The Genealogy of Moral Memes

A New Synthesis for Evolutionary Ethics

by

Scott F. Woodcock
University of Toronto

November 24, 2000

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Philosophy
University of Toronto

© Copyright Scott Woodcock 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

0-612-59050-X
The Genealogy of Moral Memes
A New Synthesis for Evolutionary Ethics

Scott Woodcock
Department of Philosophy
Ph.D. 2001
University of Toronto

Abstract

This thesis is an investigation of the consequences that arise if we take up an evolutionary perspective towards ourselves as moral agents. In particular, it examines the implications of modeling cultural evolution for prescriptive ethical theory. I argue that if the newly emerging science of memetics (the study of evolutionary models of cultural information transmission) develops into a mature science, it will have a significant impact on our ability to systematically justify the authority we attribute to our ethical judgments. This impact is made possible by the fact that memetics will allow us to very precisely trace the cultural genealogy of our moral intuitions. In doing so, memetics will help us to identify situations where our intuitions are supported by recognizably defective cultural selection mechanisms. The underlying message of the thesis is that we cannot trust cultural evolution to produce outcomes that necessarily promote our best interests, and examples of anomalous selection mechanisms are provided to illustrate that even cultural natural selection often leads to spectacularly undesirable results. Such results are shown to be inevitable because of the overpowering tendency for cultural evolution to economize on information costs. Hence, the ability to identify moral intuitions that persist because of anomalous selection mechanisms is a valuable asset, for it provides us with an additional means of evaluating the trustworthiness of our intuitions and the prescriptive authority they ought to possess. Finally, I also argue that seeking to eliminate distortions in our intuitions reveals something important about what we should expect from a well-functioning moral psychology. I claim that the prospect of taking up a critical standpoint towards our evolved moral intuitions leads us to encourage two separate levels of moral thinking within the psychology of each virtuous moral agent.
Acknowledgments

I would like to thank the University of Toronto and the Ontario Graduate Scholarship program for their financial assistance during the writing of this thesis. Without the funds provided by these institutions (and their federal source), the ideas expressed in this thesis would never have seen the light of day.

I would also like to thank the department of Philosophy at the University of Toronto for accepting me into their program and providing me with a friendly, creative environment in which to carry out my doctoral research. I have found the faculty in this department to be cooperative and highly dedicated to helping those graduate students who seek to gain the benefits of their experience and wisdom.

I am, of course, especially thankful for the advice given to me over the past three years by my thesis committee. Paul Thompson has consistently come through with support and encouragement whenever it mattered most, and he has graciously allowed me to take this project in directions that must have often seemed to him to be deeply perplexing. Joseph Heath has provided me with the benefit of honest, forthright criticism, and his animated enthusiasm for philosophy is both inspiring and wonderfully contagious. Finally, Wayne Sumner has displayed a heroic amount of patience with my project, and his understated but acutely perceptive advice has taught me a vast amount about ethics and the precision required to clearly communicate one’s ideas.

I also owe many thanks to the graduate students with whom I have had discussions over the years. In particular, I am grateful that Violetta Igneski, Emer O’Hagen and Anthony Skelton have allowed me to ‘talk shop’ with them for as long as they would listen.

On a more personal note, I must thank my parents for their unwavering emotional support and for the fact that they have never once second-guessed my somewhat quixotic choice to pursue a doctoral degree.

I am, however, most deeply thankful for the support and patience shown to me by my best friend, roommate, partner, companion and wife to be, Anna Mason. She alone had to watch this thesis be written and thus had to endure a seemingly endless sequence of late nights, panic attacks and nearly glacial progress. Without her help this thesis would have quite simply been impossible. She has my immeasurable love and gratitude.

Finally, I feel I should thank Hal 9000, my MacIntosh LC520, for persisting long enough to let me complete this thesis. I am sure it often wanted to quit almost as bad as I did, but thankfully it survived long enough to finish the job.
When will we complete our de-deification of nature? When may we begin to "naturalize" humanity in terms of a pure, newly discovered, newly redeemed nature?

Nietzsche, p. 169
## Contents

Abstract ii

Introduction 1

**Chapter One. A Critique of Current Evolutionary Ethics** 10

Introduction 10

I. Michael Ruse: Error Theory and Epigenetic Rules 15
   * Innate Objectivity 16
   * Avoiding Hume’s Law 17

II. Levels of Moral Agency and Prescriptive Authority 21
   * Specificity and Ethical Sophistication 22
   * Two Responses 26

III. A Valuable Empirical Resource 29
   * The Genealogy of Social Dispositions 31
   * A Valuable Flexibility 32

Conclusion 34

**Chapter Two. Cultural Evolution and the Move Towards Memetics** 36

Introduction 36

I. Cultural Inheritance 40
   * Two Developments 41

II. Modeling Cultural Evolution 44
   * Three Challenges 44
     * Adaptive Origins 45
     * Explanatory Independence 48
     * The Right Tools for the Job 56

III. The Move Towards Memetics 58
   * Memetic Independence 58
   * Dawkins’ Legacy 59
   * Influence and Potential 61
   * Technical Obstacles 64
   * Potential Responses 69

Conclusion 72

**Chapter Three. Ethics and Memetics: A Best-Case Scenario** 74

Introduction 74

I. Anticipating the Best-Case Scenario 78
   * Cultural Fitness 80

II. Ethics: Initial Reactions 81
   * Memetic Engineering 82
   * Memetic Determinism 86
   * Memetic Skepticism 91
Chapter Six. Striking a Balance: The Virtues of Reflective Instability 199

Introduction 199

I. Three Options 201
   A Ruling Elite 202
   Bury our Heads in the Sand 205
   A Delicate Internal Balance 212

II. Triggers, Anchors and Moral Alarm Clocks 214
   A Peculiar Balancing Act 216
   The Need for Reflective Instability 218
   Two Separate Functions 223

III. Memetic Schizophrenia 226
   Explaining our Unified Experience 227
   An Emerging Awareness 230
   Psychological Harmony as Virtue in Practice 236

Conclusion 237

Conclusion: Moral Monkey Traps and Naturalized Perfection 240

Bibliography 244
Introduction

It remains an open question whether the study of ethics has fully come to grips with the evolutionary origins of human nature. On the one hand, the fact that evolution has, in some form or another, painstakingly constructed the natural world is a conviction more widely accepted now than ever before. Charles Darwin's The Origin of Species is over a hundred and forty years old (Darwin 1859), and its legacy continues to revolutionize the natural sciences on a seemingly daily basis. On the other hand, the implications of adopting an evolutionary view of ourselves as part of the natural world remain stubbornly indeterminate. Evolutionary theory carries the potential to dramatically transform our conceptions about what it means to be human, but it has so far proven unable to definitively shape our philosophical convictions about what this transformation will entail.

This thesis is an investigation into the consequences that arise if we take up a specifically evolutionary perspective towards ourselves as moral agents. It attempts to answer the question, How will the study of ethics be affected if we view ourselves as Darwinian creatures who have somehow developed uniquely moral capacities? The question is an important one, for debate over the impact evolutionary theory ought to have on ethics continues to stir up controversy among philosophers who are otherwise willing to agree on a naturalistic picture of our evolutionary origins. Some theorists have suggested that Darwinism will completely restructure ethics.¹ Others, however, remain perfectly content to accept that we have been designed by natural selection without believing that this information ought to have any significant impact on the way we approach ethical inquiry. The result is that current research is characterized by a strange

¹ Ruse and Wilson (1986), for example, go as far as to claim that evolutionary theory will turn moral philosophy into an applied science.
contradiction: philosophical interest in topics like the structure of evolutionary theory or the possibility of naturalized moral epistemology is currently stronger than ever, yet the perception within contemporary analytic ethics is that evolutionary based methodologies are not worth serious consideration. The aim of my thesis is to overcome this perception and to demonstrate that evolutionary theory can contribute something meaningful to the way we approach substantive ethical theory.

There are two specific motives that govern my approach to the question of how evolutionary theory ought to affect ethics. First, with the exception of chapter one I am principally concerned with the implications of cultural evolution for ethics. Biological and cultural evolution are no doubt intimately connected within an overall coevolutionary process. However, my focus is on the cultural forces working within this coevolutionary process, and I primarily discuss the ethical implications of memetics, the study of evolutionary models of cultural information transmission. Second, I am exclusively interested in the impact our evolutionary origins ought to have on the prescriptive justification of our ethical intuitions. I therefore address questions related to how evolutionary theory will affect practical questions in ethics - questions where we must decide whether the prescriptive character of certain specific intuitions is genuinely warranted. Evolutionary theory is perhaps also well suited to investigating how the prescriptive character of our moral intuitions has originated from the raw material of a naturalistic social system, but I confess that I find questions of moral ontology like this one exceedingly dull. My interest instead lies in how to most effectively legislate the existing prescriptivity in our intuitions in order to best achieve whatever foundational ethical objectives we choose to pursue.

The most important concept in my thesis is that natural selection, whether it be biological or cultural, cannot be trusted to promote our best interests. In the case of our
inherited genetic dispositions, we acquire many beneficial social traits in the fabric of our biological construction. However, we cannot trust that these traits will always work to our advantage. Though it is sometimes tempting to assume that we have been designed by evolution to fulfill some function from which we will surely benefit, we have no guarantee that natural selection has left us with dispositions that necessarily promote our well-being. Natural selection operates according to its own independent logic, in that it promotes *whatever happens to replicate most effectively*. Thus, our genetic social dispositions have evolved because they are adept at passing on copies of themselves to future generations; they have not evolved because they necessarily promote human flourishing or our current social objectives.

Similarly, in the case of cultural evolution, we cannot trust the products of natural selection to secure our best interests any more than we can trust their biological counterparts. It may seem odd that we must also be suspicious of cultural evolution, because as intentional creatures we are able to play an active part in directing the selection forces that shape cultural evolutionary outcomes. Can we not proactively guide cultural evolution to ensure that it promotes our interests? In many ways we can. However, we cannot trust that our efforts will be successful in shaping the final products of cultural evolution. The reason we can have no such guarantee is not only a factor of the tremendous complexity present in our cultural systems. There are also, I argue, *persistent selection anomalies* built into the structure of cultural evolution. These anomalies create distortions in what we would otherwise expect to emerge from the ordinary functioning of our cultural systems. Like the peculiar selection forces that lead to maladaptive biological traits (e.g. peacocks' tails or the panicked urge to conform in lemmings), cultural selection can fall victim to similar traps. The aim of my thesis is therefore to describe the nature of these anomalous forces and the unusual kinds of cultural distortions that they generate.
The upshot of these cultural anomalies for ethics is that we cannot trust our moral intuitions to necessarily serve our best interests. This is a significant result, for we must ultimately rely on our intuitions to settle the conflicts that arise in our ethical judgments. We know already that we must guard against the possibility that our moral intuitions reflect mere prejudices, but the fact that cultural evolution is susceptible to chronic structural irregularities suggests that even commonly accepted intuitions may be the products of anomalous selection mechanisms. If so, then even our deeply held moral intuitions may be distorted somehow, and as they continue to evolve over time they may become dramatically inconsistent with our basic ethical objectives.

The potential for these anomalies will be difficult to conceptualize at this point without further elaboration. I can ask only for the reader’s patience and hope that the details will become clear farther along in the text. The key idea is that cultural selection mechanisms are imperfect, and that they sometimes supply us with defective intuitions which do not deserve our prescriptive endorsement.

After establishing this fact about our intuitions, I argue in the remainder of the thesis for two interconnected conclusions. First, I argue that the most profitable application of memetics to ethics is to use the empirical information generated from studying cultural evolution to identify distortions in our moral intuitions in order to better systematize the prescriptive authority we attribute to our ethical judgments. To accomplish this, I recommend that we investigate the cultural genealogy of our moral intuitions to determine whether they have been produced via defective selection mechanisms. By employing the genealogical information supplied by memetics as an aid to critical reflection, we can use evolutionary theory to more effectively scrutinize our common-sense moral intuitions and legislate the prescriptive influence that they exert over us.
The second conclusion I argue for in the thesis is that the possibility of eliminating distortions in our intuitions reveals something important about what we ought to expect from an ideal moral psychology. Specifically, I argue that the best way to allow ourselves the opportunity to identify defective moral intuitions (without exposing ourselves to the corrosive forces associated with overzealous critical reflection) is to promote two levels of moral thinking within what is on the whole a reflectively unstable moral psychology. To explain this idea I draw upon literature in ethics related to indirect forms of consequentialism, self-defeating strategies and 'satisficing' approaches to optimization. This may seem like an eclectic mix to prove a point about the moral psychology required to deal with cultural selection problems. Nevertheless, by appealing to these concepts I attempt to demonstrate that an awareness of the evolutionary forces acting on our intuitions ought to persuade us to accept a delicate balance between separate components operating within our more general capacity for moral thinking.

If nothing else, then, the project is original. Memetics has become exceedingly fashionable and the issues surrounding indirect strategies have been extensively discussed in ethics literature for the past twenty-five years. But no one, to my knowledge, has attempted to bridge these two areas so explicitly in order to explain the importance of the latter in terms of the former.² In the process of blending these two disparate areas, I am unfortunately forced to set aside many compelling issues and provide only cursory answers to others. Yet, I hope this project can nonetheless weave a coherent path through these areas, and that it can serve as a pioneering attempt to uncover what will someday be considered an obvious connection between the two.

---

² Dennett (1995) comes very close, and he does explicitly note the link between natural selection and Herbert Simon's satisficing view of rational decision making (Simon 1959). But as I explain in chapter six, he does not take the next step and connect evolutionary theory to the reflectively unstable psychology required to implement ethical theories like indirect forms of consequentialism.
The thesis is organized as follows. Chapter one introduces the idea that we can take up a critical perspective towards our evolutionary dispositions. In this chapter I offer a critique of prescriptive forms of Evolutionary Ethics, which attempt to ground a direct source of ethical value on the sociobiological facts of our evolved social dispositions. Using the philosophy of Michael Ruse as a test case, I argue that even though there is a strong genetic component to our social behaviour, we can nonetheless distinguish between social dispositions we ought to promote and those we ought to suppress - a fact which indicates that our critical abilities, not our genetic dispositions, are doing all the justificatory work in prescriptively authorizing our ostensibly innate moral nature. Hence, I claim that evolutionary theory is better suited to serve as a valuable source of empirical information than as a direct source of prescriptive ethical value.

After this first chapter, the structure of the thesis begins to take on an hourglass shape. Chapters two and three discuss the prospects for modeling cultural evolution and the potential consequences that a successful science of memetics will produce for the study of ethics. These chapters address a wide range of possibilities, and they resemble the first half of the hourglass because they are largely concerned with narrowing these possibilities to applications that will have a significant philosophical impact. The bottleneck of the hourglass is chapter four, which describes the one application of these possibilities that holds the most promise (the use of cultural genealogy as a means of systematizing our moral intuitions). At the other end of this bottleneck, chapters five and six open up the implications of what has been established in chapter four. These chapters attempt to show that using memetics to reflect on the authority of our intuitions leads to a much more general conclusion about the nature of our moral psychology.

Chapter two is therefore a very broad introduction to the newly emerging science of memetics - one that I hope will be accessible to those unfamiliar with the prospects of modeling cultural evolution. The chapter also presents some of the methodological problems facing memetics. In doing so, it attempts to give the reader a feel for both the
potential memetics seems to possess and the difficulties that arise for the aim of mapping out cultural patterns as if they were genetic units. The conclusion reached is that despite a number of very serious methodological obstacles, memetics may be able to provide us with spectacular new insights into the way cultural information is transmitted and selected.

Chapter three outlines the ethical implications we ought to expect if memetics overcomes its current methodological problems and succeeds at developing into a mature science. The chapter therefore adopts a 'best-case scenario' approach and looks at various potential applications of memetics to ethics. The most viable of these applications is determined to be one that is strikingly similar to the strategy recommended in chapter one: using memetics as a valuable source of empirical information about our moral intuitions to help us reflect on the prescriptive authority these intuitions ought to possess.

This possibility is explored in more detail in chapter four. After defending the premise that we can take up a critical perspective towards the contents of our own minds, the chapter argues that we can use the genealogical information supplied by memetics to weed out moral intuitions that have been allowed to propagate because of anomalous cultural selection forces. One such anomalous selection force is described in detail: the propensity for cultural evolution to excessively favour *indirect indicators* of ethical value. Using examples drawn from cases in ethics where our deeply internalized intuitions conflict with our judgments about what serves our ethical objectives, the chapter closes with the conclusion that memetics can help us prevent instances where the normative authority of our intuitions has become exaggerated beyond its justifiable limits.

In chapter five, however, a problem is introduced for the aim of critically reflecting on the genealogy of our moral intuitions. The problem is that the implementation of this aim has an inevitably corrosive effect on the prescriptive authority of many justifiable moral intuitions. Thus, the process of weeding out the products of
defective cultural selection forces cannot be achieved without considerable cost, for it has the unfortunate consequence of undermining many important social norms and virtuous character traits that must be insulated from rational scrutiny in order to function effectively.

Finally, chapter six attempts to reconcile the objectives of chapter four with the corrosive implications of chapter five. I argue that the tension between these two chapters is ultimately unavoidable. However, I also argue that the most effective way to manage this tension is to promote two separate levels of moral thinking within the psychology of each individual agent. The result is that each agent must possess a reflectively unstable moral psychology: when operating at an intuitive level of thinking, one simply cannot be aware of the considerations driving our critical level of thinking without inviting the corrosive effects outlined in chapter five. This is an unusual conclusion for the way we conceive of our moral psychology, but I believe it is the best option we have available to us considering our status as evolutionary moral creatures.

Considering the scope of the conclusion reached at the end of chapter six, one might ask whether memetics, or even evolutionary theory, is necessary to point out what I want to argue about the study of ethics. It may seem as if memetics is being used as a kind of Trojan horse to push through an independent hypothesis about the nature of critical reflection and our moral psychology. This is not my intention. It is certainly possible that what I want to reveal about the structure of moral thinking can be explained without the help of memetics. The conclusion reached in chapter six is general enough that every reasonable ethical framework ought to eventually recognize its merits without having to resort to the specialized terminology of evolutionary theory. However, appealing to evolutionary theory serves as an especially clear method of articulating the basis for this conclusion. It explains why our moral psychology has developed in the way that it has, and it explains why we cannot avoid either allowing distortions in our
intuitions to arise or else inviting the corrosive side-effects that are produced if we use critical reflection to try to curb these distortions. In this respect, I believe the language of memetics provides us with an invaluable resource to draw upon in order to better understand ourselves and more effectively manage the prescriptive justification of our ethical judgments.
Chapter 1

A Critique of Current Evolutionary Ethics

Introduction

The hesitancy displayed on the part of analytic ethics to investigate evolutionary methodologies is understandable. With the connotations of Social Darwinism, early twentieth century eugenics projects and such seemingly ruthless maxims as "survival of the fittest" left lingering in our cultural memory banks, it is no wonder many academics are reluctant to mix evolutionary theory with ethics. Yet the most significant obstacle standing in the way of evolutionary approaches to ethics has not been these shadowy remnants of past mistakes. Instead, the most serious problem preventing evolutionary approaches to ethics from gaining acceptance has been a philosophical concern over whether or not evolutionary theory is qualified to have an impact on the prescriptive justification of ethical claims.

