels of reactive aggression are common among many psychological disorders, (such as depression, bipolar disorder, post traumatic stress disorder, for instance), but only psychopaths have elevated levels of instrumental aggression.

- Impaired ability to distinguish moral transgressions (harming other people) from conventional transgressions (breaking social rules).

- Impaired processing of fear-related stimuli. Psychopaths tend to exhibit very little fear and are unlikely to feel threatened. Studies indicate that fearful children develop higher levels of moral development.[2]

---

[1] Blair, 2006, page 415

[2] Blair, 2006, page 419

- Reduced comprehension of situations likely to induce guilt. Psychopaths do show appropriate comprehension of happiness, sadness, and even complex social emotions, such as embarrassment.

- Pronounced impairment in recognizing submissive cues, which normal functioning humans use to moderate anger and aggression. Submissive cues include the visual and audio signs of distress, fear and submission, such as a sad facial expression or the sound of whimpering.

It is important to emphasize that, despite these severe impairments, psychopaths appear generally lucid and rational. This description indicates that the disorder affects a specific set of neural functions, and that such neural functioning plays a role in the development of behaviour which is conducive to living a moral life. Although the precise cause of psychopathy is unknown, the problem appears to be centered on a number of functional aspects of the amygdala. Importantly, Blair's research indicates that the cause of these amygdala-based deficiencies is genetic, as opposed to environmental: adults who encounter similar amygdala damage do not show psychopathic tendencies.[1] This is also supported by developmental research which shows children with psychopathic tendencies are unable to form normal associations between certain stimuli and punishment:

> In order to achieve successful socialization, the child needs to form associations between representations of moral transgressions (acts which harm others) and the aversive 'punishment' caused by the victim's distress.

[1] Blair, 2006, page 416

> This allows the child to learn to avoid actions that will harm others. As individuals with psychopathy find the distress of the victim significantly less aversive, they are less likely to learn to avoid actions that will harm others. [1]

Psychopathy is one of a number of cognitive disorders that appear to have both genetic and environmental causes, and can account for some, probably small, percentage of the immoral behaviour that humans display. Although more needs to be learned in this area, our current understanding of the disorder is broadly supportive of the modularity hypothesis discussed in chapter 4, as evidenced by the selective set of neural functions impacted. Also, this learning suggests that good reasoning ability is not sufficient for living a moral life: certain innate behavioural traits supporting, or motivating, the moral use of good reasoning appear to be necessary. This is consistent with the notion that human nature - as manifest in the general population - may well be, not just conducive to living a moral life, but necessary.

4.1 Chapter Summary

Although it is not possible to provide a comprehensive rationalization of immoral behaviour, two contributing factors to immoral behaviour, both of which can be seen as generally consistent with this thesis, have been identified. First, the phenomenon of ingroup bias can account for many examples of immoral behaviour. Ingroup bias is best understood as a behavioural trait which encourages *positive* relations among those with whom we have regular interaction, rather than *poor* relations with outgroup members. Such a trait would

[1] Blair, 2006, page 435

likely encourage only moral behaviour among small hunter-gatherer groups. As populations expand, however, it appears that this trait may form the basis upon which immoral behaviour can develop, if other environmental factors are conducive to such behaviour. Second, new learning about psychopathy is broadly supportive of the modularity hypothesis and also indicates that certain innate behavioural traits, which support the moral use of good reasoning, appear to be necessary to lead a moral life.

Chapter 5
The Moral Faculty

As mentioned in the introductory chapter, Marc Hauser has recently made some of the strongest claims for innate morality. Specifically, he claims that we have a moral faculty operating in much the same way as the, generally accepted, linguistic faculty, which was first posited by Noam Chomsky in 1957.[1] Further, Hauser claims that we are on the verge of a science of morality, which he suggests might make it unnecessary to involve philosophy in moral discussions in the future.

In this chapter I examine his claims in more detail, review the evidence supporting them and evaluate their merit. I find his theory to be lacking in two important respects. First, the analogy to the linguistic faculty is weak. Second, he does not adequately account for the importance of the role played by our analytic intellect, our free rational capacity, which I argue, via Stanovich, is necessary for any comprehensive account of human moral decision making.