As the philosophy of Herbert Spencer dramatically illustrates,¹ even the most well intentioned attempts at blending evolutionary theory and ethics tend to project inappropriate values onto the morally neutral machinery of nature. Hence, the most serious problem blocking evolutionary methodologies in ethics is that the facts of biology appear ill-equipped to help us resolve questions related to normative moral principles. Even after we clear away the misconceptions about contemporary forms of Evolutionary Ethics (EE) being associated with unsavory political agendas, a legitimate question remains about whether the empirical data provided by evolutionary theory is properly suited to contribute to the prescriptive task of justifying our ethical judgments.

¹See (Spencer 1893).
The problem of applying evolutionary theory to ethics revolves primarily around the is/ought distinction - an axiom that has come to be known commonly as Hume's Law. Hume's Law states that one cannot derive prescriptive 'ought' statements by appealing directly to descriptive 'is' statements. Although originally not meant as an absolute prohibition, Hume's message is that the jump from facts to norms requires something more than an unsubstantiated appeal to empirical information if our justifications for normative principles are to stand on secure foundations.\(^2\) Hume's Law forces us to clearly articulate the means by which prescriptive claims are generated, and it seems to present a special challenge for the aim of drawing prescriptive conclusions based on evolutionary facts. If the gap between is and ought must be carefully accounted for, then something more than a direct appeal to evolutionary information is required in order to convincingly justify our ethical judgments.

Despite the stark challenge presented by Hume's Law, several prominent writers in the philosophy of biology have attempted to show that evolutionary theory can, in fact, make a significant contribution not only to our understanding of ethics but to its justification. In the mid 1970's, the move towards EE gained considerable momentum (and no small amount of notoriety) with the publication of E.O. Wilson's ground breaking work *Sociobiology* (1975), a book that examined the evolutionary causes of social behaviour in animals and included a final chapter on the genetic basis for human behaviour. Wilson's book was initially met with venomous resistance, but it had a profound influence on those willing to set aside the political dimensions of sociobiology in order to sort out the underlying philosophical implications of the project.\(^3\) The result was a flurry of activity in the area and the emergence of a contemporary approach to EE

---

\(^2\) In fact it seems fairly ironic that the injunction against deriving 'ought' from 'is' has come to be known as "Hume's Law", since Hume himself clearly did not believe that this axiom ought to be unconditionally imposed. Moreover, one might just as well invoke Hume to ground an injunction against deriving 'is' from 'is', since Hume was, of course, also a skeptic about induction and natural causation.

\(^3\) For discussion of the controversy that surrounded sociobiology when it first emerged as a field of study, see the extensive anthology *The Sociobiology Debate* (Caplan 1978).
that examined the ethical consequences of evolutionary theory with a far greater sense of awareness of the danger of violating Hume's Law by moving too quickly from descriptive facts to prescriptive moral obligations.

The most common methodology among contemporary versions of EE has been to abandon the idea that the structure of evolution itself provides us with a normative example to follow. Instead, this methodology shifts our attention to the evolutionary forces responsible for constructing the unique character of our human moral capacities, and it investigates our biological propensities to behave in morally noteworthy ways. Thus, rather than looking to evolution as a direct source of moral guidance, recent articulations of EE attempt to draw prescriptive ethical conclusions based on the existence of certain key aspects of human nature that can be considered constitutive of a biologically grounded sense of human moral agency. By drawing attention to the evolutionary basis for our moral capacities, current forms of EE attempt to ascribe a sense of prescriptive authority to our underlying genetic capacities in virtue of the fact that they represent an inalienable part of human nature that enables us to be uniquely moral beings. This methodology represents a boldly naturalistic form of ethical justification, but contemporary proponents of EE hope to demonstrate that this form of justification is all that is required (and perhaps all that is possible) in order to secure the normative status of our moral commitments.

In this chapter, I argue that examining the evolutionary origins of human moral agency could prove to be an interesting and very profitable exercise for the study of ethics. However, I also aim to demonstrate in this chapter that the attempt to derive explicitly prescriptive versions of Evolutionary Ethics ought to be abandoned. By this I mean that we ought to give up on the attempt to ground a direct sense of prescriptive ethical authority on the existence of our evolved predispositions for moral behaviour. Instead, I hope to show that although a naturalistic explanation of our moral capacities
constitutes a promising source of empirical information for ethics, the sociobiological details of these capacities are not nearly as relevant to the justification of our ethical principles as are our cultural contexts and our ability to critically evaluate the legitimacy of our genetic inclinations. I thus intend to demonstrate that attributing justificatory power to our inborn social capacities is a misguided enterprise, despite the fact that contemporary proponents of EE have taken great care to avoid the more obvious mistakes made in the name of EE in the past.

The reason why the recent move towards attributing justificatory power to our genetic moral capacities is misguided can be broken down into a simple two step argument. First, although a naturalistic account of moral agency will certainly be illuminating, it is also quite likely that this account (if it is to be satisfactory) will be extremely complex and therefore far too intricate to allow direct ascriptions of justificatory power to any clearly definable set of natural capacities. It may be easier to speak in terms of some kind of discrete "moral sense" until we achieve a more advanced understanding of the mechanisms involved in producing our moral capacities, but we should not let this provisional shortcut allow us to think that we possess a clearly demarcated genetic moral nature that is responsible for guiding our thoughts and actions in ethically appropriate ways. If we acknowledge the fact that a naturalistic account of our moral capacities will likely provide a more complicated picture of human nature than is often assumed, it becomes evident that prescriptive theories of EE face a difficult task attributing prescriptive authority to a distinct set of genetic traits within what is an exceedingly complex system of normative control.

The second step of the argument is to question why our genetic social dispositions are singled out from this complex system of normative control as uniquely capable of grounding prescriptive ethical claims. It is important to be clear about the specific mechanisms responsible for generating the prescriptive authority we attribute to our ethical judgments, because it is far from obvious what makes our genetic social capacities
the definitive source of authority within the scope of a complex naturalistic account of human moral nature. The (albeit fascinating) existence of these capacities is often tacitly relied upon to fulfill a prescriptive ethical role. But if we look closely at what is doing the philosophical 'work', so to speak, in questions of ethical justification, the details of our cultural surroundings and our collective decisions to promote certain social dispositions over others seem to hold an equal, if not greater, amount of weight in terms of explaining why we ought to respect the specific capacities we identify as being prescriptively valuable. If this is the case, then the sociobiological details of our broadly social genetic dispositions seem much less like descriptions of our fully developed moral capacities and therefore much less vital to questions of ethical justification than is often supposed.

To demonstrate that this argument represents a decisive refutation of prescriptive forms of EE, this chapter examines the philosophy of Michael Ruse, a vocal proponent of the idea that ethics ought to be profoundly influenced by the evolutionary origins of moral agency. Ruse's work is addressed not only because he is one of the more prolific and widely recognized writers in the area, but also because his views are carefully constructed in order to take the challenge of providing ethical justifications seriously. Ruse often writes bold assertions about the potential for evolutionary theory to revolutionize philosophy, but he is also careful about minding the gap between "is" and "ought" and makes noticeable efforts to avoid violating Hume's Law by bridging this gap too hastily. For this reason, Ruse serves as a useful test case in order to demonstrate that even sophisticated versions of prescriptive EE remain deeply problematic.

The chapter is organized as follows. Section one summarizes Ruse's position with particular attention given to the way in which the theory generates its subtle justificatory power. Section two offers reasons why Ruse fails to appreciate the complexity of our capacity for moral behaviour and argues that he makes the mistake of attributing
prescriptive authority to our genetic social dispositions without acknowledging the role our more sophisticated critical (and therefore only trivially evolutionary) abilities play in determining the ethical value of these dispositions. Finally, section three argues for a revised approach to EE that uses information about the evolutionary origins of our moral capacities as an important empirical resource without attributing direct justificatory power to the capacities themselves. By briefly sketching how evolutionary theory can act as a source of genealogical information about our moral nature, the hope is that a viable alternative to current EE can be offered that remains consistent with a naturalistic understanding of moral agency without trivializing the complexity of how this unique moral capacity has evolved.

I. Michael Ruse: Error Theory and Epigenetic Rules
Michael Ruse has written extensively on the subject of evolutionary theory and its potential implications for ethics, but his most significant contribution to the area is Taking Darwin Seriously (TDS), an examination of the philosophical consequences of contemporary Darwinism. In this book, Ruse offers what is perhaps the most definitive and clearly articulated statement of his views on Evolutionary Ethics. The underlying message in TDS is that the human capacity for moral behaviour is a biological adaptation that has evolved because it allows humans to reap the benefits of cooperative interaction. However, Ruse also makes the claim that we have developed an innate propensity to think and act according to objective moral standards because these standards serve as an especially efficient means of coordinating human interaction. By causing us to believe in binding, objectively authoritative rules of moral conduct, our genes are said to have found a convenient way of programming us to internalize limits on our otherwise selfish motives so that we can achieve the lucrative benefits of collective action. Thus, Ruse's

---

4 See (Ruse 1984), (Ruse 1986a), (Ruse and Wilson 1986) and (Ruse 1990).
5 See (Ruse 1986b).
message is not only that we share an inborn set of primary moral directives. His claim is also that we share the unique genetic ability to approve and disapprove of social behaviour according to what we perceive to be external, objective rules of conduct.

To defend this claim, that morality is another biological adaptation, "just like hands and feet" (TDS, p. 222), Ruse relies heavily on the work of E.O. Wilson for the details of how our moral nature is generated by evolutionary dispositions. According to Ruse (via Wilson), our moral capacities are grounded in 'epigenetic rules' - developmental biases controlled by our genetic makeup that predispose us to favour certain types of behaviour over others. Epigenetic rules do not directly control our actions; they instead set broad constraints on our behaviour, the details of which are then filled in by cultural information and environmental circumstance. The metaphor cited to describe this relationship is that genes and culture are held together by "an elastic but unbreakable leash." Environmental and culture admittedly produce profound effects on our moral dispositions, but they do so only within the constraints imposed by an underlying biological framework.

**Innate Objectivity**

What Ruse brings to this preexisting theory of the genetic basis for moral behaviour is the specific conclusion that we have been designed with the unique ability to think in terms of external rules of conduct. According to Ruse, the epigenetic rules responsible for our moral nature program us to have altruistic impulses, but they also impel us to evaluate social interactions according to (seemingly) objective standards of value. Thus, our genes have designed us in such a way that we are genetically predisposed to believe in the objectivity of moral judgments even if no such objective moral standards exist. Ruse claims that this aspect of the genetic basis for moral agency must be present, because if our genes merely provided us with something like a

---

propensity for sympathetic emotions it would not be enough to ensure a stable basis for cooperation over time. Sympathy might be enough to hold a family unit together (or perhaps a small tribe or close circle of friends) but in general something more is required in order to persuade us to think in terms of duty and obligation - sentiments that reach beyond the limits of our volatile emotions and the narrow confines of our personal commitments. Hence, Ruse claims that the epigenetic rules responsible for our moral nature have carefully designed our ethical intuitions to believe in the existence of objective sources of authority because this design dramatically increases the adaptive value of cooperative behaviour, even if the standards we are compelled to believe in are entirely illusory.

Ruse's contribution to Evolutionary Ethics is therefore an 'error theory' with respect to the foundation of morality.\(^7\) The theory asserts that a detailed understanding of the evolutionary origins of our moral intuitions demonstrates that ethics is ultimately (what Ruse refers to as) a subjective enterprise, for the true source of authority attached to moral principles is located within ourselves rather than any external standard. However, it also asserts that if our ethical principles are to function effectively, moral agents must be misled into thinking that ethics is not merely a subjective enterprise. The objective character of morality is therefore an illusion, but it is an illusion that is necessary for humans to successfully cooperate with one another. (TDS, p. 253)

*Avoiding Hume's Law*

Much of the criticism directed towards Ruse's views has taken issue with the details of appealing to an error theory of moral obligation or the conclusion that a naturalistic understanding of our moral capacities invariably supports a subjectivist meta-ethical standpoint.\(^8\) The literature written on these topics is intriguing. However, I want

\(^7\)Ruse is aware of the debt owed to (Mackie 1977).

\(^8\)For discussion on these aspects of Ruse's philosophy, see (Rottschaefer and Martinsen, 1990), (Barrett 1991), (Collier and Stingl 1993), (Woolcock, 1993), (Waller 1996), (Campbell 1996) and (Ryan, 1997).
to focus on a different aspect of Ruse's proposal in order to address a problem that seems to me to be more endemic to modern attempts to defend prescriptive versions of EE. The question I want to focus on is how Ruse's theory generates the sense of *justificatory power* it attributes to our genetic social dispositions. By defending an error theory of moral judgment, Ruse is clearly not attempting to give a descriptive account of our evolved moral capacities with the intention of stripping these capacities of their prescriptive force. It is therefore worth looking closely at how Ruse expects to derive a binding form of prescriptive authority from a theory that states that "morality is a collective illusion foisted upon us by own genes." (TDS p. 253)

In particular, if Ruse takes the challenge of Hume's Law seriously, how does he manage to avoid its constraints? Ruse is careful not to commit himself to any premature claims about evolutionary theory's prescriptive influence, and continually makes a point of noting that issues of justification deserve special attention if the mistakes made by EE in the past are to be avoided. But after setting up an impressive case against older forms of EE being able to generate prescriptive authority, Ruse leaves himself a difficult task explaining how *any* theory can draw prescriptive conclusions from the empirical facts of evolutionary theory. In effect, Ruse asks us to take Hume's Law seriously enough that his own project is at risk of being stripped of any meaningful justificatory power.

The solution to this problem, according to Ruse, is to avoid confronting Hume's Law directly and to instead perform an "end run" around the is/ought barrier. (TDS p. 256) The method Ruse suggests for doing so is a familiar one in terms of contemporary

---

9 To contrast his own proposal with older, Spencerian forms of EE, Ruse calls his view "Darwinian Ethics" and condemns "Evolutionary Ethics" for both ignoring the difference between prescriptive and descriptive facts and relying on evolution as a guide to life. I will resist using Ruse's terminology here, and instead refer to all attempts to blend evolutionary theory and ethics as 'Evolutionary Ethics'. When referring to theories that (mistakenly) attempt to use evolutionary facts as sources of guidance I will try to either use the term 'prescriptive forms of EE', or 'directly prescriptive forms of EE'. The distinction to which Ruse wants to draw attention is important (although perhaps not quite as cut and dry as he assumes), but appealing to Darwin's name to help explain the distinction only adds unnecessary confusion to the area by setting up an odd contrast between "Darwinism" and 'ordinary' evolutionary theory. The contrast may make sense in the context of Ruse's book, but once this context is removed the contrast becomes potentially misleading.
approaches to uniting evolution and ethics: instead of using evolutionary theory as a guide to justified moral behaviour, he suggests that we turn inward and examine the evolutionary origins of our own unique capacity for moral sentiment. By focusing our attention on the mechanisms within human nature that allow us to think in terms of prescriptive judgments, Ruse hopes to gain access to a form of justification that is capable of avoiding the sharp effects of Hume's Law.

However, as Ruse admits, it is not immediately obvious how examining the origins of our dispositions for moral behaviour can contribute to the justification of these dispositions. One might be willing to acknowledge that epigenetic rules exist which help to shape our moral capacities yet claim that something more is required in order to provide a credible justification for the authority of these capacities. No matter how much information we have about the composition of our moral nature, it seems we need to appeal to some other authority to tell us why we ought to obey the impulses that our genes place within us.

Ruse's answer to this challenge is to assert that no further authority need be invoked in order to justify our basic moral traits. Once we have a detailed understanding of our evolutionary moral nature, then, according to Ruse, Hume's Law can no longer be wielded with the same ferocity. The reason this is so, he maintains, is that discovering the evolutionary origins of our moral capacities confirms that nothing more than a naturalistic understanding of these capacities is possible in terms of meta-ethical justification. By demonstrating that our genes are responsible for making us believe ethics requires external standards, an understanding of our evolutionary moral nature demonstrates that any such standards must be illusory (or at least superfluous), and this move subsequently removes the expectation of a prescriptive authority existing beyond the sense of duty generated by our own genes. Thus, the "end run" Ruse hopes to achieve is to identify our genetic social dispositions as the ultimate source of morality's uniquely prescriptive character and then firmly assert that this aspect of human nature gives us "all
the justificatory insight possible" given what we now know about morality's biological origin. (TDS p. 256)

However, there still seems to be room to feel uncomfortable with Ruse's method of sidestepping Hume's Law. The idea that no external authority exists to justify our moral capacities may be plausible, but are we not still in a position to reject the dispositions our genes produce in us if we choose to do so? Ruse claims we are not. He argues that it will be exceedingly difficult for humans to purge themselves of their sense of moral obligation. Moreover, he argues that it is unlikely we would ever want to rid ourselves of this part of human nature.

Morality is part of human nature, and [...] an effective adaptation. Why should we forego morality any more than we should put out our eyes? I would not say that we could not escape morality - presumably we could get into wholesale, anti-morality, genetic engineering - but I strongly suspect that a simple attempt to ignore it will fail.

(TDS p. 253)

Thus, we may wonder why it seems like such a struggle to live up to our moral commitments if ethical dispositions are a direct function of human nature, but Ruse argues that no matter how much we stray from the common moral path we will always return to certain universal moral imperatives (and a shared belief that objective justification of these imperatives is possible) when we evaluate codes of moral conduct. We may be in error when we act as if there is some external standard grounding our ethical judgments, but it is an error deeply embedded in our biological makeup and one we cannot easily escape. In the same way that we have opposable thumbs, Ruse believes that we have an innate capacity to think and act as though ethical obligations carry binding prescriptive force, and this fact about human nature is ultimately all that can, or need, be offered as justification for living up to our moral commitments.
II. Levels of Moral Agency and Prescriptive Authority

Ruse's approach to Evolutionary Ethics is provocative and engaging, since it provides a bold but especially clear articulation of the opinion that our moral capacities are biologically adaptive aspects of human nature. Yet his theory is nonetheless problematic. Is it really the case that the illusion of objectivity is absolutely necessary for successful moral behaviour to occur?\(^\text{10}\) And is there really nothing more we can appeal to in matters of ethical justification beyond an evolutionary account of our social dispositions?

In this section I want to set aside the issues related to Ruse's appeal to error theory in order to examine the more innovative thesis underlying his theory: the idea that an evolutionary account of moral agency can serve as the basis for a meaningful source of justificatory authority. I take this to be the critical thesis operating within Ruse's work that deserves careful attention as an example of the more general aim of constructing prescriptive versions of EE. It is this underlying theme that situates Ruse firmly within the contemporary literature being written on EE and therefore deserves close inspection.\(^\text{11}\) This theme is, of course, closely connected to his use of error theory and the claim that morality requires the semblance of objectivity to be adaptive. But the core supposition that the evolutionary forces responsible for shaping our social nature can ground the prescriptive authority we expect from justified moral claims is the hypothesis that makes Ruse's theory characteristic of current approaches to EE and thus worthy of serious consideration.

Unfortunately, this core supposition - that prescriptive authority can be attributed to our inborn sense of moral agency - is in Ruse's work impressively difficult to draw out in detail. Ruse is an incredibly clear writer, but he is evasive when it comes to providing

\(^{10}\)See both (Waller 1996) and (Campbell 1996) for examples of philosophers who have attempted to salvage the potential in Ruse's work without committing themselves to the claim that we need to believe in false objectivity.

\(^{11}\)This theme can be seen in the work of the two most prominent advocates of EE other than Ruse: Robert J. Richards and William A. Rottschaefer. See Richards (1986) and Rottschaefer (1991), (1998), with David Martinsen, (1990).
a step by step account of the way in which our moral capacities acquire justificatory power. Instead, the strategy implied by Ruse's methodology is one of default: we have no external standards to appeal to in ethics, yet we clearly do not want to lose the benefits of thinking that ethical judgments demand prescriptive authority; hence, the prescriptive authority we seek must be situated in the genetic dispositions that make this aspect of our moral intuitions possible. The best interpretation of Ruse's argument is thus that our own capacities for thinking in terms of binding ethical commitments acquire legitimate prescriptive authority because we would not want to (and cannot easily) relinquish this invaluable, deeply embedded part of ourselves.

Specificity and Ethical Sophistication

Attributing prescriptive authority to our genetic social capacities because nothing else meets the proper requirements is not an immediately compelling justificatory strategy. However, there is an important kernel of truth to Ruse's methodology. If a naturalistic understanding of human moral agency can explain the full extent of those underlying intuitions that serve as the foundation of our capacity for ethical judgment, then in a sense this understanding captures something so essential about human nature that it may rightfully deserve to carry prescriptive weight. Ruse's method of circumnavigating the is/ought distinction seems dubious when we think of 'mere' descriptive facts about our evolutionary origins taking on a prescriptive role, but if these facts represent a part of human nature that is utterly fundamental to our survival, integrity and general well-being as biological agents, then the prescriptive authority of our moral dispositions may be self-explanatory. In effect, a naturalistic account of moral agency

---

12 Moreover, this is essentially the methodology employed by Hume in An Enquiry Concerning the Principles of Morals. (1975) The evolutionary theorist may be working with a more sophisticated set of empirical tools, but the project of uncovering the fundamental empirical attributes of human moral nature as a means of justification is strikingly similar. The comparison is not lost on Ruse; see TDS (p. 87) and (Ruse 1990, 70-74).

13 This seems especially true if we follow Ruse in believing that it is nearly impossible to escape the constraints imposed on us by our genetic moral sense. Like the Kantian view of our capacity for practical
arguably describes something so compelling about what it is to be human that the facts it uncovers ought to surely command our respect and obedience.

Yet this approach is misleading in at least one crucial respect. The problem emerges when we attempt to identify the particular genetic mechanisms operating within moral agency that are ultimately responsible for securing the prescriptive authority of our moral intuitions. Ruse's solution is to suggest that we share a common set of epigenetic rules, often referred to as a single, universal "moral sense", which is then described as having all the justificatory power we can hope to achieve considering what we know about our evolutionary origins. But it seems highly implausible that we possess any one distinct, cleanly delineated set of genetic rules for moral action that can be uniquely marked as having its own justificatory power. It seems much more likely that the kind of moral agency to which we ought to attribute this power will be the result of a highly integrated system of genetic, cultural and psychological forces that combine to create the dispositions that we subsequently describe as being moral. The problem, then, for prescriptive versions of EE that rely on a principally sociobiological understanding of moral agency is that the larger the role we attribute to forces other than our genetic dispositions, the less reason we have to think that the genetic basis for our moral capacities is ethically sophisticated enough to deserve a directly prescriptive role in the justification of moral claims.¹⁴

¹⁴Note that I am not claiming that our genetic capacities are incapable of producing sophisticated social behaviour. They most certainly can. The point is that it is extremely unlikely that the behaviour they produce - no matter how spectacular it is - will match up exactly with what we collectively refer to as specifically ethical behaviour. It is instead broadly social behaviour and therefore what I describe as ethically unsophisticated. This point will, I hope, become more clear as the chapter progresses.
Consider the structure of Ruse's account of moral agency. Ruse adopts E.O. Wilson's sociobiological perspective wholeheartedly enough to interpret our genetic moral capacities as a coherent set of epigenetic rules that set limits on how cultural factors can affect our behaviour. For Ruse, the limits our epigenetic rules impose upon us are impressively sophisticated: as human beings we share a set of distinct, recognizable moral sentiments, and we are constrained to think in terms of objectively binding prescriptive norms (presumably no matter what culture we find ourselves in). This description of our moral nature may seem plausible since Ruse is cautious enough to speak in terms of broad genetic constraints and admit, when necessary, that culture plays an important role in the final product of our moral judgments. But despite Ruse's cautious language, his account requires a disturbing amount of complexity to be built into our basic genetic dispositions if they are to end up constraining our behaviour in very specific ways. Culture is left to fill in certain key details, but the constraints within which culture and other environmental considerations are allowed to operate must be suspiciously narrow if such recognizable traits as thinking in terms of objective moral judgments emerge from our shared genetic heritage.15

Once we appreciate the level of sophistication being attributed to our genetic capacities, it becomes clear that Ruse is proposing a suspiciously naive conception of how human moral agency has evolved. It may be convenient to think that we have a discrete, recognizable sense of moral agency at the genetic level, since this initial moral capacity can then fulfill the role of an essential (and largely inescapable) part of human nature. However, the idea that we share a philosophically equipped genetic moral sense is surely a descriptive shortcut that covers over the complexity involved in our moral development and the interactions between our genes and what eventually comes to be recognized as our final moral capacities. The shortcut itself is not grounds for complaint,

15At one point Ruse even goes as far as to speculate that we "have the Categorical Imperative, or something very much like it, embedded in an epigenetic rule." (TDS 244)
for one cannot address the intricacy of ontogeny at every possible moment without drowning in detail. But because this shortcut is responsible for lending credibility to Ruse's claim that we can attach prescriptive force to our genetic social dispositions, it incorrectly leads us to believe that we can ascribe justificatory power to a convenient abstraction dressed up as an essential part of human nature.

Ruse's willingness to rely on this shortcut is indicative of the fact that the methodology of attributing prescriptive authority to our genetic social capacities runs the risk of providing an overly intellectualized account of the genetic contribution to moral agency. Consider an alternative approach to the idea that we possess recognizable genetic moral capacities: what if our genes simply provide us with more abstract emotional control mechanisms? Rather than setting constraints at the level of advanced ethical concepts, our genes may be able to significantly affect our behaviour by equipping us with powerful, but far less intellectually sophisticated, moral sentiments such as sympathy, pride and shame. In fact, it seems likely that the constraints we associate with ethical behaviour are grounded in these less scholarly emotions, even if we tend to classify their combinations and the results they achieve in complex theoretical ways. In short, epigenetic rules may be capable of building an effective basis for moral agency out of impressively unsophisticated emotional components - components that are broadly social and only later selectively interpreted as ethical traits and then retroactively interpreted as uniquely specialized moral dispositions.

This possibility creates a significant problem for the strategy of grounding a plausible sense of justificatory authority on a distinctly genetic basis for moral agency. How can we attach a meaningful sense of prescriptive authority to the sociobiological origins of our moral traits if these origins exist only as indistinct emotional constraints? Our genes may supply us with remarkably complex social capacities, but it is unlikely that epigenetic rules produce anything as awe-inspiring as distinct, immediately recognizable sets of decidedly moral dispositions.
At present, there may be no way to settle the matter. But a healthy sense of caution when dealing with complex biological systems ought to prevent us from appealing to a Panglossian solution that locates the bulk of what needs to be explained within convenient genetic dispositions. Such solutions are seductive, but they blind us to the complexity lurking underneath our fully developed moral capacities. Thus, if we carefully consider the evolutionary origins of moral agency, we ought to be suspicious of evolutionary accounts of human nature that draw on pre-equipped genetic moral capacities in order to establish a plausible sense of prescriptive authority.

Two Responses

Ruse has two possible responses to the charge that his account of moral agency is too simplistic to command prescriptive authority. The first response is to accept that epigenetic rules set more diffuse moral constraints on behaviour and still retain the overall structure of the argument that our genetic capacities are capable of holding prescriptive weight. Thus, Ruse might be willing to back away from the claim that our genes affect us in ethically sophisticated ways,16 without abandoning the idea that our genes provide us with something valuable enough to act as a direct source of prescriptive authority. This is a concession Ruse may be willing to make, since he sometimes claims that he is exclusively interested in meta-ethics and the ultimate foundations of morality when it comes to matters of justification. The justificatory authority generated by this response may be more abstract, but if Ruse is only interested in meta-ethical justification he may not see abstractness as a significant problem.

This is perhaps the most plausible option available to Ruse. But even this more abstract formulation of his position underestimates the obscurity of the prescriptive conclusions that follow from a sober interpretation of our genetic moral capacities. Even

---

16E.g. that our genes carry direct implications for questions of substantive ethics related to utilitarianism or Kantianism. See (Ruse 1986, 235-50).
at the level of meta-ethics, Ruse faces a serious underdetermination problem if the genetic dispositions supporting our sense of moral agency consist only in primary emotional responses (e.g. empathy or feelings of kin-related sympathy). These basic emotional responses are undoubtedly essential for the development of any well functioning sense of moral agency, but in terms of their specific contribution to the justification of moral claims, these responses offer very little guidance until they are focused and interpreted by our higher levels of reasoning and our cultural contexts. In short, if we stripped our genetic 'moral' capacities of all scholarly reconstructions, they would likely reduce to general social capacities that lead to a broader range of behaviour than we would want to refer to as even meta-ethically justified. The underdetermination problem facing this response is thus that our genetic social capacities are, by themselves, unsuitable candidates for legitimate justificatory authority.\(^{17}\)

The second potential response Ruse could employ against the challenge that he relies on an overly simplified conception of our 'moral sense' would be to extend the scope of the prescriptive authority in question beyond the genetic components of moral agency. While retaining a thoroughly naturalistic account of the forces leading to our developed moral capacities, Ruse could choose to avoid the underdetermination problem associated with our genetic capacities by accepting the fact that the complicated interactions between epigenetic rules, cultural forces and the higher levels of our evolved psychology play a key role in constructing the characteristics of human nature that deserve justificatory power. This option involves conceding that epigenetic rules alone do not set constraints specific enough to lead to distinctly ethical responses, forcing Ruse to admit that speaking in this way perpetuates a convenient but misleading shortcut. However, the response still allows him to defend a source of prescriptive authority that is

---

\(^{17}\)For example, the genetic disposition for kin-based sympathy is only as morally justified as our willingness to selectively disregard the closely related disposition to ignore, or even seek to harm, non-genetically related dependents.
capable of operating at an appropriately sophisticated ethical level - a significant accomplishment if EE is to provide a credible basis for moral justification.

This second response is an interesting option, as it allows EE the ability to attach prescriptive authority to meaningful sets of ethical constraints rather than less clearly defined (pre)moral sentiments. However, the response is less viable than it appears. The reason it is inadequate is that once the focus on primary genetic dispositions is dropped it is no longer clear that the resulting justifications for moral claims would be in any way uniquely evolutionary. Of course, in a certain sense anything we appeal to can be considered evolutionary, since every aspect of human nature (and every aspect of culture and cognition) is arguably the result of some general evolutionary process. In other words, every aspect of the human species and the natural world can be said to have evolved from an initial primordial soup. But if the only sense in which a theory of EE can be said to be drawing prescriptive conclusions based on evolutionary theory is by reverting to a general faith in Darwinism, then there is hardly anything meaningful left to separate prescriptive forms of EE from other naturalistic ethical frameworks. By allowing sophisticated levels of moral agency to enter into what becomes identified as the elements of human nature deserving justificatory authority, this second strategy succeeds only in watering down the impact of prescriptive EE to such an extent as to render the project unremarkable.

This is not to say that the broader details of how cultural and our higher level cognitive capacities have evolved are not intriguing. The issue is whether the levels of biological machinery involved in constructing moral agency can act as a uniquely evolutionary yet direct source of prescriptive authority. Once the question of whether we can attach prescriptive authority to the mechanisms of moral agency is separated from our general interest in investigating how these mechanisms function, it quickly becomes apparent that the only aspects of moral agency that are legitimately capable of supporting
prescriptive ethical authority are those that are sophisticated enough to render their evolutionary status banal.

Thus Ruse's project fails in one of two ways. Either the components put forth as moral standards are primary genetic dispositions, in which case ascribing justificatory authority to these capacities faces a severe underdetermination problem and ignores the cultural and critically reflective faculties involved in interpreting and identifying our dispositions as discernibly ethical.\(^\text{18}\) Or else the components proposed are made up of genetic, cultural and rational factors, in which case the strategy of grounding a prescriptive source of authority on evolutionary theory reduces to a common form of naturalism and effectively removes what is supposedly unique to the project. Either way, no matter how interesting the sociobiological forces underlying our moral agency may be, the sophisticated character of moral judgment suggests that questions of justification remain beyond the immediate reach of distinctly evolutionary facts.

\textbf{III. A Valuable Empirical Resource}

If a position as carefully articulated as Ruse's version of EE is unable to successfully ascribe a sense of prescriptive authority to the evolutionary origins of moral agency, then I think we have good reason to suspect that the enterprise is generally destined to share a similar fate. However, writing off the evolutionary origins of our moral dispositions as being wholly irrelevant to the generation of prescriptive claims seems suspiciously underwhelming considering how informative the sociobiological details of human behaviour might be if the discipline reaches its full potential. If evolutionary theory manages to uncover intriguing facts about our shared social capacities, will these facts really have no direct impact on the way we ought to construct effective ethical principles?

\footnote{\textit{Gewirth (1986)} articulates a similar point about underdetermination in a response to Robert Richard's (1986) version of Evolutionary Ethics.}
The key to alleviating the disappointing flavour of this conclusion lies in recognizing that evolutionary theory can make a significant contribution to the sanctioning of our ethical claims without acting as a direct source of prescriptive authority. Since the debate over EE has thus far revolved almost entirely around the justificatory role of evolutionary facts, one is often left with the impression that EE can not serve any worthwhile function unless it is capable of providing its own unique source of justificatory power. But this assumption is clearly false. To deserve to be included in the category of 'Evolutionary Ethics', a theory need not appeal to a specifically evolutionary source of prescriptive value. Instead, a meaningful evolutionary approach to ethics can operate within a more general naturalistic perspective that draws on rationally determined sources of prescriptive value yet appeals to the evolutionary origins of human nature as an especially powerful source of empirical information. A convincing theory of EE can adopt a thoroughly naturalistic conception of the way human moral agency has evolved, yet nonetheless claim that the levels of moral agency we recognize as being manifestly evolutionary are best suited to informing our ethical decisions in specially insightful ways rather than acting as authoritative moral standards.

This more cautious approach to EE has been proposed by several writers in the course of objecting to the use of evolutionary facts as a source of moral value.\(^\text{19}\) However, the option is usually offered as a kind of lame concession to the attempt to use evolutionary theory in more ambitious ways and is rarely discussed in much detail.\(^\text{20}\) The result is that the prospect of developing a substantial theory of EE using evolutionary theory as an empirical resource is often set aside as a comparatively dull outcome. Yet this impression is only deserved if one continues to hold unrealistic assumptions about

\(^{19}\text{See (Duff, 1988), (Bateson, 1989), (Rachels, 1990), (Farber, 1994).}\)

\(^{20}\text{The exception seems to be philosophers especially interested in the welfare of non-human animals. Peter Singer (1981) and James Rachels (1990) maintain a clear distinction between ethical value and evolutionary fact, yet they both defend strong positive theses concerning the moral status of non-human animals based on the empirical facts of our evolutionary origins.}\)
EE's ability to provide a unique means of justification. Moreover, the assumption that this more modest approach to EE is a comparatively dull outcome fails to realize just how useful evolutionary theory might be in terms of helping us identify the unacknowledged ways in which our social dispositions affect our moral judgments and then allowing us to selectively promote or resist these evolved dispositions according to our heightened awareness of their ethical implications. The most viable form of EE may therefore be carrying an unjustly banal reputation because of our tendency to underestimate the impact evolutionary forces have on our social behaviour.21

The Genealogy of Social Dispositions

In fact, the strategy of using evolutionary theory as an empirical resource could prove to be tremendously helpful in terms of enabling us to consistently systematize the authority we attribute to our social dispositions. Evolutionary theory may 'only' be able to act as a 'mere' source of empirical information. But if we can develop plausible explanations for the evolutionary origins of our social dispositions, this information will supply us with a useful source of knowledge about ourselves as moral creatures. By revealing the biological history of those dispositions that serve as an underlying platform for our moral faculties and intuitions, evolutionary theory provides us with a source of information that will allow us to organize and reevaluate our genetic social attributes in potentially dramatic ways. This might be accomplished, for example, by examining the adaptive functions of our genetic capacities (and the conditions that led to their selection)

---

21 As an example of the impressively subtle ways our genes can affect our behaviour, consider the following experiment performed by Claus Wedekind et al. (1995) to investigate the relationship between mate selection and human major histocompatibility complex (MHC). MHC gene products play an important role in immune response, and the immune systems of individuals heterozygous at MHC loci are thought to have a higher fitness because of their advantage in fighting disease. Wedekind's experiment had men sleep in a fresh new cotton shirt for two nights in a row. The shirts were then placed in individual cardboard boxes with holes in the lids through which the shirts could be sniffed. When women were asked to sniff the T-shirts of six different men (three with MHC genotypes similar to her own and three with MHC genotypes dissimilar to her own), the results showed that women preferred the odors of men whose MHC genotypes were dissimilar to their own by a statistically significant margin! Though this example is not related to ethical behaviour, it speaks volumes about the potential for underlying sociobiological factors to explain our seemingly spontaneous behaviour.
in order to decide whether these capacities ought to be promoted within contemporary social environments. We may have a general idea of the social capacities evolution has implanted within us, and we may normally assume that these capacities are beneficial enough to be considered justifiably moral. But a detailed account of the evolutionary history of these capacities will allow us to distinguish more accurately between those aspects of human nature that legitimately deserve to be promoted and those that have evolved for reasons that are no longer compatible with our considered beliefs about what constitutes acceptable ethical conduct.

The information supplied by evolutionary theory can therefore act as an elaborate *genealogical* resource for ethics by exposing the evolutionary reasons why we possess the social intuitions we find deeply embedded in human nature. By exposing the history and adaptive characteristics of our social dispositions and the intuitions these dispositions facilitate, this evolutionary genealogical information could allow us to discriminate between aspects of our evolutionary heritage possessing clear ethical value (e.g. the capacity for empathy) and those we ought to consider natural aspects of human nature that are nonetheless ethically unacceptable (e.g. the capacity for vengeance or sexual dominance). Considering the powerful grasp our social dispositions undoubtedly exert over our intuitions, the ability to examine the evolutionary genealogy of these dispositions will be an extremely useful tool for the task of systematically organizing the authority we attribute to our biological moral nature.

*A Valuable Flexibility*

The key advantage to using evolutionary theory as a source of empirical information is the ability to recognize the unique nature of our evolved social dispositions without abandoning the capacity to distinguish between those parts of our genetic heritage we accept as being morally valuable and those we think ought to be suppressed. Thus, examining the genealogy of our evolutionary origins allows us to incorporate *all*
the insight worth salvaging from prescriptive theories of EE without committing ourselves to assigning a justificatory role to every aspect of human nature we identify as a genetic social disposition. This more viable approach to EE provides a crucial sense of flexibility to the project of interpreting moral agency from a naturalistic perspective. In the case of obviously beneficial genetic dispositions, we can simply confirm the moral status of these dispositions and attribute as much prescriptive authority to them as we see fit. A limiting case might involve dispositions so obviously advantageous and deeply ingrained in human nature that they represent the genetic moral capacities described by Ruse as carrying "all the justificatory insight possible" considering what we know about our biological origins (e.g. our propensity to care for our children). But if the difference between Ruse's project and the more moderate approach to EE seems trivial in this instance, it quickly becomes important when the relevant aspects of human nature become more complex and we want to avoid mistakenly attributing ethical value to social dispositions that are connected to moral traits but are nonetheless unsuitable candidates for carrying prescriptive authority (e.g. a propensity for xenophobia that is intimately linked to praiseworthy dispositions for group loyalty).