5.1 The Case for a Moral Faculty

Hauser bases his claim for a moral faculty on the ability to demonstrate that many of the elements of Chomsky's methodology for claiming a linguistic faculty can also be fulfilled for the claim of a moral faculty. Chomsky's linguistic faculty can best be understood as a set

[1] Chomsky, 1965

of principles and parameters that enable humans to take a set of words and construct an essentially infinite number of meaningful expressions. The basis upon which Chomsky claims that these competencies are innate is referred to as the poverty of the stimulus argument, which requires the following four-step process:

1. Identify a particular piece of knowledge in mature individuals.
2. Identify what kind of input is necessary or indispensable for the learner to acquire this piece of knowledge.
3. Demonstrate that this piece of knowledge is not present or available from the environment.
4. Show that the knowledge is nonetheless available and present in the child, at the earliest possible age, prior to any relevant input.[1]

In addition, Hauser identifies a broader list of expectations that should be met in order to claim the existence of a moral faculty. This framework for characterizing the moral faculty is based largely on Chomsky's poverty of the stimulus process, but is put in the context of what Hauser claims to be a Rawlsian view. According to Hauser, John Rawls and Noam Chomsky both proposed "that there may be deep similarities between language and morality, including especially our innate competencies for these two domains of knowledge."[2] Hauser's Rawlsian framework is as follows:

1. The moral faculty consists of a set of principles that guide our moral judgements but do not strictly determine how we act. The principles constitute the universal moral grammar, a signature of the species.

[1] Hauser, 2006, page 66

[2] Hauser, 2006, page 37

2. Each principle generates an automatic and rapid judgement concerning whether an act or event is morally permissible, obligatory, or forbidden.
3. The principles are inaccessible to conscious awareness.
4. The principles operate on experiences that are independent of their sensory origins, including imagined and perceived visual scenes, auditory events, and all forms of language – spoken, signed and written.
5. The principles of the universal moral grammar are innate.
6. Acquiring the native moral system is fast and effortless, requiring little to no instruction. Experience with the native morality sets a series of parameters, giving birth to a specific moral system.
7. The moral faculty constrains the range of both possible and stable ethical systems.
8. Only the principles of our universal moral grammar are uniquely human and unique to the moral faculty.
9. To function properly, the moral faculty must interface with other capacities of the mind (e.g., language, vision, memory, attention, emotion, beliefs), some unique to humans and some shared with other species.
10. Because the moral faculty relies on specialized brain systems, damage to these systems can lead to selective deficits in moral judgements. Damage to areas involved in supporting the moral faculty (e.g., emotions, memory) can lead to deficits in moral action – of what individuals actually do, as distinct from what they think someone else should, or would do.[1]

With this framework in mind, Hauser attempts to show how the available evidence fulfills the individual elements. Unfortunately, he does not do this in any sort of systematic manner, so it is necessary to make certain assumptions about which evidence fulfills which element. He does identify five broad principles, and, although he does not directly state that they are principles upon which the moral faculty operates, it appears that this is what he intends to communicate:

---

[1] Hauser, 2006, page 53

1. A principle to care for and not abuse children.

2. A prohibition of intentional battery.

3. A principle of double effect.

4. Moral conventions are more important than social conventions.

5. Intentional actions are more concerning morally than accidental actions.

Hauser shows that Principle 1 is evident across the human population, but that exceptions vary dramatically by culture. "For Americans, (infanticide) is a barbaric act, characteristic of a group that requires a moral tutorial on child care. For the Eskimos, and several other cultures, infanticide is morally permissible, and justifiable on the grounds of limited resources and other aspects of parenting and survival."[1]

Principle 2 is supported by a variety of examples from different cultures, but more specifically from a large scale international study, which Hauser is overseeing. This study employs a variety of moral dilemma scenarios, such as the Trolley test, which was discussed in chapter 2.

Principle 3 has also been addressed in the Trolley test discussion in Chapter 2. This principle states that it is permissible to cause harm as a by-product of a greater good, but it is not permissible to cause harm as a means to a greater good. Hauser provides ample evidence

[1] Hauser, 2006, page 44

to support that this principle appears to be at work in our unconscious, immediate response to a variety of moral dilemmas which are similar to the Trolley test.[1]

Principle 4 is addressed, in part, by research Hauser cites showing a marked difference in the nature and levels of emotions attributed to moral, versus amoral situations. For example, testing respondents' level of disgust at certain types of moral and amoral scenarios shows that disgust associated with moral concerns is typically much stronger and much less subject to mitigation.[2] One such test compares levels of disgust at having to lick a toilet seat with that of having to commit incest. There are many unanswered questions about the implications of this research, however. Firstly, in the case of incest, an innate aversion to it is prevalent in many animal species, so it is not clear what this indicates about human morality. The main issue, however, is that disgust appears to be automatically triggered in situations relating to food, rather than morality. The association of disgust with immoral acts appears to be an environmental attribution, according to research by anthropologist Daniel Fessler, which Hauser does acknowledge.[3] Hence, it is questionable if there is a sound basis for claiming this to be an innate principle of morality.