This is an important option when we realize just how complex our moral agency is once it is broken into smaller components. It may be convenient to speak as if we possess distinct genetic moral dispositions, but as we have seen this is certainly a heuristic shortcut that covers over the detail lurking beneath the surface of our fully developed moral traits. Even seemingly straightforward social dispositions, such as 'caring for one's family' and 'willingness to respect community standards', contain components we would not want to sanction as ethically justified. (The former may include dispositions for excessive partiality at the expense of others and the tendency to disregard stepchildren as less valuable offspring. The latter can lead to cases where an eagerness to obey community standards manifests itself in uncritical support for abhorrent political or cultural practices.) The capacities we describe as being the
distinctly moral aspects of human nature are thus almost certainly linked to other closely related social capacities we would be much less enthusiastic to support. The ability to apply a flexible source of prescriptive authority to the evolutionary details of moral agency is therefore an indispensable tool if we want to unravel the puzzle of determining which aspects of our genetic social dispositions truly deserve justificatory power.

**Conclusion**

Although the significance evolutionary theory holds for ethics is often dismissed by philosophers for undeserved reasons, the strategy of grounding a directly prescriptive form of Evolutionary Ethics on the details of our evolved sense of moral agency is nevertheless deeply misguided. Using Michael Ruse as a representative of the current approach to EE, I have argued that ascribing a sense of prescriptive authority to the biological basis of moral agency will either force us to propose unrealistically sophisticated genetic dispositions which surely ignore the intellectual reconstructions involved in identifying specific parts of human nature as distinctly moral, or we will be forced to allow culture and the higher levels of our evolved psychology to play a role in what we distinguish as having prescriptive value - a move that gives up what is unique to the project of outlining a distinctly *evolutionary* foundation for ethical justification. To remedy this dilemma, I have briefly sketched an alternative approach to EE that uses the evolutionary details of moral agency as a source of genealogical information in order to capture the insights of prescriptive forms of EE without committing the same justificatory mistakes. This more moderate approach to EE may seem comparatively dull compared to the bold claims made in the name of EE, but if evolutionary theory continues to uncover extraordinary details about the origins of human nature I think this approach carries the

---

22For example, in their recent work on the biological basis for altruism, (Sober and Wilson, 1998) point out that the conditions necessary for providing a setting in which individuals can develop helping behaviour towards members of their group can also lead to contexts in which it is adaptive to *hurt* members of other groups (p. 9).
potential to dramatically affect the way we construct justified ethical systems without succumbing to the temptations that disrupt this otherwise compelling area of research.

It is important to note, however, that prescriptive versions of EE fail to generate an independent sense of normative authority because they tacitly rely on sophisticated cultural components, and not because they necessarily violate a fundamental difference between 'is' and 'ought'. It is often assumed that EE collapses under the weight of Hume's Law before it even deserves serious consideration, but the argument I have provided against prescriptive EE does not depend on our being able to wield Hume's Law in such a resolute manner. Instead, my argument has the advantage of remaining agnostic about the general meta-ethical question of whether some form of naturalistic ethical justification can close the gap between facts and norms. I claim only that justification based specifically on sociobiology will fail to identify the right kinds of facts to close this gap, because the genetic dispositions they employ cannot capture the customized quality of what we distinguish as discernibly moral. Hence, current versions of prescriptive EE fail, but they do not necessarily fail because they can be dismissed with a passing nod to Hume's Law. They fail because they cannot draw a direct connection between our broadly social genetic tendencies and the particular character of our justified ethical intuitions.
Chapter 2

Cultural Evolution and the Move Towards Memetics

Introduction

Despite the bold claims and heated political controversy surrounding the promise of sociobiology, chapter one has shown that the details surrounding our genetic capacities for social behaviour will not lead to radical changes for the way we justify our ethical principles. Instead, the advances we achieve in identifying our sociobiological dispositions will supply ethics with a useful source of empirical data - one that will enable us to recognize the genealogy of the genetic forces contributing to human social action. These genetic forces do not map on to our informed moral intuitions accurately enough to support an independent basis for prescriptive authority, but they almost certainly influence the content of our ethical claims in subtle but remarkable ways.

However, if sociobiology provides only a partial explanation (albeit a very interesting one) for the rich detail present in our ethical judgments, will evolutionary theory understood more generally be similarly limited from accounting for this detail? In chapter one the key focus of the argument against prescriptive Evolutionary Ethics was that sociobiology is too blunt an instrument to serve as a foundation for the ethical sophistication of our moral intuitions. But if our intuitions cannot be fully captured by genetic dispositions alone because of the input required from cultural and psychological factors, then must we assume that our moral capacities cannot benefit from evolutionary analysis because the cultural and psychological components they possess remain beyond the immediate reach of genetic explanation?

This assumption becomes questionable once the general structure of evolutionary theory is disconnected from the limitations of sociobiology. The cultural and psychological components of our moral capacities are related to evolutionary theory in a most unremarkable manner when it comes to the issue of whether or not their
evolutionary origins can support a unique source of direct prescriptive justification. However, if we take up a naturalistic approach towards cultural and psychology as empirical subjects, it is nonetheless possible that evolutionary theory will be a helpful explanatory tool for investigating the form and construction of these important aspects of moral agency.

In fact, new research suggests that cultural and psychological systems operate according to an independent system of evolutionary transmission. If this is the case it is no longer clear that ethics ought to remain unaffected by broadly evolutionary concerns. If we can extend evolutionary theory beyond the confines of sociobiology and successfully apply formal evolutionary principles to the cultural and psychological aspects of our moral intuitions, we will have good reason to think that ethics ought to be deeply influenced by evolutionary (but not strictly genetic) considerations. This would be a significant advance for our understanding of the origins of our moral capacities. If investigating our genetic social dispositions provides valuable empirical information about the genealogy of our moral intuitions, then an investigation into the evolutionary character of our culture and psychology will surely yield equally profitable results.

Thus, in this chapter I explore the existing literature on cultural evolution, the hypothesis that cultural information can be transmitted socially and selected according to rules implied by the formal structure of (biological) evolutionary science. The study of cultural evolution opens up an intriguing area of research based on the idea that certain abstract evolutionary principles, such as selection, heredity, and coded information transmission, can be applied to more than just biological phenomena. Rather than focusing exclusively on the transmission of genetic information, the premise supporting cultural evolution suggests we can examine the structure of cultural transmission as a parallel system of inheritance in which discrete patterns of cultural information are passed on to our (social) offspring and then selected for according to the pressures
implied by cultural standards of fitness. This system of inheritance will definitely be closely linked to our genetic inheritance system. (E.g. cultural traits may depend on genetic dispositions for their cultural fitness.). But if cultural evolution can demonstrate an explanatory independence from its underlying biological platform (i.e. if it produces outcomes that cannot be explained using only biological evolutionary considerations), then this process will most certainly be worth investigating if we want to capture the full scope of the evolutionary forces influencing human social behaviour.

Current attempts at modeling cultural evolution have proven moderately but not positively successful. Questions remain with respect to how faithfully we can apply the science of evolutionary theory to the conceptual medium of culture. The use of evolutionary terminology may seem alluring, since we often speak as if changing cultural patterns 'evolve' over time. However, the intuitive appeal of this terminology must be more substantially supported by further evidence. As critics have been quick to point out, the similarities between the mechanisms responsible for biological evolution and those responsible for cultural change may not be strong enough to support the hypothesis that both media operate according to the same set of transmission and selection rules. The language of cultural evolution carries a seductive potential, but it remains to be seen whether the benefits of an evolutionary cultural perspective outweigh the problems associated with mapping a complex set of biological principles on to what may be a very different realm of phenomena.

Chapter two therefore examines some of the obstacles facing current approaches to cultural evolution and outlines some of the solutions proposed in its defense. Moreover, the chapter attempts to capture the fascinating nature of cultural evolution's potential by isolating exactly what seem to be at stake in the debate over whether or not modeling cultural transmission ought to be considered a worthwhile enterprise. What exactly can we gain by being able to describe cultural phenomena in evolutionary terms? To grasp the significance of the work being done on cultural evolution, it is helpful to
keep in mind what the implications might be if we succeed at producing credible evolutionary models. (The specific implications of these models for ethics are set aside until chapter three.) Hence, the chapter introduces some of the issues surrounding cultural transmission and attempts to draw out how this field of research might have a profound effect on the way we look at cultural information.

Particular attention is given to the newly emerging science of memetics. Memetics represents the most ambitious attempt to model cultural phenomena according to the principles of evolutionary biology. Memetics claims that the basic units of cultural selection, termed "memes", can be identified as discrete packages of cultural information that replicate and then distribute copies of themselves in the abstract environment of cultural infosystems. Intended as functional analogues of our genes, memes represent the basic building blocks of cultural evolution - cultural replicators that multiply, mutate, and are ultimately selected for according to the unique evolutionary pressures present in our social systems. Memes can be traditions, fashions, catch-phrases, melodies, inventions, or lasting intellectual concepts.\(^1\) They represent the raw material of our cultural and psychological experience: primary conceptual units shaped by selection pressures to form our ongoing cultural history. The study of memetics may seem farfetched (and is often poorly represented by popular media), but it has nevertheless become a serious area of research for academics who suspect that it may emerge as a theory with tremendous explanatory power. By examining this hopeful science, chapter two attempts to capture both the problems and spectacular potential theories of cultural evolution seem to simultaneously possess.

The chapter is organized as follows. Section one provides a very brief history of the conditions that led to the study of cultural inheritance as a formal system. Section

\(^1\)See Dawkins (1976, chap. 11) and Dennett (1995, chap. 12) for more examples.
two examines some of the general methodological problems facing the aim of describing culture as an evolutionary phenomenon. It looks at the basic question of why we would want to model cultural systems in terms of evolutionary theory and considers the explanatory advantages this approach can provide over sociobiology. Section three focuses specifically on memetics and describes both its specific methodological problems and intended benefits. This more ambitious approach to cultural evolution is presented as a very immature science, but one with considerable potential and a great deal of heuristic value.

I. Cultural Inheritance

The origins of the hypothesis that culture operates according to an evolutionary system can be traced back at least as far as the nineteenth century. In fact, if one accepts a charitable interpretation of what constitutes cultural evolution, it may be impossible to determine the origins of the hypothesis, since countless authors in the history of philosophy have suggested that cultural ideas develop over time in a discernibly structured manner. However, if we limit the scope of what we mean by cultural evolution to a definition that requires at least a direct connection to Darwinism and natural selection, the hypothesis has obviously only appeared within the past one hundred and forty years. Moreover, if we restrict ourselves to theories attempting to draw a systematic parallel with the structure of biological phenomena, we find that cultural evolution has really only taken on a recognizable shape in the past forty years or less.

The premise that cultural phenomena can be faithfully modeled according to the structure of evolutionary theory is therefore a fairly recent development. After Darwin published *The Origin of Species*, political and sociological writers began to make use of evolutionary language to describe what they perceived to be progressive transformations in cultural organization. But these early theories often fell short of investigating the similarities between the structure of cultural and biological transmission in literal detail.
Of course, this was largely due to the fact that our understanding of the mechanisms driving biological evolution was still not very well defined. It is easy to look back at early conceptions of cultural evolution and conclude (anachronistically) that they presented a distorted view of the analogy between the structure of cultural and biological evolution. Yet early articulations of cultural evolution often reflected the level of understanding of evolutionary theory available at the time. The parallel between biological and cultural evolution may have been taken very seriously, but usually all this parallel produced was a general emphasis on progressive transformations towards increasing cultural complexity. Thus, cultural evolution was presented as a fairly mysterious process, but this was largely due to the fact that the structural forces driving biological evolution were no less mysterious.

Two Developments

Two key developments paved the way for more detailed conceptions of cultural evolution and the emergence of a systematic approach to the parallel between cultural and biological phenomena. First, our understanding of the mechanisms responsible for biological heredity became more sophisticated as scientific interest in the mechanisms of inheritance began to develop over the first half of the twentieth century. Darwin himself acknowledged that heredity plays an essential role in evolutionary theory, but the means by which offspring acquire the characteristics of their parents remained distressingly unclear during the course of his lifetime. Though Gregor Mendel had conducted experiments on the rules of heredity as early as 1865, the importance of these experiments were not fully recognized until the beginning of the next century.

Once the implications of Mendel's work were finally appreciated, a new period of research on the biological basis of heredity began which dramatically altered the nature of

---

2The exception here is perhaps Durkheim who proposed that the evolution towards the cultural division of labour proceeded in precisely the same way as the organs of the body evolved towards increasing interdependence and specificity of design. (Durkheim 1893)
evolutionary biology. Within fifty years, Darwin's theory of 'descent with modification' had been greatly enhanced by several significant discoveries: the detection of chromosomes as the physical agents of heredity, the subsequent development of population genetics and the theoretical concept of the gene, and the discovery of the biochemical basis of information coding within the unique self-replicating structure of DNA. What was once a somewhat abstract account of the mutable nature of species had now become an extremely rigorous area of scientific investigation.3

The second key development that contributed to a more detailed parallel being drawn between the structure of cultural and biological evolution was the emergence of ideational theories of culture in sociology and anthropology. In the 1960's and 1970's, the social sciences began to turn towards increasingly conceptual definitions of culture as academics began to acknowledge the semiotic nature of cultural information. The most prominent advocate of this approach was (and still is) Clifford Geertz, who defined culture as:

an historically transmitted pattern of meanings embodied in symbols, a system of inherited conceptions expressed in symbolic forms by means of which men [sic] communicate, perpetuate, and develop their knowledge about and attitudes toward life.

(Geertz 1973, p. 89)

Culture came to be recognized as a system of conceptual control mechanisms that could be passed on as representational symbols and used to guide our beliefs (and ultimately our behaviour) in socially significant ways. The ideational understanding of culture was originally developed much earlier than the second half of the twentieth century (with such prominent sociological figures as Emile Durkheim, Max Weber, and Talcott Parsons).4 However, symbolic interpretations of culture did not gain widespread popularity until much later when the social sciences themselves began to attract greater academic

---

3 For an account of the history of this development, see (Judson 1979) or (Mayr 1982).
4 See (Durkheim 1895), (Weber 1964) and (Parsons 1937).
attention. Before the rise of symbolic definitions, culture was often viewed as a system of concrete actions and behavioural patterns - a methodology that ignored the fact that these patterns contained meaning via their representational content. Thus, the switch to ideational theories allowed those studying culture to focus explicitly on the beliefs being passed to further generations, instead of strictly concentrating on propagation of complex combinations of social behaviour.

It may seem odd that the move towards studying cultural ideas rather than behaviour would constitute a major advance for evolutionary approaches to culture, because from a biological perspective evolutionary theory seems far better suited to dealing with distinct actions than the more abstract landscape of cultural ideas. However, the development of ideational theories in the social sciences allowed evolutionary principles to be applied to a whole new system of inheritable information and greatly strengthened the idea that evolutionary rules of transmission could be generalized beyond their biological platform. Culture had so far presented only an obstacle for sociobiological accounts of human behaviour because it interfered with straightforward genetic predictions, but this obstacle soon became an asset once evolutionary principles were extended to the coded information present in cultural symbols. As William H. Durham describes this transformation:

The development of ideational theory in anthropology re-emphasizes that human beings are possessed of two major information systems, one genetic, and one cultural. It forcefully reminds us that both of these systems have the potential for transmission or "inheritance" across space and time, that both have profound effects on the behavior of the organism, and that both are simultaneously co-resident in each and every living human being.

(Durham 1991, p. 9 original emphasis)

In effect, once culture came to be understood as the transmission of meaningful information, a whole new perspective became available where cultural concepts could be
interpreted as Darwinian units operating under a parallel but potentially independent system of evolutionary inheritance.

II. Modeling Cultural Evolution

With a more technical understanding of the mechanisms responsible for genetic heredity available and the suggestion that cultural transmission provides us with a separate set of inheritable instructions, the 1970's and early 1980's saw a virtual explosion of interest in the idea that culture can be accurately modeled as an evolutionary phenomenon. After a long period of gestation, cultural evolution had finally come to be studied in a more exacting manner as theorists began to examine the evolutionary dynamics of social transmission in much greater detail. In fact, even though this research was conducted by individuals trained in a diverse range of disciplines (including anthropology, evolutionary biology and ecology), the accounts of cultural transmission developed during this period display a notably similar interest in achieving an unprecedented level of precision. Whereas prior accounts of cultural evolution used evolutionary language quite liberally, modern accounts attempt to provide methodical, often quantitative, descriptions of how cultural transmission might operate according to a strictly Darwinian perspective.

Three Challenges

The move towards methodical accounts of cultural evolution opens up an intriguing potential for the study of social inheritance systems, but it also faces some important challenges before these accounts can claim to have successfully outlined the full scope and character of cultural transmission. Some of these challenges involve technical difficulties - specific problems that arise for the task of building working models of cultural units and the transmission functions these units obey. Others represent

---

more general methodological problems that raise questions about the utility of trying to model cultural phenomena from an evolutionary perspective.6

Of these challenges, three have emerged as being especially deserving of attention: 1) Can cultural evolution be reconciled with biological evolution so that the former is consistent with a Darwinian understanding of the origins of our system of cultural inheritance? 2) If cultural evolution is consistent with our genetic system of inheritance, then can cultural evolution achieve an explanatory independence from biological standards of fitness, or will it provide us with the same explanations for human social behaviour as sociobiology? and 3) If cultural evolution can achieve explanatory independence, does the structure of biological evolution provide the appropriate tools to investigate cultural phenomena? The first two challenges are closely related (the second being the more important of the two). The third involves technical issues which will be set aside until section three, where they will be discussed in the context of memetics.

By drawing attention to these challenges it may seem as if my aim is to give a pessimistic forecast for the explanatory potential of modeling cultural evolution. This is not the case. Modeling cultural evolution is a fairly recent enterprise, and I think it shows considerable promise given its current stage of development. Rather, I think that by examining the challenges and the ways in which models of cultural evolution respond to them, we can gain a better idea of how the study of cultural evolution can contribute to our understanding of our social capacities.

Adaptive Origins

The initial challenge for the aim of modeling culture as an evolutionary system is to demonstrate that cultural transmission has evolved from what was originally a purely genetic system of inheritance. We may appreciate the current benefits associated with

6See (Durham 1991) for an impressively detailed overview of the challenges and controversies facing the different methods of modeling cultural inheritance.
our status as cultural beings (to put the point mildly), but how did this capacity first arise? To be consistent with a Darwinian picture of culture having emerged from biological origins, cultural transmission must be understood as having evolved for fitness-enhancing reasons. In other words, if a cultural system of inheritance has evolved like other naturalistic traits, it must have conferred a reproductive advantage to those individuals who first acquired the ability to pass social information to their peers. It cannot be useful for its own cultural reasons - it must also be biologically adaptive.