Principle 5, which states that intentional actions are more concerning morally than accidental actions, is demonstrated most vividly in Hauser's discussion of developmental evidence for moral traits. "Studies starting in the 1980's showed that four- to five-year-olds

---

[1] Hauser, 2006, page 128

[2] Hauser, 2006, page 196-199

[3] Hauser, 2006, page 197

generate different moral judgements when an individual carries out the same acts, with the same consequences, but with different motivations, with some evidence for a bias to consider actions worse than omissions."[1]

There are no other principles identified by Hauser, or in evidence in his work. The social exchange reasoning module, discussed in chapter 3, may be seen to represent a principle in this context, but this is not clear. Hauser does discuss this subject and is generally supportive of the theory, but he does not directly claim that it represents one of the principles of our moral faculty. I believe that our special ability to reason through social exchange situations does not represent a principle of morality, but rather a short cut for reasoning through one specific type of morally relevant issue.

How well do these five principles meet the other nine requirements Hauser has set out? The answer is not clear, because he does not attempt to rationalize the principles against the other expectations he set out in any direct way. I should emphasize also, that the five principles listed are not specifically identified as meeting the criteria for a principle in this construct, though he does refer to all of them as principles affecting morality at different points in his book.

The focus of Hauser's case for a moral faculty, and the principle that gets the most attention from Hauser throughout his book is that of the double effect, Principle 3. Hauser

[1] Hauser, 2006, page 207

does show that this principle meets all of his nine other expectations for a moral faculty, although again, this is not done in any sort of clear or direct manner.

In addition to the principles outlined here, Hauser discusses a variety of capacities and competencies that, in his view, are part of the moral faculty, or at least play an important role in the moral faculty:

- an ability to feel emotions that impact morality (such as pity, pain, empathy, pleasure)
- a predisposition to cooperate (similar to that discussed in Chapter 2)
- an ability to comprehend goal directed actions
- an empathy competence (much of which stems from the mirror neurons discussed in Chapter 3)
- an ability to pass up initial benefits for longer term rewards
- a "mind reading" ability (similar to that described in Chapter 3)
- an ability to distinguish between intended and foreseen consequences

This collection is not actually rationalized against either the Rawlsian framework for a moral faculty, or the general construct of principles and parameters. I suggest that it is fair to consider them important enabling competencies.

As for parameters, Hauser outlines a wide variety of examples of cultural variations. One example shows how the notion of shame in Western cultures is typically associated with a situation where one is exposed for violating a rule, whereas in some non-Western cultures,

such as Indonesia, shame is felt when in the presence of people of higher status.[1] Another example Hauser cites is that tests of fairness using a device called the ultimatum game are appropriate for a wide selection of cultures, including traditional tribal cultures, but not for the Au and Gnau of Papua New Guinea. For these people, accepting a gift is equivalent to assuming a debt, as all gifts offered and accepted are expected to be reciprocated with gifts of at least equal value.[2]

The above provides an outline of Hauser's case for a moral faculty. Overall, the case can be summarized as follows. He bases his claim for the moral faculty on the general construct of principles and parameters set out for the linguistic faculty, which uses the poverty of the stimulus basis for claiming certain traits to be innate. The moral faculty, broadly speaking, is seen as working along the lines of John Rawls' description of the process for moral decision making, which starts with an analysis of the actions involved, determines a judgement about whether or not such actions are obligatory, permissible, or forbidden and only then, if at all, generates emotions, which may play a role in modulating the actions taken. This Rawlsian approach is elaborated in a framework which represents a set of expectations and criteria for the moral faculty.

---

[1] Hauser, 2006, page 189

[2] Hauser, 2006, page 84

5.2 Critique of the Moral Faculty Case

I will discuss two major areas of concern with Hauser's moral faculty hypothesis. The first relates to problems with the parallel drawn between the linguistic faculty and the moral faculty. The second deals with the role of our analytic intelligence in moral decision making.