The question is important because if cultural inheritance cannot be shown to be biologically adaptive, we may have less reason to attempt to model it as an evolutionary phenomenon. For example, if culture simply emerges from its biological substrate as a contingent supervenient event (one that perhaps arises once a certain amount of brain mass is achieved), then it seems less likely that it operates according to rules of distinctly evolutionary transmission. In the extreme, if one views culture as a miracle that has dropped out of the sky, then one would likely see no reason at all to expect it to conform to a Darwinian structure. Similarly, even if culture can be reconciled within a naturalistic framework, if it is a wholly supervenient process it may be less likely to function according to transmission and selection principles that we observe in evolutionary biology. If, on the other hand, culture is an adaptive characteristic that has evolved over many thousands of years, it is more likely to have an intimate connection to the underlying biological principles we recognize in the structure of evolutionary theory. The link here is not decisive. (Culture might be biologically adaptive and still have a wholly different structure, or it might be a miracle and still happen to have a structure similar to biological evolution.) But proving that culture has adaptive origins makes it that much more likely that it is a distinctly evolutionary system of inheritance - one that is fluidly integrated within the overall structure of a comprehensive coevolutionary process.

Hence, many theorists have sought to establish the adaptive value of cultural inheritance as an important first step in the process of modeling culture as an
evolutionary system. The challenge, however, can be met without much difficulty.\(^7\) It is resolved by the fact that the capacity to transmit information to one's offspring via a mechanism more immediate than genetic transmission is extremely advantageous for the reproductive fitness of humans and other social species. The reason for this advantage is the fact that the capacity for cultural transmission allows individuals to economize on the often deadly learning costs associated with identifying adaptive behaviour in hazardous environments.

If organisms are constrained to determining the most adaptive behaviour in their environment through individual learning, then the successful responses discovered by this learning process can only be transmitted to future offspring through natural selection and the survival of those individuals fortunate enough to pass on their genetic legacy. This Darwinian process will eventually produce individuals who have developed adaptive responses to their environment, but it remains an inefficient means of achieving results, since each new individual must once again decipher the hazardous rules of her surroundings through a process of private trial and error. Individuals with parents who have survived this process may be more likely to survive because of their ancestry, but since their parents can only provide them with genetic information, the characteristics they acquire must either be hardwired (and therefore a liability in rapidly changing environments) or too general to help them avoid the costs of having to learn on their own.\(^8\)

Imagine the following hypothetical example: a group of savannah dwelling hominids with moderate social capacities living in an environment that has many

---

\(^7\)Lack of controversy, however, has not prevented theorists from constructing highly detailed models demonstrating that cultural transmission is adaptive. See (Lumsden and Wilson 1981), (Cavalli-Sforza and Feldman 1981) and (Boyd and Richerson 1985).

\(^8\)For example, individuals whose parents survive the learning process in a hazardous environment may have received genes that make them generally more intelligent - a trait that would certainly be beneficial - but this trait would still not prevent them from having to engage in a dangerous trial and error process for themselves.
complex obstacles and dangers. There are particular plants that are poisonous and others that are a good source of food. There are particular snakes that are poisonous and others that are a source of food. There are certain nuts that can be nourishing but must be opened in a particular way. In short, there are a great many useful tricks to be learned in order to live and successfully procreate. If all the hominids have is a genetic system of inheritance, a vast number of hominids will die, generation after generation, attempting to determine which plants are poisonous and which are not, which snakes are poisonous and which are not, and how to open nutritious nuts. Eventually, a strain of hominids might evolve that carries the innate knowledge to stay away from the right kinds of plants and snakes and how to open nuts, but this would take a very, very long time. (Moreover, the hominids are likely to turn out like many of us - afraid of all snakes - and miss out on a potential source of food.) The advantage of a cultural system of inheritance is in this case unmistakable. If parents can teach their offspring what they have learned about their environment, they can allow these offspring to avoid the deadly costs of individual trial and error exploration.

Thus, a system of cultural inheritance allows organisms to escape from an inefficient genetic transmission loop by providing them with a means of directly transmitting learned information to their offspring. Instead of sacrificing vast numbers of individuals to the perils of trial and error learning until one's genes eventually find a way to transmit the appropriate warnings, cultural inheritance allows those who have discovered critical facts about their environment to transmit this information to naive individuals and therefore save them from the danger of having to test their environment for themselves.

*Explanatory Independence*

If cultural inheritance is (generally) biologically adaptive, can the results it produces be accounted for using the same explanatory criteria as sociobiology? The next
challenge for modeling cultural evolution is the question of whether it constitutes a meaningfully independent system of inheritance. Certainly, cultural inheritance is independent of genetic inheritance in the sense that it operates on a different substrate (our minds and institutions rather than our DNA and cells). However, if cultural inheritance has emerged because it offers a survival benefit to social individuals, we may still wonder whether this system produces results that deviate substantially from those that are generated (albeit more slowly) by biological inheritance.

The question is important because if we are ultimately interested in the evolutionary forces shaping our moral capacities, then the value of investigating the structure of cultural transmission as an independent system of inheritance depends on whether it produces unique empirical outcomes. If cultural evolution produces adaptive results more often than not, it may not be worth building elaborate models of cultural transmission if these models end up teaching us the same lessons as existing sociobiological theory (i.e. if they tell us that culture, like many other human qualities, promotes particular beliefs and desires because they are ultimately biologically fitness enhancing). The unique means by which culture arrives at this result may be interesting as a matter of empirical science, but if we go to the trouble of constructing complex models of cultural transmission only to rediscover this basic sociobiological maxim, the result will be underwhelming for our interest in the evolutionary influences acting on human moral agency. Thus, cultural inheritance is certainly a valuable area of empirical research in its own right, but if it cannot achieve an explanatory independence from our genetic inheritance system it will be a less interesting factor in examining the peculiarities of human social behaviour (as well as how these peculiarities affect our basic moral dispositions).

The argument against cultural evolution being able to exhibit explanatory independence is impressively simple: if biologically disadvantageous cultural information distributes itself within a population, then this population will be
reproductively disadvantaged and may not survive long enough for its cultural information to be successfully passed on to further generations. Culture can work against biology for a limited time, but, according to this view, it cannot do so for long enough to make a significant difference for the content of adaptive explanations. Hence, the predictions generated by cultural evolution will not differ on the whole from those offered by sociobiology, and cultural transmission need not be modeled as an independent system of inheritance.

The most prominent articulation of the argument against the explanatory independence of culture has been given by Charles Lumsden and E.O. Wilson in their treatise *Genes, Mind and Culture*. (1981) In this work, Lumsden and Wilson argue that although cultural patterns cannot always be immediately explained by genetic factors, gene-culture coevolution is nonetheless ultimately reducible to genetic fitness because the forces driving cultural evolution are originally based on genetic considerations. The argument is built on the idea that biological processes direct the assembly of the mind, and that the rules governing these biological processes (referred to as "epigenetic rules", as one may remember from Ruse's use of the term in chapter one) impose constraints on the cultural patterns our minds will accept and subsequently propagate. It may seem as if culture has a life of its own, but in the end this seemingly autonomous existence is determined by the underlying biological forces that create the mind and make cultural transmission possible.

Lumsden and Wilson's argument against the explanatory independence of cultural evolution may seem incredulously rigid. However, they do not defend this position without making some important concessions to culture's strength and influence.\(^9\) First,

\(^9\)Plus, *Genes, Mind and Culture* is primarily meant to argue against what Lumsden and Wilson call "the promethean-gene hypothesis" (p. 1-2), the idea that biology creates the capacity for culture but that, once created, culture no longer depends on biology in any significant respect - a view that strips the mind of its genetic predispositions and allows culture to advance in any direction whatsoever. Lumsden and Wilson may be pushing the genetic basis for culture especially hard to demonstrate that gene-culture coevolution involves at least somewhat of an extension of biology into the social sciences. (p. 26)
the view of gene-culture coevolution they defend provides culture with a substantial role in producing evolutionary results. On their view, culture is shaped (and ultimately determined) by biological imperatives, but biological evolution is, in turn, deeply affected by culture's ability to alter the landscape on which genetic selection operates. Culture is said to be dependent on biology, but it remains a key part of an intricate coevolutionary cycle and carries the potential to dramatically alter the genetic fitness parameters acting on a social population. Second, Lumsden and Wilson are careful to acknowledge the obvious fact that culture can vary a great deal within the limits imposed by biological constraints. Cultural patterns may be biologically adaptive, but this fact is consistent with an impressively diverse range of cultural means of achieving adaptive results.

Despite these concessions, Lumsden and Wilson remain firm on the idea that culture is ultimately reducible to biological fitness and that the potential eccentricities present in the structure of cultural transmission are not significant enough to pose problems for the predictions of sociobiology. Culture is said to be tied to biology with an "elastic but unbreakable" leash, which allows some room for variation but otherwise guarantees that cultural inheritance will not deviate from genetic inheritance in any meaningful way. (1983, p. 60)

The argument for culture's dependence on genetic evolutionary forces may seem convincing after the detailed treatment it receives in Genes, Mind and Culture, but an important response has since been offered by Robert Boyd and Peter J. Richerson in their ground breaking book Culture and the Evolutionary Process (1985). In this work, Boyd and Richerson outline a rigorous account of how cultural transmission can be modeled as an evolutionary system. Using the tools of population genetics, the book draws on mathematical models of cultural transmission to examine the consequences of what is labeled a "dual inheritance theory" of gene-culture interaction. In many respects, the theory is not radically different than other approaches to cultural evolution. However, it is unusually clear about the fact that the evolutionary determinants of human behaviour
can be transmitted via two 'structurally different systems of inheritance', (p. 2) and that this unique form of coevolution allows culture to produce outcomes that cannot be explained by a straightforward sociobiological perspective. Boyd and Richerson are meticulous about demonstrating that cultural inheritance is consistent with our underlying biology, and they model it as an adaptive trait that has evolved from natural origins. Nevertheless, they make a powerful case for the hypothesis that cultural transmission can achieve an important form of explanatory independence from the predictions of sociobiology, given the proper circumstances.10

The most convincing results of *Culture and the Evolutionary Process* appear later in the book where specific cultural transmission mechanisms are introduced that can lead to genetically maladaptive outcomes. By establishing detailed models of these mechanisms, Boyd and Richerson demonstrate that the unique structure of cultural heredity allows cultural transmission to operate as a separate evolutionary system, and that in certain situations this system can lead to behaviour that cannot be explained solely by the underlying biological factors that allow this system to be generally adaptive.11

The key to these situations is the adaptive value of social learning and the emergence of *indirect indicators of fitness* that become associated with useful patterns of

---

10 In Boyd and Richerson's terminology, they demonstrate that culture has "population-level consequences" without violating the assumption that cultural inheritance has evolved from natural origins (because of its adaptive value). This argument is meant to contrast with the common sociobiological claim that culture is kept on a long biological leash - a claim they call "the argument from natural origins" which states that "The organic origin of the human capacity for culture under the influence of natural selection acting on genes guarantees that, however it may work in detail, culture will usually enhance genetic fitness." (p. 13, my emphasis)

11 Note, however, that one need not appeal to genetically maladaptive outcomes in order to establish the explanatory independence of cultural transmission. Strictly speaking, one need only demonstrate that cultural transmission can produce different results than a purely genetic system of inheritance. Hence, these results do not necessarily have to be maladaptive. One could instead, for example, appeal to outcomes where cultural transmission corrects the deficiencies of genetic inheritance. (E.g. social norms that prevent violent disputes over mate selection that arise because of genetic dispositions to fight for sexual access to females, or even cultural norms promoting the use of sunscreen, since the use of sunscreen helps to correct genetic dispositions for skin cancer.) It is more dramatic to use maladaptive outcomes to make the point that cultural inheritance leads to situations that cannot be explained using genetic inheritance alone, but such outcomes represent only one type of possible example to display to the explanatory independence that is at stake.
behaviour. If cultural evolution is biologically adaptive because it allows individuals to learn behaviour that enables them to survive and reproduce without having to test their environment on their own, then it is reasonable to expect that forces within the evolution of social learning will favour individuals that are especially adept at determining which sets of social instructions will benefit them the most. If this is the case, evolutionary pressures can emerge that favour individuals capable of acquiring the most adaptive responses possible from their (cultural) parents while at the same time exposing themselves to the least amount of risk associated with trial and error evaluations of the different patterns of behaviour they observe. These pressures are important, because they open up the possibility that individuals will attempt to economize on their evaluations of the demonstrative behaviours presented to them via cultural transmission.

Once this is the case, a situation is created where individuals begin to use indirect, but in many cases more efficient, markers to judge the most adaptive behaviour being taught to them. Examples of these indirect markers might be 'the behaviour exemplified by the majority of observed members of the group', or 'the behaviour exhibited by the most successful (or dominant) members of the group'. These guidelines will not always provide the naive individual with the most fitness-enhancing behaviour, but if the environment is dangerous enough to select against privately trying out several different behaviours before deciding on one, then following indirect markers will often lead to adaptive behaviour with a significant savings in time and personal risk.

What makes the switch to indirect indicators of reproductive success so critical is the possibility that the asymmetry between the life cycles of cultural and genetic inheritance will allow the characteristics driving cultural fitness to separate from the ordinarily parallel characteristics of genetic fitness. Thus, indirect forces within cultural evolution allow the mechanisms within cultural inheritance to achieve a level of explanatory independence that must be taken into consideration, because once individuals begin to rely on fitness indicators that are not necessarily genetically adaptive, then
sociobiological explanations can no longer capture the full scope of evolutionary forces acting on human behaviour.

An example may help to demonstrate this point. Imagine our hominid clan dwelling on the plains of the savannah. Imagine also that our hominids have a choice between a variety of ways of removing bananas from the few trees that border their savannah.\textsuperscript{12} One method is to shake the trees until the bananas fall to the ground. Another method is to climb the trees and gather the bananas by hand. Finally, a further different method is to throw stones from the ground in order to knock off bunches of bananas from their branches. In this situation it is more efficient for naive individuals to adopt the behaviour of more experienced members of the clan instead of individually evaluating each different method on their own. Furthermore, if more than one method is displayed by the experienced members of the clan, it is in the interest of naive individuals to attempt to economize on the costs of deciding between the different methods presented to them. Hence, they will adopt \textit{indirect indicators} of success, such as the strength or prestige of the members displaying each of the methods. Being guided by these indirect indicators allows naive hominids to learn the most adaptive method of acquiring bananas with the least amount of risk and effort.

However, indirect indicators are not without their own risks. Imagine that our naive hominids emulate the behaviour of the most powerful member of the clan because he or she has gained the respect and admiration of the other members in the clan. Imagine also that this powerful individual favours climbing trees to acquire bananas because of his or her strength. Naive individuals will adopt this method of retrieving bananas because it appears to be the most adaptive behaviour to emulate, but there is \textit{no guarantee that this is the case}. In fact, climbing trees to pick bananas by hand may be the \textit{least} efficient way of retrieving bananas. The individual that others emulate may

\textsuperscript{12}I should point out that I have no idea whether early hominids lived in clans or ate bananas. The example is meant to serve as a thought experiment rather than an inept attempt at armchair anthropology.
simply be strong enough to compensate for the fact that this method is inefficient, or perhaps he or she has an eccentric preference for climbing trees (or an eccentric aversion to shaking them). The result is that naive individuals are misled into adopting a maladaptive pattern of behaviour because of an indirect indicator that is, in this case, faulty. (Emulating an individual that has gained the clan's respect might, however, be adaptive and represent a substantial savings in time and energy in many other cases.)

What is important is that the drive to submit to indirect cultural indicators is in many cases strong enough to cause social groups like our clan of hominids to embrace maladaptive patterns of behaviour despite the biological selection forces working (more slowly) against such patterns evolving. Without a system of cultural inheritance, genetic selection alone would presumably favour either shaking banana trees or throwing stones at bunches, whichever is more fitness enhancing. The difference is significant because it means cultural inheritance produces outcomes that demand different evolutionary explanations than cases where outcomes are produced by a purely genetic inheritance system. In short, if the cultural evolutionary forces generally favouring indirect indicators are stronger than the genetic forces favouring biological fitness in cases where such indictors are adopted, then explanations based on our two different inheritance mechanisms will diverge. In the extreme case, the economical quality of indirect methods of cultural transmission creates a situation where a generally adaptive capacity for acquiring fitness-enhancing behaviour can be hijacked and left open to promoting gravely maladaptive outcomes.13

The implication of this result is extremely significant for the way in which we approach cultural evolution. The details of cultural transmission will no doubt be of

---

13Remember that outcomes need not be maladaptive to show the explanatory difference between genetic and cultural inheritance. Maladaptive outcomes make especially good illustrative examples, but any deviation from sociobiological predictions is enough to demonstrate that cultural transmission requires its own independent explanations. In the case of our hominids, even if climbing trees was indeed the most adaptive method of retrieving bananas, indirect cultural indicators can cause naive clan members to be disproportionately likely to adopt this method in a way that cannot be adequately explained using sociobiological predictions. See (Boyd and Richerson 1985, p. 206).
empirical interest whether this system meets the challenge of achieving an explanatory independence from sociobiology or not. But if we are specifically interested in accounting for the evolutionary forces contributing to human social behaviour and ultimately our moral capacities, then the question of explanatory independence takes on a profound importance. For if the structure of cultural heredity is such that we cannot safely ignore cultural deviations from genetic fitness (and we therefore cannot rely on the predictions of sociobiology), then a whole new domain of evolutionary considerations suddenly becomes open for investigation. Once the indicators of fitness driving cultural evolution are no longer connected to biologically adaptive traits, we must begin to examine the cultural selection pressures acting on the traits acquired by individuals participating in cycles of social transmission. These selection pressures may be biologically adaptive in most instances, but Boyd and Richerson's results demonstrate that this will not always be the case, and that we need to study the "internal logic" of a cultural system that operates according to its own unique evolutionary principles. (pp. 277-9)

The Right Tools for the Job

Thus far we have seen that cultural evolution is both consistent with our genetic system of inheritance and yet independent of this system in the sense that it provides separate explanations for social behaviour than those we would otherwise derive from a purely biological perspective. Cultural transmission can thus be plausibly understood as an important part of an overall coevolutionary process, and one that merits further investigation as a separate evolutionary system of inheritance. The third challenge, however, for the project of modeling cultural evolution is the question of whether we ought to model the mechanisms within this separate system of cultural transmission using the very particular tools of evolutionary biology. We may agree that cultural inheritance forms part of an intricate coevolutionary process, and we may agree that because cultural
inheritance evolved from biological origins it is more likely to function according to evolutionary principles. But if we consider how to most accurately describe the way cultural inheritance functions it remains an open question whether the similarities between the genetic and cultural elements of coevolution are strong enough to sanction modeling the latter in the image of the former.

Hence, it may be doubted whether the tools of evolutionary biology are genuinely appropriate for mapping out what may be a very different system of inheritance.\(^\text{14}\) The danger here is that if the structure of cultural evolution is different than that of genetic transmission, then we will end up retarding our understanding of the way cultural inheritance functions by constructing models biased towards an unjustifiably strict parallel between cultural and biological phenomena. The rules of genetic transmission certainly seem to capture the essence of many intriguing cultural processes, but if the analogy between culture and biology is drawn too tightly it can definitely become misleading. Thus, the adequacy of using genetic modeling techniques for cultural transmission depends on the benefits gained from similarities between the two systems outweighing the potential distortions associated with applying a potentially inappropriate set of evolutionary tools to a structurally distinct system of inheritance.\(^\text{15}\) This hopeful compromise looms as a source of uncertainty for current theories of cultural evolution and ought to be addressed with a sense of caution and humility until the true nature of cultural transmission is revealed. Moreover, as we shall see in section three, the technical problems facing models of cultural evolution become particularly acute in the case of memetics.

---

\(^\text{14}\)This is the argument most commonly invoked by critics who dismiss the project of modeling cultural evolution (often in ten words or less). See (Gould 1996, pp. 219-220) and (Midgley 1994).

\(^\text{15}\)This tradeoff is not lost on the proponents of using evolutionary theory to model cultural evolution. As Boyd and Richerson note, "... we will use the synthetic theory as a source of analogies and formal mathematical machinery with which to build a theory of the evolution of culture. There are important differences between the genetic and cultural inheritance systems, and the theory will by no means neglect them. However, the parallels are profound enough that there is no need to invent a completely new conceptual and mathematical apparatus to deal with culture." (1985, p. 4)
III. The Move Towards Memetics

As a newly emerging science, memetics endeavors to drive the parallel between genetic and cultural inheritance to its farthest extreme by modeling cultural transmission according to a strict analogy with the structure of our genetic system of inheritance. Just as it sounds, the term 'meme' is meant to serve as a direct analogue of the discernible units of our DNA. Thus, memetics hopes to construct a workable theory of cultural evolution based on the idea that the basic units of social transmission can be thought of as functionally equivalent to the units responsible for genetic transmission. According to this view, both genes and memes are instances of a more general classification (they are both replicators) and both can therefore be expected to obey the same general evolutionary rules in terms of copying, propagation and the pressures of natural selection. Obviously, theorists working in memetics do not believe that every mechanism involved in genetic transmission has a direct analogue in the cultural realm (it is doubtful, for example, that memes are transmitted in conceptual double helixes). However, memetics aspires to draw as much insight as possible from the apparent similarity between genetic and cultural phenomena.

Memetic Independence

If the conclusion drawn from contemporary attempts at modeling cultural transmission is that cultural inheritance constitutes an independent evolutionary system, then the study of memetics takes this conclusion and pushes it to the edge of its explanatory possibilities. Memetics assumes a bold standpoint with respect to both the evolutionary structure of cultural transmission and the fact that cultural evolution produces outcomes that require independent explanations from the predictions of sociobiology. The former issue is discussed at length in the literature written on memetics; the latter issue is often tacitly assumed to be a consequence of the hypothesis
that memes exist in an entirely parallel evolutionary environment. By working with a precise analogy between cultural and genetic forces, memetics virtually defends the strong interpretation of culture as an independent system of inheritance by definition, since it views cultural patterns to be primarily subject to the forces relevant to the cultural side of the parallel between culture and biology. Hence, memes and genes are thought to exist in systems of inheritance with similar structures, but each information system is expected to operate for the most part according to its own corresponding sets of evolutionary pressures. In this way, memetics represents the most ambitious articulation of the view that cultural transmission operates according to its own unique evolutionary logic. It is this ambition that makes memetics especially intriguing, but it also leaves the area open to more objections and controversy than it may be able to resolve.

**Dawkins' Legacy**

The term 'meme' was originally introduced by Richard Dawkins in the final chapter of his influential book *The Selfish Gene.* (1976) Although the idea that replication, or imitation, represents a kind of fundamental universal force can be traced back as far as the 19th century,16 Dawkins supplied the hypothesis with a renewed sense of vitality by connecting it to modern evolutionary theory and attaching an especially memorable name to the basic unit responsible for cultural transmission. (Thus, one could say that Dawkins introduced a mutation in the 'imitation' meme which gave it a much greater cultural fitness.) Part of the reason Dawkins' work has had such a powerful impact on the study of cultural evolution is the fact that he is a prolific and accessible author whose books have been read by vast numbers of readers who would otherwise never come into contact with evolutionary biology. Dawkins clearly has a skill for explaining the complexities of evolutionary theory in a straightforward, non-technical

---

16See (Tarde 1895).
manner, and his ability to clearly communicate difficult concepts has surely helped the concept of memes (the 'meme' meme, if you will) to spread more effectively through our cultural 'infosphere'.

Yet, another reason Dawkins' articulation of memetics has seen enduring success is most likely the challenge associated with mapping out a completely new evolutionary system. Dawkins is quite clear in *The Selfish Gene* that he is not simply drawing a loose comparison between genetics and culture, nor is he presenting a mere metaphor for cultural analysis using the tools of evolutionary terminology. Indeed, memetics is described as a genuinely separate evolutionary system that obeys the laws of natural selection in precisely the same way as do biological phenomena. This strategy constitutes a controversial approach to modeling cultural transmission, but it holds a seductive appeal for those drawn in by the prospect of developing an entirely new scientific discipline. Dawkins' original unblinking stance with respect to the ontological status of memetics has certainly led to difficulties for the science to sort out, but the alluring nature of this position has also dramatically helped memetics to emerge as a thriving new area of investigation.

Ironically, Dawkins' own interest in memes is entirely perfunctory. His motive for introducing memetics as a legitimate evolutionary system has from the beginning been to use memetics as a rhetorical case study of what he calls 'Universal Darwinism' - the view that natural selection is a substrate neutral law of the universe of which biology is but one relevant example. Hence, Dawkins' animated argument for the existence of memes is in fact only an instrumental step towards a broader thesis that is his primary interest. Nevertheless, his writing and potent terminology greatly amplified the

---

17 The word 'infosphere' is Daniel Dennett's term for the medium in which cultural evolution occurs. A more intuitive, although perhaps less accurate, description might be "conceptual environment" or else "cultural ecosystem". I tend to prefer the compromise 'infosystem'.

18 See (Dawkins 1983) and (Dawkins 1986) for more on Universal Darwinism.

19 Moreover, Dawkins seems to have retreated somewhat on the status of memetics in his later writings, such as (Dawkins 1982) and (Dawkins 1986). Although Dawkins claims to have been falsely accused of
existing interest in cultural inheritance. Whether intentionally or not, Dawkins played a key role in initiating the growing trend toward evolutionary approaches to cultural phenomena.

Influence and Potential

Today, the study of memetics includes a broad range of research projects and a diverse scale of precision and complexity. References to memes turn up loosely in works of fiction, pop-culture and magazines, yet one also finds them carefully discussed in noteworthy books and respected scholarly journals. The number of results for internet searches using the keyword 'meme' increases exponentially every year, but it is difficult to determine whether this fact ought to be interpreted as a sign of growing acceptance or the rush towards a fleeting, technology-induced fad. Plus, the quality of the material available on memes is equally as questionable as its quantity: some sources display all the subtlety and craftsmanship of a common self-help book (Brodie 1996), while others attempt detailed treatments of the subject with intentions to build in formidable degrees of complexity in the future (Lynch 1996). Thus, memetics currently seems to embody a unique mixture of pop intuitionism and extremely thorough academic research. Which quality will eventually prevail as memetics' deserved reputation remains an open question.

The easiest way to get an intuitive feel for memetics is to think of a catchy tune and its mysterious sense of longevity. Whether it be Thus Spake Zarathustra from the backtracking on the meme hypothesis and now maintains that his original intentions were more modest than some wished to believe (1999), it is not clear how his original argument for cultural replicators constituting a veritable example of Universal Darwinism can be reconciled with such statements as "Cultural 'evolution' is not really evolution at all if we are being fussy and purist about our use of words, but there may be enough in common between them to justify some comparison of principles." (Dawkins 1986, p. 216)

20Three cases in point respectively: Neal Stephenson's Snow Crash (1993), Malcolm Gladwell's The Tipping Point (2000) and Wired magazine's "overhyped meme of the month" column. (Gladwell actually explicitly rejects the terminology of memetics, but his approach to describing cultural phenomena relies on a similar premise.)
film 2001: A Space Odyssey, the whistling tune from Bridge on the River Kwai or else some anonymous pop song from the radio that repeats endlessly in one's head despite all efforts to remove it, a catchy tune represents a meme with a striking degree of cultural fitness. In the case of music, a meme replicates itself in recording equipment and the brains of (often unsuspecting) hosts and then distributes copies of itself to whoever happens to be within earshot. The qualities that help music to permeate our minds are fairly easy to predict: tunes that are short, repetitive, rhythmical and distinct will be more memorable than those that are long, complex, irregular or mundane. Yet the differential selection of various musical patterns - the proficiency of certain tunes to flood our cultural environment when so many others fail - is nothing short of spectacular. The survival of certain musical patterns over others is thus a graphic example of natural selection at work in the cultural realm.

For a number of reasons, catchy tunes serve as a particularly illustrative example of how memetic models function. First, the way music gets picked up and reproduced in our heads provides a helpful example of meme replication in a familiar cultural setting. Second, the fact that music can be instantiated in a number of different media (e.g. sound waves, tapes, compact discs, digital mp3 files, etc.) displays the wide range of transmission mechanisms available to memes. Third, music demonstrates that memes can be transmitted horizontally rather than vertically with respect to biological heredity. To be transmitted horizontally in this sense is to be distributed across generations (e.g. from brother to sister) rather than being passed directly from parents to genetic offspring.21 Fourth, musical tunes serve as illustrative examples for memetics because the reasons for their cultural success are predictable, but they are not so predictable that we are not curious about the kinds of forces at work shaping our cultural landscape.

---

21This terminology originates from (Cavalli-Sforza and Feldman 1981). Music is actually a particularly good example of horizontal transmission, because it often has considerable difficulty being passed on vertically from parents to offspring...
Finally, the fact that catchy tunes get stuck in our brains against our will nicely demonstrates the fact that cultural evolution does not always proceed according to our conscious intentions. Instead, cultural evolution proceeds on the basis of a complex system of selection mechanisms, of which our explicit objectives are but one.22

Now that we have an initial sense of what a memetic perspective is all about, the next important step is to note that memetics is intended as a means of modeling all cultural patterns, not just trivial items like catchy tunes, urban legends and fashion trends. Thus, instead of describing the cultural selection of competing music patterns, memetics can also help us to examine the changing circumstances of religious beliefs, economic trends, social policies and academic scholarship. The applications of memetics used to initially communicate the persuasive nature of the theory are often trivial because we understand enough about the structure of these minor applications to recognize the distinctly evolutionary reasons for the cultural units in the examples to possess their baffling ability to replicate and distribute copies of themselves. (E.g. annoying pop songs, chain letters, email viruses, pyramid schemes, etc.) Nevertheless, the promise of memetics is that it can work with more sophisticated applications and provide us with useful information about the cultural selection mechanisms driving more important social phenomena.

Memetics is therefore suited to investigating a broad range of issues in various disciplines. In sociology one could use memetics to investigate the specific mechanisms responsible for socializing children in society. Children must absorb a massive amount of information to function successfully in social systems, and memetics could help us to decipher the cultural selection forces that determine the types of information and guiding social norms that tend to get transmitted and successfully adopted by impressionable youth. Evolutionary psychology, on the other hand, could use memetics to map out the

---

22Potential case in point: is the theme from Bridge on the River Kwai still playing in the background of your mind?
selection forces operating within the brain that structure conscious experience. In the case of anthropology, one could apply memetics to the task of cataloguing the different selection pressures that make individual cultures susceptible to their own particular kinds of belief systems.

In each of these disciplines, memetics can provide a systematic method of investigating how cultural information tends to replicate and propagate itself according to predictable evolutionary rules. My own interest in this thesis is the prescriptive justification of ethical claims, as we shall see in chapter three. By outlining the cultural selection forces acting on our moral intuitions, I think that memetics can tell us something interesting about the evolutionary influences on our ethical judgments in the same way that it can help to explain why we are influenced by certain kinds of memorable (but not necessarily pleasant) musical tunes. Thus, despite its current limitations, I believe memetics carries a very noteworthy potential.

A similar 'cautious optimism' is exhibited by most theorists currently working to develop a credible science of memetics. The hope is that memetics will bring about a striking perspective shift for the way we view cultural transmission, yet it is certainly recognized that memetics presently holds a comparatively limited amount of explanatory value. As a result, the existing academic literature on memetics usually involves either speculative explorations of how memetics might enable us to better understand particular cultural or psychological phenomena (e.g. mental illness, imagination, policy making) and philosophical problems (e.g. the origin of language, the nature of cognition). Or else this research is devoted to the slow work of advancing memetics as a science by attempting to solve the very detailed technical problems currently facing the subject.

**Technical Obstacles**

As we have seen, the challenges facing theories of cultural evolution can be grouped into three basic areas of concern: the question of whether cultural transmission
is biologically adaptive (not especially controversial), the question of whether cultural inheritance can obtain a meaningful sense of independence from genetic inheritance, and the question of whether the benefits associated with using the tools of evolutionary biology to describe cultural transmission outweigh the drawbacks created by following a potentially misleading metaphor too closely. It is this last area of concern that poses the greatest challenge to the study of memetics. Even if memetics is in fact an example of a genuine evolutionary system, it is perilously easy to attribute inappropriate details from the biological facts with which we are most familiar to the structure of a newly emerging cultural science. Memetics must therefore be meticulously scrutinized, despite its rising popularity, to guard against allowing a seductive analogy to override our sense of mindful inquiry.

To determine how precisely the parallel between cultural and biological transmission can be drawn, numerous technical dilemmas must be unraveled in order to establish the fidelity between culture and what we know about genetic inheritance. Many of these dilemmas involve cultural processes that are difficult to measure with any degree of certainty. Hence, the project of cataloguing the similarities and dissimilarities between cultural and genetic transmission often seems interminable. However, the main questions that have been discussed in the literature so far include the following:

1. **What exactly is a meme?**

The first and by far the most widely debated problem for memetics is the question of what we ought to identify as the basic unit of cultural replication. Memes are often casually defined as units of *imitation* (Dawkins 1976, p.192) or "*distinct memorable units*" (Dennett 1995, p.344). But units of what? Granted, the definition of a gene is in many ways equally indistinct, but we at least know that genes represent the synthesis of

---

23 For a detailed review the problems facing memetics, see (Rose 1998).
particular polypeptides and that they are composed of specific, statistically defined nucleotide sequences found within the structure of DNA. No parallel substrate has been agreed upon to ground a definition of memes. Many view memes as abstract elements of cultural information that can be instantiated in any one of a number of different substrates, whereas others see memes as physical units of mental representation that can only be located in the brain. If memetics is to emerge as a consistent science, agreement must be reached as to whether memes designate the physiological replicators involved in neural transmission or the phenotypic effects (often labeled interactors) produced by replicators within a complete cultural system. Both options have advantages and potential pitfalls, but a consensus of some kind must be reached on the issue before memetics can advance in any meaningful direction.24

2. Are they memes or meta-memes?

A problem related to the fact that no precise definition of memes exists is the question of what counts as a basic meme versus what counts as a more complex combination of several different memes - something we ought to call a 'meta-meme' (or perhaps a "pheme")25. The examples offered as instances of memes often vary immensely in terms of the representational content to which they refer. Sometimes simple ideas, such as "wheel" are given as examples of memes, but often more complex units of information such as "the idea of God" (Dawkins 1976, p. 192) or

24 This issue, more than any other, is currently being fought out in the pages of the leading (electronic) journal in the area: Journal of Memetics - Evolutionary Models of Information Transmission; http://www.cpm.mmu.ac.uk/jom-emit/. For a thorough articulation of the view that we should resist the urge to think that memes necessarily refer to natural objects and that they can be functionally defined as replicators that are "not instantiated in any exclusive kind of physical array or system", see (Wilkins 1998). For the opposing view that memes can be defined as brain states, one can refer to (Dawkins 1982) where Dawkins clarifies his earlier description of memes by stating that, "A meme should be regarded as a unit of information residing in a brain. It has a definite structure, realized in whatever physical medium the brain uses for storing information. [...] This is to distinguish it from its phenotypic effects, which are its consequences in the outside world." (p. 109)

25 Wilkins (1998) uses the word "pheme" in order to link the unit for expressed packages of memetic properties to the phenotype of memetic individuals.
"deconstructionism" (Dennett 1995, p. 344) are listed as examples of memes. It seems dubious, however, whether a consistent science of cultural evolution be developed if the scope of the meme is left this indistinct. Certain theorists seem to take pride in the fact that a meme can refer to any cultural pattern capable of replicating itself with sufficient fidelity, no matter how broad or complex this pattern may be. But this flexibility in representational scope becomes a burden when we attempt to construct working models of how memes are transmitted and selected for within their cultural environment. Thus, deciding on a more consistent fundamental unit of transmission (independent of the issue of substrate) will be immensely helpful if memetics is to become a more accurate cultural science.

3. Is memetics Lamarckian?

A constant source of criticism for memetics is the fact that it seems to break from accepted evolutionary theory by allowing the inheritance of acquired characteristics. New memes can be acquired at any point in our life cycle and still be passed on to further generations since they are not bound to the fixed life cycle of biological reproduction. This seems to pose a problem for a strict analogy with biological inheritance: when we are exposed to new memes they immediately alter our 'memetic makeup', whereas it is a serious mistake to think that environmental changes to our phenotype can have any direct effect on our genes. (As the often quoted criticism of Lamarckian inheritance points out, evolution does not create giraffes with long necks because each one reaches as high as it can to get food and then transmits this behavioural information to further generations.) Hence, memetics is often accused of operating according to Lamarckian selection forces that have been forcefully discredited in modern evolutionary theory.26 This criticism

26 Stephen Jay Gould (1997, p.52) makes this point about modeling cultural evolution in the second part of his review of (Dennett 1995).
may not be insurmountable, but it points to potential discrepancy between biological and cultural inheritance that needs to be carefully investigated.

4. Is intentionality a problem?

The next main problem for memetics is the fact that we can consciously direct the process of meme propagation. Our status as intentional creatures seems to strain the parallel between cultural and biological evolution since the latter is based on the natural selection of random variations, while the former is capable of accommodating purposefully designed mutations. Evolutionary theory does not explicitly depend on a random source of variation in order to occur, but the introduction of consciously directed variations may constitute a radical enough break from what we know about genetic examples of evolutionary theory that the comparison between biological and cultural processes may no longer be worthwhile. Moreover, the thesis that we can actively control the structure of memetics seems to rely on a false dichotomy between our thoughts and ourselves as independent creatures. If the meme hypothesis is taken to its logical conclusion there is little left of the human mind other than collections of memes\(^\text{27}\), so it becomes problematic how we can remove ourselves from the contents of our own minds long enough to contemplate directing the flow of meme propagation from an independent standpoint. The tension between our consciously directing memes and these very memes being responsible for our consciousness can most likely be resolved by some pragmatic balance (chapter four addresses this balance), but the issue nonetheless raises methodological concerns for memetics' ability to model cultural patterns accurately as an external instance of information transmission.

\(^{27}\text{This is the view defended by Dennett in Consciousness Explained (Dennett 1991) and Susan Blackmore in The Meme Machine (Blackmore 1999).}\)
5. Is memetics prohibitively misleading?

By working to resolve the problems standing in the way of a consistent interpretation of how memes operate, we may be able to fit memetics into a general evolutionary framework that meets the requirements necessary to be classified as a Darwinian system of inheritance. However, it might still be the case that the structure of cultural transmission deviates from the structure of genetic transmission substantially enough to strain the parallel between the two beyond the point where this parallel remains useful. If we observe the analogy between genes and memes too closely, we may therefore inadvertently construct a distorted picture of an otherwise legitimate evolutionary system of transmission. Hence, the metaphor between cultural and genetic information may be illuminating in certain cases yet still be severely misleading if drawn on too heavily. Furthermore, the language used in studies of memetics is often charged with potentially inflammatory rhetoric that helps fuel the worry that the genetic/cultural metaphor is misleading. The terms "thought contagion" and "viruses of the mind" may capture certain unique insights about the propagation of cultural patterns, but they also take away from memetics' credibility when they are carelessly employed.