5.2.1 Moral Faculty Comparison with Linguistic Faculty

As noted earlier, Noam Chomsky is credited with first identifying the existence of what is commonly called the "linguistic faculty." As an innate, universal generative grammar, this faculty provides the set of rules by which all meaningful sentences can be constructed and understood. It consists of a finite set of principles and parameters which direct and essentially generate any and all meaningful utterances.[1] Though widely accepted by many in the field of linguistics, this faculty is still controversial in the fields of cognitive science and psychology.[2] I will proceed to evaluate Hauser's moral faculty claim, however, on the assumption that Chomsky's linguistic faculty theory is sound. This avoids a lengthy discussion of issues related to the linguistic faculty, which are not particularly relevant to this specific project. Also, given that I find Hauser's analogy to the linguistic faculty to be weak, this charitable perspective on the linguistic faculty does not undermine the critique for the purpose of this thesis.

---

[1] Chomsky, 1965

[2] See Everett, 2005, for an outline of the continuing discussion about Chomsky's linguistic faculty.

Psychologists Paul Bloom and Izzat Jarudi identify three issues with the parallel that Hauser tries to draw between his moral faculty and Chomsky's linguistic faculty: the first relates to the impact of emotions; the second and third stem from fundamental differences in the structure and use of language versus morality.[1]

Bloom and Jarudi claim that the linguistic faculty is not dependent on emotions. We can use language to express emotions, of course, but our use of language is not generated or mediated by emotions. The same cannot be said for morality, which they claim is deeply impacted by emotions. On the surface, this appears to be a valid concern, but I believe that it is actually accounted for by Hauser. Hauser acknowledges that emotions play a significant part in morality. He demonstrates that their impact is not at the judgement phase of moral decision making, however, but at the action phase. In taking this Rawlsian view of morality, Hauser is claiming that what actually happens when we are confronted with a moral problem is, first an analysis of causes and consequences, then a judgement, and only then, at the action phase, an emotional reaction. If Hauser is correct, and he does show that this appears to be true with the Trolley testing case, then it does not appear that the issue of emotions weakens the parallel between moral and linguistic faculties.

I find Bloom's and Jarudi's other two criticisms more compelling, however. They address more fundamental systemic issues. The first issue stems from the fact that languages are combinatorial symbolic systems, which involve applying a relatively short set of

[1] Bloom and Jarudi, 2006

principles (syntax) to a much larger set of words. The few principles and many words can be employed to generate an essentially infinite number of sentences. In this context, language can be seen to resemble mathematics and music: it does not, however, resemble morality. Moral decisions do involve applying principles to individual instances or scenarios, but the focus of the activity is in finding the closest match between a given principle and a particular situation. This is a fundamentally different process than applying words to rules of syntax. A considerable difference can be seen by examining the types of principles employed in each area, within this context. Language involves principles about what can be done when combining any item in the set of possible words. Morality has principles which essentially direct that, if a particular set of data matches this general type, it should be labeled thusly. Bloom and Jarudi state that morality "might be better characterized as a small list of evolved rules, perhaps simple (such as a default prohibition against intentional harm), perhaps complex (such as some version of the doctrine of double effect)."[1]

The other related concern is with the difference in degree of parametric variation between language and morality. In the case of language, the parameters vary only moderately, whereas the range is quite dramatic in the case of morals. An example of parametric variation in language is that, while all languages have verbs and they are used for the same purpose, the object to which the verb relates may be placed before or after the verb itself. Parametric variation in the moral domain shows a range so extreme that in some cases the same act interpreted in the same culture can be deemed so egregiously immoral that it

[1] Bloom and Jarudi, 2006

warrants the death penalty by some, and so minor that it is not worth acting upon by others. Bloom and Jarudi cite Hauser's example of the high incidence of honour killings in Pakistan, which are used to punish women accused of infidelity, even though many Pakistanis are strongly against such action. They make the point that, in contrast to this dramatic moral variance, all Pakistanis appear to agree happily on what constitutes a well-formed Urdu sentence.