Potential Responses

In light of the preceding points of criticism, the future of memetics may appear somewhat uncertain. If the problems facing the area cannot be resolved as memetics develops into a mature field of study, then there is still a very real possibility that the difficulties it faces will be insurmountable and memetics will eventually vanish as a passing terminological fad. However, it ought to be noted that the evidence against memetics is not conclusive, and certain valuable attempts have been made to show how memetics may be able to overcome the obstacles it currently finds in its path. David Hull

---

28See (Sperber 1996, chap. 5) for a Darwinian critique of the idea that cultural evolution can be described using selection models drawn from genetic evolution.
(1982), for example, has provided an influential defense of the prospect of modeling 'socio-cultural evolution' which carefully addresses many of the concerns raised against memetics.

Hull points out that much of the criticism surrounding the technical issues related to differences between genes and memes rests on a failure to be clear about whether one is working on the biological or cultural side of the analogy upon which memetics relies. Memetic inheritance may appear both Lamarckian in structure and unsuitably directed by our conscious intentions, but it will only appear this way if we fail to be consistent about defining the appropriate units and environmental conditions that are relevant to the cultural side of the parallel between genetic and cultural inheritance. Cultural transmission is bound to appear Lamarckian from a biological standpoint. The question is whether cultural replicators acquire characteristics from environmental changes to their cultural phenotype - a situation that would truly be a problematic example of Lamarckian inheritance. Similarly, our ability to consciously direct memes appears to pose a problem from the phenomenological perspective to which we are accustomed, but if we view the mind as a collection of memes competing for neural resources in our brains and cultural institutions, then the problem dissolves entirely.

In fact, the argument underlying Hull's thesis can be generalized to many of the controversies currently facing memetics. What seems to be an insoluble methodological problem may actually be a failure to fully appreciate the new terms and conditions that apply to a completely new system of inheritance. If we knew exactly what the fundamental units of memetics were, we could be more clear about these terms and conditions, and many of the problems in contemporary memetics would resolve themselves (or sink the area once and for all). In the meantime, since we do not yet possess a firm grip on the intricacies of meme transmission, it is simply too early to tell whether or not memetics will eventually prevail as a workable science. The cautious optimism displayed by meme theorists therefore seems appropriate, for the area currently
faces an alarming number of difficulties yet none so obviously insurmountable as to represent a prohibitive methodological barrier.

Moreover, the ambitious nature of memetics tends to amplify the existing problems associated with modeling cultural evolution, but it also captures our intuitions about cultural inheritance in a unique and especially clear manner. The reason so many theorists are willing to sort out the difficulties facing memetics is due to the fact that the area carries the potential to provide us with a tremendous amount of explanatory power. The fact that cultural patterns obey some sort of evolutionary transmission process seems inevitable, but there is something profoundly intuitive about the hypothesis that ideas act as functional analogues of our genes. We often speak of traditions "dying out" in the same way we speak of the loss of an endangered species. We trace the spread of fashions, cults, urban legends, religions and political ideologies in terms of their cultural robustness in much the same way we trace genetic characteristics in terms of biological fitness. And we understand the selection pressures working within cultural infosystems (e.g. "that doesn't make any sense!", "that's too long to remember!", or "how can you put your faith into that?") in a way that is strikingly similar to our understanding of biological selection pressures. Thus, the ambition of memetics undoubtedly carries disadvantages, but it also possesses an undeniable appeal in terms of being able to clearly articulate our intuitions about the evolutionary nature of cultural transmission.

Memetics' general appeal notwithstanding, the area possesses two particularly useful characteristics which may help to tip the scale towards the science's advantages outweighing its potential drawbacks. The first is that memetics carries the potential for a startling level of precision in terms of being able to map out the evolutionary forces associated with cultural transmission. If we can identify particular memes and the forces acting upon them, we would be able to generate an amazing sense of accuracy in terms of understanding the details of cultural inheritance. This accuracy could provide us with a far better conception of the forces responsible for the evolution of cultural patterns. In
particular, an increase in accuracy would allow a more detailed understanding of the kinds of evolutionary pressures that influence the content of our ideas and cultural institutions.

The second useful characteristic memetics possesses is an especially clear theoretical framework for working on issues of cultural transmission. The term 'meme' provides a very useful provisional focal point for examining cultural phenomena, and it helps to conceptualize the hypothesis that cultural information is capable of replicating, propagating and being selected for within a conceptual environment. Defining this unit is problematic, but the term nonetheless serves as a helpful place-holder for starting discussion about issues of transmission fidelity, cultural fitness, and the unique selection pressures operating within cultural inheritance systems. This is perhaps the most inspiring capacity of memetics, since it holds the promise of providing the tools to investigate a revolutionary new domain of evolutionary forces. These tools may be rudimentary at present, but they provide us with an initial means of contemplating the idea that cultural patterns reproduce themselves according to the internal logic of an independent system of inheritance. Thus, despite its deficiencies, the language of memetics provides us with considerable benefits, since the area is capable of dramatically altering the way we think about the evolution of cultural information.

**Conclusion**

This chapter has endeavored to provide an introductory overview of the issues surrounding cultural evolution and the emergence of a science of memetics. It is by no means a complete summary of the area. To provide such a summary (or alternatively to provide a detailed account of any one particular issue within the area) is beyond the aims and scope of the project at hand.\(^{29}\) Instead, the chapter has aspired to capture the overall

\(^{29}\text{To obtain a comprehensive summary of the work done in modeling cultural evolution or to investigate the details of a specific issue in memetics, the reader is encouraged to consult (Durham 1991) and Journal}
character of attempts to model cultural phenomena from an evolutionary perspective. This more modest objective has been achieved by outlining both the motivation behind efforts to develop theories of cultural transmission and the problems these theories currently face in order to ultimately succeed. The underlying theme of the chapter has been that theories of cultural evolution - those of memetics in particular - have numerous obstacles in their path which must be overcome if the area is to become a practicable scientific discipline. However, I also hope to have conveyed the message that none of the obstacles facing this area are insurmountable.

Hence, the most important conclusion reached in this chapter is that theories of cultural evolution carry a tremendous amount of potential despite their current difficulties. I hope to have shown why it is that so many theorists have taken the time to try to resolve the problems surrounding cultural evolution, and why it seems that something important is at stake when issues related to cultural transmission are discussed. If we achieve the ability to accurately map out the evolutionary forces acting on our cultural ideas, the consequences could be extraordinary. With a new system of inheritance to investigate, we could discover selection pressures and evolutionary fitness considerations that are presently recognized only as diffuse cultural tendencies. Theories of cultural evolution therefore hold the potential to articulate the forces driving cultural change with a precision that has up to now been unimaginable. This potential may never be realized if the area eventually succumbs to its methodological dilemmas, but if it is successful the results may be well worth the wait.

of Memetics - Evolutionary Models of Information Transmission; http://www.cpm.mmu.ac.uk/jom-emit/ respectively.
Chapter 3

Ethics and Memetics: A Best-case Scenario

Introduction

The preceding chapters have established two important conclusions about the combination of evolutionary theory and philosophy. Chapter one revealed that current articulations of Evolutionary Ethics fail to account for important cultural details in their efforts to supply ethics with a naturalistic source of prescriptive value. It concluded that the sociobiological facts to which these theories appeal are better utilized as a source of genealogical information about the biological origins of our moral capacities. Chapter two demonstrated that contemporary approaches to cultural evolution have the potential to significantly alter the way we look at cultural information, even if these theories currently find themselves addressing the complex methodological problems associated with constructing reliable models of cultural transmission.

The discussion in chapter two reveals that the potential contemporary theories of cultural evolution possess is dependent on two key features. First, for these theories to achieve a significant degree of explanatory power (in terms of our interest in the evolutionary forces influencing human social behaviour), cultural transmission must operate according to an inheritance system that is meaningfully independent of our genetic system of inheritance. Unless cultural evolution operates according to its own unique 'evolutionary logic', it will not provide explanations for social action that differ from existing sociobiological predictions. Second, the explanatory potential associated with new theories of cultural evolution is closely linked to the promise that these theories will be able to model cultural forces in increasing detail. We presently have an adequate common-sense understanding of the kinds of ideas that persist in our cultural infosystems and the reasons why these ideas prove to be adept at replicating themselves. But the appeal of new theories of cultural transmission is that they will be able to provide us with
a greater level of precision in terms of identifying the selection mechanisms and fitness defining characteristics that drive cultural evolution. The combination of increased precision and explanatory independence suggests that models of cultural transmission will be able to supply us with a striking new means of discovering the evolutionary forces acting on our minds and our cultural landscape.

Considering the outcomes of chapters one and two, the natural question at this point is to ask whether an advanced understanding of cultural transmission can supplement the inadequacies of existing theories of Evolutionary Ethics and in some way have a meaningful impact on contemporary ethical theory. It would be overly optimistic to think that an understanding of cultural transmission will provide a straightforward remedy for the ills of EE. But the fact that cultural evolution takes us one step closer to the level of detail required for an adequate treatment of our moral intuitions (as opposed to our underlying, potentially immoral, genetic social dispositions) suggests that reliable models of cultural transmission will be able to contribute something important to the project of developing a meaningfully evolutionary approach to ethics.

Chapter three therefore examines the consequences for ethics if cultural evolution establishes itself as a profitable area of study. By ‘profitable’ I mean capable of providing us with substantial new insights into the ways in which culture replicates and distributes itself over time. How substantial? How new? My intention is not to set the bar too high and foster the expectation that only some wildly futuristic theory of cultural evolution could have an effect on ethics. (Quite the opposite, I suspect that current research into cultural transmission could have a significant impact on ethics.) Nevertheless, I address the hypothetical situation where the research on cultural evolution acquires an indisputably credible status in order to narrow the focus of the chapter to a specific discussion of how ethics will be changed by an understanding of cultural
evolution, rather than continuously re-evaluating the scientific standing of cultural transmission models at each step along the way.

Using memetics as an example of the more viable theories of cultural evolution, the chapter therefore addresses the following question: *If memetics ultimately succeeds at producing credible evolutionary models of cultural phenomena, how would this newly emerging science contribute to our understanding of ethics and its justification?* At this point, memetics is an uncertain, albeit compelling, area of study. However, we presently have enough of a sense of what it would mean to successfully uncover the evolutionary status of cultural phenomena that we can reasonably contemplate the consequences memetics will have for ethics at this early stage in our understanding of cultural evolution.\(^1\)

To do this, the chapter adopts a 'best-case scenario' approach. Rather than ignoring memetics until it emerges as a fully established science, the aim is to investigate the implications for ethics if memetics successfully lives up to our considerable expectations. We may not know exactly how the future of this science will unfold, but we know enough to draw interesting conclusions about how memetics will change the way we look at cultural phenomena. One might object that memetics faces severe enough methodological problems that it will never develop into a mature field of study. But as we have seen in the previous chapter, we have good reason to believe that the problems facing memetics are surmountable despite their daunting appearance. Moreover, most disputes in the area revolve around internal disagreements over the *most effective* means of modeling what is already assumed to be an evolutionary system. While these internal disagreements about the technical modeling issues at stake may be interesting, the lessons memetics can teach us are not always critically dependent on the

\(^1\)Moreover, uncertainty and an early stage of development have certainly not dissuaded us from examining the implications of neuroscience for the philosophy of mind. Memetics, I submit, ought to therefore be given the same opportunity.
intricacies which must remain ill-defined until we can pinpoint the best way to model cultural replication and transmission.

Thus, I wish to remain neutral about the prospects of any one conception of memetics succeeding as the most viable theory of cultural evolution. Instead, I am convinced there is enough overlap among contemporary models to ensure that we can speak meaningfully about what it would be like to have access to a well developed theory of cultural evolution. Without knowing exactly what kind of evolutionary system is at work or exactly how this system functions, we can appeal to certain modest assumptions about what to expect from theories of cultural evolution to generate some very interesting results.

The chapter is structured in four sections. Section one briefly sketches what a best-case scenario for memetics will look like. Section two addresses three different initial reactions that arise when the question of how memetics might affect ethics is first considered. These reactions predict drastic changes to the nature of ethics, yet section two also describes why these interpretations of the impact memetics will have on ethics are misguided. Section three addresses a more practicable application of memetics to ethics: the use of cultural evolution to more clearly specify the standards of moral realism. This option is presented as being more plausible than the initial reactions described in section two, yet it is ultimately shown to be either unacceptable or uninformative when it comes to issues of prescriptive justification. Finally, section four proposes an application of memetics to ethics that is directly parallel to the conclusion drawn in chapter one. The proposal is that memetics will have a significant impact on ethics by enabling us to trace the evolutionary genealogy of our moral intuitions and thus uncover the unique cultural forces that cause certain moral memes to propagate more effectively than others. By revealing these otherwise unacknowledged cultural selection pressures influencing our moral intuitions, it is argued, memetics provides us with a
means of reevaluating the prescriptive authority of memes that manage to propagate for reasons that are not related to the characteristics we ordinarily attribute to their ethical standing. The conclusion drawn is that this ability to reveal the genealogy of our moral memes suggests that memetics will have a subtle yet exceedingly useful impact on the nature of ethical justification.

I. Anticipating the Best-Case Scenario

What exactly would it mean to have access to memetics and to therefore fully recognize the evolutionary nature of cultural transmission? As we have seen, memetics attempts to model the way in which culture is transmitted and selected for in terms of the evolutionary principles abstracted from biological instances of Darwinian natural selection. Memetics draws on the generalizable aspects of evolutionary theory in order to model the processes responsible for passing certain patterns of identifiable ideas on to our (cultural) offspring more effectively than others. Having access to a fully developed science of memetics would therefore allow us to identify the unique evolutionary pressures present within our cultural infosystems and predict which memes will be most likely to thrive within particular conceptual environments.

What would it be like to possess this knowledge? It seems difficult to imagine having access to so much information about ourselves and our cultural environment, and the prospect of acquiring a sophisticated awareness of the evolutionary forces acting on each thought permeating *our own minds* is not an easy concept to grasp. We can begin by imagining what it would be like to have an understanding of the factors that make us more likely to acquire and possibly disseminate certain cultural units over others. Presumably, memetics would provide detailed information about the psychological mechanisms responsible for accepting, modifying and potentially distributing different types of memes. Plus, a fully developed memetic science would reveal the workings of the institutional mechanisms that allow certain memes to spread through society more
effectively than their competitors. At the very least (i.e. even if we set aside aspirations of uncovering the intimate physical structure of our brains and cultural institutions), memetics would be capable of providing a statistical analysis of the characteristics that make memes more or less likely to survive the cultural selection processes that exist within their respective cultural environments.

For example, memetics might provide quantitative evidence for the fact that ideas containing memorable or appealing words like 'freedom' or 'natural' (and more recently 'paradigm') seem especially adept at replicating themselves in our current North American cultural environment. Furthermore, memetics could explain why memes are more likely to survive if they attach themselves to such familiarly seductive meme categories as sex, wealth, power or faith. It is a common truism that 'sex sells' or that the thirst for power is insatiable. But memetics will presumably be able to provide a sophisticated analysis of the conditions that allow these memes to distribute copies of themselves so effectively. Rather than relying on our indistinct impressions of the appeal these memes possess, memetics could supply us with quantitative methods of determining why these memes are so strongly selected for and the ability to measure exactly how powerful an influence they exert on our cultural surroundings. In effect, memetics could provide solid empirical evidence for the cultural forces that we only diffusely perceive to be shaping the composition of our ideas.

More importantly, however, memetics might reveal other cultural biases of which we are not yet aware. We may find, for example, that memes appealing to hopeful spiritual entities are extremely difficult to extinguish within even the most skeptical of cultural niches. Similarly, we may find that concise memes are disseminated far more readily than we otherwise expect. Stranger still, we may find that unique rules of cadence and syllable emphasis cause the linguistic presentation of certain memes to strongly affect their ability to replicate in our minds. This is not an unreasonable possibility considering the significance of language in human cognition, yet the implication that the
cultural success of our ideas is determined by such semantically insignificant characteristics as syllable emphasis and cadence is nonetheless startling. Moreover, there are undoubtedly unique biases built into the fabric of each individual type of cultural media, such as newsprint, radio, television or the internet. Each of these substrates has its own tendencies to promote memes with particular characteristics, and each one undoubtedly possesses its own unexpected biases towards the promotion of particular memes within our cultural networks.

**Cultural Fitness**

Developing an accurate model of the ways in which the biases of these cultural media interact seems a mind-boggling task. However, by observing the differential selection of memes within each substrate we can begin the process of assigning cultural units varying degrees of cultural fitness depending on their ability to survive the selection pressures present in our cultural environments. This would provide a tangible starting point for evaluating how much more likely it is that some memes will continue to exist rather than others. In theory, we could eventually develop a running catalogue of the cultural fitnesses associated with different types of memes. Moreover, we could obtain interesting results by observing the ways in which these fitnesses are altered if we combine several memes together or place single memes in varying cultural environments. (E.g. the faith meme may be extremely resilient in theological environments, yet it probably has a low cultural fitness within skeptical environments - especially if it is combined with other, highly implausible, memes.) A familiarity with cultural fitness would not constitute a revolutionary advance for our understanding of cultural information if the results simply confirm what we already know about certain ideas being easier to transmit than others. But if we discover that the fitness of certain memes is
significantly different than what we would otherwise expect, then this familiarity could prove to be exceedingly useful.2

Thus, a detailed outline of the characteristics contributing to cultural fitness would constitute a significant advance in our understanding of the forces acting on human thought and behaviour by making transparent the qualities that lead to our ideas' successful propagation of our ideas and precisely cataloguing their relative cultural fitnesses. The qualities responsible for the cultural fitness of one meme may be very different than those responsible for the cultural fitness of another (the pop song that sticks in my head for a week probably retains an annoying permanence for reasons that are quite different than the lasting quality of disturbing images on the news), but memetics could conceivably unravel these complicating factors and provide us with a detailed analysis of the memes to which we are most susceptible and the mechanisms that most efficiently enhance their propagation. Having access to this information would constitute an exciting new form of knowledge about ourselves, since we are inescapably creatures who depend heavily on the cultural environments in which we exist. Considering this potential, the question for the remainder of this chapter is how this striking new information about ourselves will affect the way we perceive the nature of ethics and its justification.

II. Ethics: Initial Reactions

An increased awareness of the evolutionary forces in our cultural environment suggests a large enough shift in our understanding of cultural phenomena that one might expect an equally significant shift in our views concerning ethics once the consequences of memetics have been fully taken into consideration. The level of precision promised by

2 As Daniel Dennett points out, the memes we expect to be good replicators (because we think they are true or beautiful, etc.) often are good replicators, because we would not have survived for long if there was no correlation between what we value and what helps us to navigate our cultural environments. Dennett therefore correctly points out that: "[memetics] becomes interesting only when we look at the exceptions". (Dennett 1990, p. 130)
memetics is enough to give the impression that revolutionary changes are in store for the structure and justification of ethical theory. Moreover, the dramatic reputation memetics carries for viewing minds as mere vehicles for virulent ideas lends itself to bold initial reactions when the combination of this theory and ethics is proposed. Three initial reactions are examined in this section. Each one leads to an application of memetics to ethics that is shown to be defective. However, the reactions are worth examining both because they arise so often in discussions of the potential connections between memetics and ethics (if you try to describe this thesis to others, be warned), and because their flaws highlight considerations that more fruitful attempts to blend ethics with theories of cultural evolution should keep in mind. Thus, taking the time to articulate exactly why these applications fail serves a worthwhile purpose by establishing a provisional context in which to discuss further, more profitable, proposals.