5.2.2 Moral Faculty and the Analytic Intelligence

Although Hauser acknowledges some role for analytic intelligence in our moral decision making, I believe that he does not properly account for its importance. In fact, much of the evidence provided to demonstrate the existence of what I have called moral phenotypes (and what he variously calls evidence of a moral faculty, or principles of a moral faculty), are better accounted for via the theory of a dual processing system (as outlined in Chapter 3) than the theory of a moral faculty. Within the context of a dual process cognitive system, the moral phenotypes are examples of our heuristic intelligence; innate predispositions, or computational abilities that impact our moral decisions. What Hauser has succeeded in doing is to demonstrate that there is good reason to believe that these behavioural phenotypes exist and that they do indeed play a role in our moral decision making. His investigative work essentially stops at this point, however, and then focuses on showing that these phenotypes collectively make up a moral faculty. Once the role of the analytic intelligence is more fully examined, I believe the dual processing theory provides a much better explanation for our moral decision making.

In Stanovich's approach to dual processing he shows that there are four computational biases exhibited consistently in human decision making:

1. The tendency to contextualize a problem with as much prior knowledge as is easily accessible, even when the problem is formal and the only solution is a content-free rule.
2. The tendency to "socialize" problems, even in situations where interpersonal cues are few.
3. The tendency to see deliberative design and pattern in situations that lack intentional design and pattern.
4. The tendency toward a narrative mode of thought. [1]

Much of what Hauser has identified can be shown to fall within these computational biases, which Stanovich views collectively as "part of the automatic inferential machinery of the brain that supplements problem solving with stored declarative knowledge, linguistic information, and social knowledge."[2] Such biases represent the workings of the heuristic intelligence. The benefit of using Stanovich's model is that it represents an explanation which accounts for both this heuristic intelligence and the analytic intelligence. Hauser's view inflates this heuristic intelligence, at least certain aspects of this intelligence, and hence understates the large role played by our analytic intelligence.

5.3 Chapter Summary

Hauser's case for a moral faculty is based on the general construct of principles and parameters set out for the linguistic faculty. The moral faculty, broadly speaking, is seen as working along the lines of John Rawls' description of the process for moral decision making,

[1] Stanovich, 2004, page 113

[2] Stanovich, 2003

which starts with an analysis of the actions involved, determines a judgement about whether or not such actions are obligatory, permissible, or forbidden and only then, if at all, generates emotions, which may play a role in modulating the actions taken. Our analytic intellect also comes into play only at this level. In this chapter I show Hauser's theory to be lacking in two important respects. First, the analogy to the linguistic faculty is weak. Second, he does not adequately account for the importance of the role played by our analytic intellect, our free rational capacity, which I argue, via Stanovich, is necessary for any comprehensive account of human moral decision making.

Chapter 6
Conclusions

The goal of this thesis has been to show that there is a middle ground between the conception of human nature posited by the scientific moral view and that posited by the dominant moral view. I believe that this goal has been achieved. The thesis has shown that a view of human nature, informed by the latest scientific learning, indicates that humans possess a number of important instincts, or innate behavioural traits, which are highly conducive to living a moral life. Specifically, there is sufficient evidence to warrant positing the existence of behavioural phenotypes for 1) motivating altruistic behaviour; 2) harm reduction reasoning; and 3) social exchange reasoning. Both of the latter two traits have been shown to manifest instinctive solutions to problems of a specific nature, and to do so with a high degree of consistency and reliability across a wide range of test subjects. As such, I argue that these traits play a significant enough role in our behaviour towards others to represent a legitimate challenge to the conception of human nature used in the dominant moral view.

Hence, I argue that it is appropriate to update our conception of human nature with such scientific learning. Once developed, this new conception of human nature should be used to review all moral theory work which presupposes a particular conception of human nature. This represents a large set, including the work of Plato, Thomas Hobbes and Immanuel Kant, to name only a few of the more notable examples.

Also, contrary to the scientific moral view, this thesis suggests the role of science is limited to identifying and explaining biological traits which may either motivate or facilitate morality among humans. I find that Marc Hauser's claim of a moral faculty cannot be supported, in part because the parallels drawn with the linguistic faculty are weak, but more importantly, because it fails to properly account for the role played by our more flexible analytic intelligence.

As an initial hypothesis, Hauser's work should not be abandoned, however. There is still a great deal to learn about the workings of the human mind, and Hauser's work provides a context within which to search for and evaluate cognitive functioning as it relates to moral behaviour. To make it more plausible, however, I suggest that it must more fully account for the sort of dualistic processes which are discussed in this thesis, via Stanovich. It appears that the only way to account for the role of the analytic cognitive processing system is by acknowledging that our morality is not governed by an innate faculty, but rather stems from the decisions made by our analytic mind, which are influenced both by innate heuristic mechanisms and environmental factors. Where Hauser's work is particularly valuable for future research, is in examining what links there are between different heuristic mechanisms themselves, and how they individually and collectively interact with the analytic intellect.

Bibliography

Axelrod, R. 1984. *The Evolution of Cooperation.* New York: Basic Books.

Baker, L., Bezdjian, S., and Raine, A. 2006. Behavioral Genetics and Crime: The Science of Antisocial Behavior. *Law and Contemporary Problems*. Vol 69 Iss. 1-2. pp. 7-46

Barkow, J., Cosmides, L., and Tooby, J. 1992. *The Adapted Mind*. New York: Oxford University Press.

Blair, R. 2006. The Emergence of Psychopathy: Implications for the Neuropsychological Approach to Developmental Disorders. *Cognition*. Vol. 101, Iss. 2, pp. 414-442.

Bloom, P. and Jarudi, I. 2006. The Chomsky of Morality? *Nature*. Vol. 443, pp. 909-910.

Buller, D. 2005. *Adapting Minds*. Cambridge, Massachusetts: MIT Press.

Cheng, P. and Holyoak, K. 1985. Pragmatic Reasoning Schemas. *Cognitive Psychology, Vol.* 17, pp. 391-416.

Chomsky, N. 1986. *Knowledge of Language: Its nature, origin, and use*. New York: Praeger.

Chomsky, N. 1965. *Syntactic Structures*. The Hague: Mouton & Co.

Cosmides, L. and Tooby, J. 2005. Neurocognitive Adaptations Designed for Social Exchange. In Buss, D. *The Handbook of Evolutionary Psychology*. New Jersey: John Wiley & Sons.

Cosmides, L. and Tooby, J. 2004. Social Exchange: The Evolutionary Design of a Neurocognitive System, In Gazzaniga, M. *The Cognitive Neurosciences III*. Cambridge Massachusetts: MIT Press.

Dapretto, M and Iacoboni, M. Dec. 2006. The Mirror Neuron System and the Consequenses of its Dysfunction. *Nature Reviews Neuroscience.* Vol. 7, Iss. 12, pp. 942-951.

Dapretto, M., Davies, M. Pfeifer, J., Scott, A., Sigman, M., Bookheimer, S., and Iacoboni, M. Jan. 2006. Understanding emotions in others: Mirror Neuron Dysfunction in Children with Autism Spectrum Disorders. *Nature Neuroscience*. Vol. 9, Iss. 1, pp. 28-30.

Dawkins, R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

Deutsch M. 1973. *The Resolution of Conflict: Constructive and Destructive Processes*. New Haven: Yale University Press.

Dunbar, R. 1996. *Grooming, Gossip and the Evolution of Language*. London: Faber and Faber.

Everett, D. 2005. The Language Organ: Linguistics as Cognitive Physiology. *Journal of Linguistics,* Vol. 41, Iss. 1, pp. 157-175.

Fiddick, L. 2004. Domains of Deontic Reasoning: Resolving the discrepancy between the cognitive and moral literatures. *Quarterly Journal of Experimental Psychology*, Vol. 57A, Iss. 4, pp. 447-474.

Fiddick, L. 1998. *The Deal and the Danger: An evolutionary analysis of deontic reasoning*. Doctoral dissertation, Department of Psychology, University of California, Santa Barbara.

Foot, P. 1967. The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review,* Vol. 5, pp 5-15.

Gigerenzer, G. 2000. *Adaptive Thinking: Rationality in the Real World*. New York: Oxford University Press.

Gigerenzer, G. and Hug, K. 1992. Domain Specific Reasoning: Social Contracts, Cheating, and Perspective Change. *Cognition,* Vol. 43, pp. 127-71.

Hauser, M. 2006. *Moral Minds*. New York: Harper Collins.

Hobbes, T. 1651/1962. *Leviathan*. London: Dent.

Kant, I. 1785/2002. *Groundwork for the Metaphysics of Morals*. Ed. Wood, A. New Haven: Yale University Press.

Leyens, J., Cortes, B., Demoulin, S., Dovido, J., Fiske, S., Gaunt, R., Paladino, M., Rodriguez-Perez, A., Rodriguez-Torres, R., and Vaes, J. 2003. Emotional Prejudice, Essentialism, and Nationalism - The 2002 Tajfel Lecture. *European Journal of Social Psychology*. Vol. 33, Iss. 6, pp. 703-717

Lieberman, D., Tooby, J., and Cosmides, L. 2007. The Architecture of Human Kin Detection. *Nature*, Vol. 445, pp. 727-731.

Lieberman, D., Tooby, J., and Cosmides, L. 2003. Does Morality have a Biological Basis? An empirical test of factors governing moral sentiments relating to incest. *Proceedings of the Royal Society of London:* Series B-Biological Sciences 270 (1570), pp. 819-826.

Mikhail, J. 2000. Rawls' Linguistic Analogy: A study of the "Generative Grammar" Model of Moral Theory Described by John Rawls in "A Theory of Justice." *Dissertation Abstract International A, The Humanities and Social Sciences*. Vol. 61 Issue 4, pp.1446.

Moll, J. et al. 2006. Human Fronto-Mesolimbi Networks Guide Decisions About Charitable Donation. *Proceedings of the National Academy of the Sciences*. Vol. 103, Iss. 42, pp. 15623-8.

Myers, D. 2001. *Psychology*. New York: Worth Publishers.

Oberman, L. and Ramachandran, V. 2007. The Simulating Social Mind: The Role of the Mirror Neuron System and Simulation in the Social and Communicative Deficits of Autism Spectrum Disorders. *Psychological Bulletin*. Vol. 133, Iss. 2, pp. 310-327.

Orend, B. 2000. *War and International Justice: A Kantian Perspective*. Waterloo, Canada: Wilfrid Laurier University Press.

Pinker, S. 2002. *The Blank Slate*. New York: Viking.

Plato. c.360 B.C./2005). The Republic, in *Plato: The Collected Dialogues, including the Letters*. Ed. Hamilton, E. and Huntington, C. Princeton: Princeton University Press.

Plomin, R., DeFries, J., McLearn, G., and McGuffin, P. 2001. *Behaviour Genetics* (3rd. edition). New York: Worth.

Pugh, George E. 1978. *The Biological Origin of Human Values*. London: Routledge.

Quartz, S. and Sejnowski, T. 2002. *Liars Lovers and Heroes*. New York: Harper Collins.

Ridley, M. 1997. The Origin of Virtue - Human Instincts and the Evolution of Cooperation. New York: Viking.

Richardson, R. 2007. The Adaptive Programme of Evolutionary Psychology. In Thagard, P. *Handbook of the Philosophy of Science: Philosophy of Psychology and Cognitive Science*. Amsterdam: Elsevier.

Rilling, J., Gutman, D., Thorsten, R., Pagnoni, G., Berns, G., and Kilts, C. A Neural Basis for Social Cooperation. *Neuron*. Vol. 35, Issue 2, pp. 395-405.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. 1996. Premotor Cortex and the Recognition of Motor Actions. *Cognitive Brain Research*, Vol. 3, Iss. 2, pp. 131-141.

Samuels, R. and Stich, S. 2004. Rationality and Psychology. In Mele, A. and Rawling, P. *The Oxford Handbook of Rationality.* Oxford: Oxford University Press.

Segal, N., Hershberger, S., and Arad, S. Meeting One's Twin: Perceived Social Closeness and Familiarity. *Evolutionary Psychology*. Vol. 4, pp. 70-95.

Sober, E. and Wilson, D. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge: Cambridge University Press.

Stanovich, K. 1999. *Who is Rational?* New Jersey: Erlbaum.

Stanovich, K. 2003. The Fundamental Computational Biases of Human Cognition: Heuristics that (sometimes) impair decision making and problem solving. In Davidson, J. and Sternberg R. (Eds.). *The Psychology of Problem Solving*. New York: Cambridge University Press.

Stanovich, K. 2004. *The Robot's Rebellion*. Chicago: University of Chicago Press.

Sunstein, C. 2005. Moral Heuristics. *Behavioural and Brain Sciences.* Vol. 28, Iss. 4.

Thorpe, W.H. 1972. *Animal Nature and Human Nature*. New York: Anchor Press.

Turkheimer, E. 2000. Three laws of behaviour genetics and what they mean. *Current Directions in Psychological Science.* Vol. 5, pp. 160-164.

Wynne-Edwards, V. 1986. *Evolution Through Group Selection*. Oxford: Blackwell Scientific.