**Memetic Engineering**

With a detailed knowledge of cultural fitness available, the most obvious application of memetics to ethics is often thought to be (for better or worse) using our understanding of how ideas propagate to spread ethical memes in the most efficient way. By identifying the ideas that cause people to behave in a suitably ethical manner and determining which of these memes propagate most effectively, we could attempt to carry out a kind of *memetic engineering* with the goal of pragmatically enhancing those ideas that lead to what is generally considered to be ethically appropriate behaviour. In other words, if we discover that certain ideas are more effective at enabling people to respect, empathize and cooperate with one another, we could choose to actively promote these memes over their less proficient counterparts.

Of course, memetic engineering might not generate results as easily as we think: our experiences tampering with biological ecosystems has taught us that making abrupt changes to the constitution of complex, highly integrated systems often leads to
unexpectedly deleterious outcomes. By analogy, our efforts to promote the most efficient memes contributing to ethical behaviour might upset the current equilibrium of ethical norms in society and open up new avenues for agents bent on ignoring ethical standards to exploit those who obey the dictates of our specifically engineered moral memes. If we take the parallel with biological evolution seriously, moral memes exist within a carefully balanced and potentially fragile cultural infosystem - one we might be better off not interfering with.

Moreover, there may be structural considerations that will interfere with memetic engineering being successful. Genetic engineering is certainly complex enough, but in this substrate we at least have stable units of DNA to work with. In the case of memetics we have no such luxury because memes have such a high mutation rate and massively variable longevity. As a result, the environment in which memetics operates exhibits an increased level of complexity that might make the underlying structure of the cultural medium too diffuse to accurately manipulate for the purpose of particular engineering goals.

Nevertheless, it is at least plausible that memetics could overcome both of these difficult engineering challenges and become a powerful resource for the aim of constructing an increasingly ethical society. But would this constitute a significant contribution to the study of ethics? Despite the fact that memetic engineering might become an extraordinary new enterprise, it has very little to offer ethical theory in any meaningful sense. This may seem surprising considering how much potential the application appears to possess in terms of manipulating human thought and behaviour.

---

3 For an interesting game-theoretic version of this scenario, see (Danielson 1992, chapter 5.2). Danielson defends the rationality of specifically reciprocal cooperation after demonstrating that the introduction of undiscriminating cooperative agents into an evolving Prisoner's Dilemma tournament population actually results in the invasion of exploitive strategic maximizers, since these maximizers are particularly well suited to profiting from the resulting moral equilibrium. (Alternatively, my favorite version of this scenario is Lisa Simpson's seemingly foolproof wish for peace on Earth. The wish leads to the destruction of all weaponry and succeeds only in leaving the Earth at the mercy of an opportunistic alien invasion.)
Yet, there are two important reasons why memetics' vast potential will fail to affect the study of ethics in a similarly dramatic way.

First, memetic engineering is not nearly as new or revolutionary as it may seem. Once we dispense with the highly specialized terminology associated with memetics, anyone with even a cursory appreciation for the strategy involved in marketing, politics or military propaganda will recognize that conceptual engineering occurs on a daily basis. What are advertisements and public service announcements but memetic engineering experiments that attempt to modify our thinking or behaviour in specifically designed ways? Clearly, the technology of memetic engineering is already upon us whether we acknowledge the fact or not. To be fair, an established science of memetics could supply a new level of precision to our current mental engineering abilities. Yet it is not at all clear that lack of precision is the limiting factor explaining why most philosophers would not be keen to investigate the use of memetic engineering as a tool to promote ethical behaviour.4 Quite the contrary, it seems we currently have more than enough psychological evidence about ourselves to begin engineering contagious ethical memes if we wished to pursue the idea. The problem, rather, lies in deciding which contagious ethical memes we want memetic engineering to promote.

Hence, besides being not nearly as futuristic as it appears, the second key problem for memetic engineering's ability to affect ethical theory is the fact that the project relies on the assumption that we can agree on a set of ethical memes worth aggressively promoting. A memetic engineering perspective may yield an impressive amount of information about the kinds of ideas most likely to distribute copies of themselves, yet it

---

4 I say "most" philosophers to leave room for Richard Rorty who seems to be recommending something very similar to memetic engineering in his article, "Human Rights, Rationality, and Sentimentality" (1998). In this essay, Rorty suggests that we give up on the foundationalist search for a means of grounding human rights in theories of human nature. Instead, he recommends that we turn to the more pragmatic strategy of indoctrinating children with empathic qualities. The idea is that it is "easier for us to be moved to action by sad and sentimental stories" than by reason. (p. 185) The article has a strong memetic engineering flavour to it since Rorty is explicit about the fact that his reasoning is not based on advancing a better epistemic foundation. Rorty is interested only in spreading those ideas that will achieve the most effective results.
remains unable to determine the ideas we deem valuable enough to be specially promoted. Memetic engineering is simply not based on the dissemination of the ideas that would be generated by careful, collective deliberation. As those uncomfortable with the idea of memetic engineering will point out, there is no necessary connection between the memes to which our cultural environments are most vulnerable and the memes we expect to result from careful philosophical inquiry. The awe-inspiring power of memetic engineering is therefore tenuously dependent on our ability to agree in advance on the ethical principles we feel we ought to implement before we can appeal to this powerful new cultural technology.

It should thus be clear why ethics will not be dramatically affected by memetic engineering: the study of ethics is primarily interested in areas of dispute (or at least uncertainty) within our moral landscape, and on this front memetic engineering has very little to offer. This is not to say that ethics is interested exclusively in moral disagreement, but memetic engineering's ability to affect less controversial areas in ethics is compromised by the fact that new levels of precision will be wasted on our efforts to promote those memes broad enough to be widely agreed upon. If we decided to promote the attitudes, emotions and general dispositions necessary for individuals to act according to widely accepted ethical norms, we could start tomorrow by launching education campaigns or simply trying to ensure that children grow up in environments that effectively nurture their basic moral sentiments. This in itself could be labeled memetic

---

5 This is not strictly true of memetics in general. If one takes seriously the idea that memetics would be capable of mapping out every idea that passes through our brains, then it could potentially give us an account of even those ideas we are most likely to accept as ethically valuable after careful reflection. This possibility is set aside until later in the chapter.

6 One could argue that memetic engineering could have a direct effect on the primary considerations of ethics if one were a consequentialist, since one's basic moral maxims would be based on whatever strategies achieve the best results - a consideration that might be deeply affected by the impact of memetic engineering. However, a consequentialist outlook would still have to decide upon some underlying normative goal (the maximization of happiness, preference satisfaction, etc.) before memetic engineering concerns could be of much use. Thus, consequentialist ethics might be more affected by memetics than other theories, but the point remains that questions related to primary ethical concerns will be unaffected by memetic engineering's philosophically subordinate potential.
engineering of a certain type, but it hardly warrants the status of a revolutionary breakthrough. We could certainly try to use memetics to enhance these projects, but the impressive accuracy that is memetics' best asset would be wastefully lost on the goal of promoting the broad dispositions of which we generally approve.

Memetics is therefore caught between a rock and a vague place, so to speak. Either the goal of engineering ethical memes is plainly subordinate to the prior philosophical task of identifying memes deserving enough to merit being promoted, or else memetics' precision is squandered on promoting broad moral dispositions that we already have the means to effectively encourage. While it may still be possible that memetics will be able to refine our efforts at promoting acceptable ethical dispositions somewhat, it is unlikely that memetic engineering will supply the revolutionary improvement its provocative title might lead one to expect. Thus, the study of ethics remains largely unaffected by the promise of memetic engineering, even if it may seem as if this practice could succeed at modifying our thoughts in spectacular ways.

Memetic Determinism

The next proposal for the combination of memetics and ethics pushes the possibilities of a fully operational science of memetics one step further: what if we gain the ability to explain the propagation of ideas in all instances, including cases of careful deliberation? In other words, what if memetics can eventually provide an exhaustive account of the evolutionary forces responsible for the existence of every meme in our minds? The idea that this kind of self awareness could be realized may seem fanciful, but if we achieve a detailed understanding of the mechanisms involved in meme transmission it might someday be possible to provide an exhaustive evolutionary explanation for every thought that passes through our heads.

The second initial reaction to combining memetics and ethics is therefore a picture of the mind so deterministic it renders ethics obsolete. This may not be an
especially appealing contribution, but if we had access to the evolutionary history of every ethical intuition entering our minds, the standard justifications we give for believing in ethical principles (and mental causes in general) might suddenly appear superfluous. What sense would it make to say "I believe I ought to do x because y" when we can trace the evolutionary history of this belief and identify the selection mechanisms responsible for our brain replicating the meme in question within our conscious mind? Whatever y may be, it will seem arbitrary once a memetic description of the event suggests that this justification is only present in our mind because of the selection forces operating at the level of our neural physiology. The realization of memetics' full potential thus provides an analysis of mind so deterministic that it appears to undermine our confidence in the kinds of explanations required for ethical theory to be of much use. Once again, however, I think the revolutionary (and in this case rather alarming) reputation associated with memetics is greatly exaggerated. The deterministic vision of memetics just described is likely what philosophers fear most from naturalistic accounts of human cognition, but their fear is I think unwarranted. Even if a deterministic vision of ourselves as memetic 'hosts' is achievable, it will not necessarily eliminate the need for conceptual theories of ethics or human agency. The reason that conceptual (or 'ideational') accounts of ourselves cannot be so easily dismissed is that when it comes to reasoning *practically* about our actions, we simply cannot escape from thinking about our choices as if our ideas and justifications possessed an independence from the predetermined processes functioning at the level of memetics. If we managed to map out a fully deterministic perspective of how memes propagate in our brains, there is still no reason to believe that we could abandon the conceptual perspective of ourselves as intentional agents. For the latter perspective is a *practical perspective* that is responsible for deliberating about the justifications we give ourselves as reasons for action. No matter what we may discover from an empirical standpoint about the memes that fill our heads, we cannot avoid deliberating *as if* we have the capacity to operate outside the
constraints of a rigidly deterministic framework. In a sense, then, because we invariably view the world from a seemingly autonomous subjective standpoint, the ability to reason practically at the level of conceptual terms (rather than the level of determined memetic forces) can be thought of as a fundamental characteristic of human experience - one that will not be stripped away by the advancement of science into the workings of the brain.

This is tough talk for what is actually a timeless and extremely messy can of philosophical worms. The issue lies at the heart of compatibilist positions in the ongoing debate over free will, and it surfaces regularly in discussions of the implications of neuroscience for the philosophy of mind. It would therefore be presumptuous to think that I can offer a definitive solution to the issue in the context of memetics. However, I think we can provide at least a plausible response to the prediction that memetic determinism will render ethics obsolete, and this response is, I think, convincing enough that we should not fear an evolutionary view of mental phenomena when we address questions of ethical justification.

As one can probably infer from the discussion above, the response to the problem of mental determinism I favour is the solution provided by Immanuel Kant. In the *Critique of Practical Reason*, Kant argues for the primacy of our capacity to employ pure practical reason in accordance with the moral law (the condition that secures our freedom) and claims that this transcendental fact about our capacity for practical reason remains true beyond all possible doubt, no matter what we may discover about the natural world through the use of theoretical reason. This core message of the second *Critique* is instructive for the problem of memetic determinism: in virtue of the fact that we inescapably experience the world as rational beings (in so far as we can choose against our desires and we seek practical reasons for our actions), the transcendental fact of our freedom is unconditionally secured from a practical perspective - this no matter what we may come to learn about the structure of our minds from a theoretical perspective (the latter being a perspective where we can consider ourselves objects of scientific inquiry).
That Kant's argument can be extended to such futuristic examples as memetics is indisputable considering the foresight he displays about the possibility of investigating the causal forces responsible for human cognition:

One can therefore grant that if it were possible for us to have such deep insight into a human being's cast of mind, as shown by inner as well as outer actions, that we would know every incentive to action, even the smallest, as well as all the external occasions affecting them, we could calculate a human being's conduct for the future with as much certainty as a lunar or solar eclipse and could nevertheless maintain that the human being's conduct is free.

(Kant 1997, p. 83.5:99)

Hence, even if we achieve a startling theoretical understanding of the memetic influences acting on our minds, Kant provides a convincing argument for the fact that we can retain a practical perspective towards ourselves as agents with the capacity to provide undetermined reasons for our actions.

It may be argued that this seemingly inescapable fact about our practical freedom is the result of a failure on our part to imagine what it would be like to think in terms of deterministic memetics, since we currently have no way to imagine what it would be like to give up the primacy of our own intentionality. However, Kant's argument is compelling in that it seems highly unlikely that we could ever give up on the primacy of what we experience as practical freedom, whether we gain access to evolutionary explanations of how this experience is generated by memes or not. Whether we follow Kant in the ethical implications he subsequently draws from this claim is, of course, an entirely separate matter. What is interesting is the hypothesis that we can have access to two separate perspectives of ourselves - one deterministic and one where the experience of practical reason secures our freedom - without these perspectives necessarily contradicting one another.7

7 Thomas Nagel has famously provided a contemporary articulation of the view that we can retain subjective and objective perspectives of ourselves and that these perspectives cannot be reduced to one another. See (Nagel 1986), especially chapter seven.
In fact, even if one was unwilling to accept the metaphysical stance Kant takes in terms of claiming we are _truly_ free, the structure of his argument can be used to ground an elegant pragmatic defense of the idea that a deterministic account of the mind would merely force us to become 'ironic' compatibilists in the sense that we would inevitably continue to use conceptual justifications for our beliefs - in spite of the evidence suggesting that such justifications are illusory in terms of their causal efficacy.8 Why should we necessarily give up on our use of conceptual justifications? The fact that we understand the chemical reactions in the stomach and the signals that it sends to the brain does not prevent us from saying "I ate a sandwich _because I was hungry". Similarly, there is no reason to think that we will necessarily discard the conceptual explanations we appeal to for acting as intentional agents once we discover the underlying cultural selection processes responsible for directing conscious experience.

Even if memetics can produce a comprehensive evolutionary mapping of our thoughts, it is therefore unlikely that we would necessarily begin to think in terms of evolving memes to such an extent as to render ethics obsolete. This is admittedly a speculative conclusion since it is impossible to know for sure what it would be like to have access to such unique information about the way the mind functions. However, it is at least clear that while we wait for the necessary empirical evidence to fall into place, memetics contributes very little to the preexisting literature written on reductionist programs in the philosophy of mind and the possibility of free will. Hence, the revolutionary appearance surrounding memetics once again crumbles under scrutiny. Memetics might demonstrate someday that the kind of compatibilism outlined above is mere wishful thinking, but for now the memetic formulation of hard determinism

---

8 I use the word 'ironic' here to draw a parallel with a similar theme in the work of Richard Rorty. One could use the term "ironic compatibilist" or "memetic ironist" to indicate that we should continue to act as if our thoughts are not merely propagating memes (even if we know otherwise) in the same way that Rorty recommends that we act as if there is a solid metaphysical foundation for our linguistic practices while we simultaneously recognize the historically contingent nature of normative language. See (Rorty 1989).
contributes little, if anything, new to our understanding of the problem of free will and its relation to ethics.

**Memetic Skepticism**

The third initial reaction to the possibility of combining memetics and ethics follows a similar line of argument as memetic determinism. But instead of giving up on human agency entirely, this third proposal suggests that acquiring a detailed knowledge of the evolutionary forces acting on our ideas will specifically undermine our trust in ethical justifications. The concern is that by providing a detailed account of the cultural selection forces responsible for ethical memes propagating in our minds, memetics will dissolve the confidence we have in the authority of ethical claims (i.e. without necessarily unraveling our conception of ourselves as intentional agents). Thus, memetics may not cause us to abandon the notion of free will, but it could still lead us to believe that the reasons we normally give as justifications for moral behaviour are 'only' patterns of evolving cultural information. If so, it may seem as if these patterns are incapable of carrying the prescriptive authority we normally associate with their conceptual instantiations.

In other words, will we really feel the same sense of moral obligation in cases where we know, for example, that the intuition telling us that we ought to not lie to one another exists in our minds because cultural selection pressures have allowed this meme to survive and replicate successfully? If we fully accept the fact that our ideas are transmitted according to blind evolutionary principles, we may feel less secure about the normative status of moral obligations like this one. Why should we feel obligated to obey our moral intuitions if we know that they are only fleeting memes that have managed to propagate because of the contingent circumstances of our cultural environment?
In the case of memetics potentially undermining our faith in free will, our continuing to act as if memetics poses no threat to our conceptual level of experience is sanctioned by the fact that (as Kant so perceptively points out) it is virtually impossible to think of ourselves as anything but free agents when we deliberate from a practical point of view. Unfortunately, in the case of memetics undermining our faith in ethical justifications, the possibility of making choices contrary to the authority of our moral intuitions is all too familiar. As a result, a detailed understanding of the evolutionary forces acting on our intuitions could result in a sense of disillusionment creeping into our views about the prescriptive status of ethical claims - an outcome that would invite moral skepticism to take the place of the feelings of duty we normally associate with these claims.

In fact, there is a legitimate problem associated with memetics potentially eroding the authority our ethical intuitions carry in certain kinds of cases. (We will return to this problem in chapter five.) But with respect to memetics' ability to dissolve our faith in the prescriptive basis for ethical justification in such a way as to lead to general moral skepticism, I think that, once again, the radical consequences foreseen for the combination of memetics and ethics are greatly exaggerated. Memetics will not lead to general moral skepticism because an understanding of our moral intuitions as evolving memes will not necessarily lead to a depreciation of the normative authority that these intuitions/memes possess. It may be startling to discover that our thoughts have evolved out of a soup of cultural mutations and selection pressures, but this discovery alone should not be enough for us to dismiss the authority that our moral intuitions otherwise possess.

The key to responding to the threat of memetic skepticism is to target the rhetorical force of the word 'only' in such phrases as, 'Our moral intuitions appear arbitrary once we discover that they are only replicating memes existing in the neural substrate of our brains'. Why should the memetic status of our moral intuitions entail the
diminishment that is implied by the use of the word 'only' in such statements? Consider again the example of eating a sandwich because one is hungry, even though one is aware of the physiological processes at work in the stomach and the subsequent signals sent to the brain. We clearly do not think, 'I seem to feel hungry, but this feeling is meaningless because it is only the result of certain biochemical processes'. Similarly, there is no reason to necessarily abandon our moral intuitions once we understand their evolutionary status. It is certainly possible to be skeptical about our ethical intuitions in a way that seems unthinkable if we try to do the same for the experience of free will, but an awareness of the evolutionary forces acting on our intuitions will not automatically invalidate their normative status.

Without a doubt, there are some complex reduction issues lurking here concerning the identification of our thoughts as instantiations of memes. But it ought to nonetheless be clear that reductive explanations of mental concepts do not always lead us to abandon the concepts themselves. Thus, even if our thoughts are completely reducible to assemblages of memes, these memes and their mental instantiations are not necessarily arbitrary in a sense that ought to rob them of their prescriptive power. Memetics might uncover evidence about the evolutionary origins of our moral intuitions that undermine our confidence in particular memes that have especially arbitrary cultural genealogies (an option examined in section four of this chapter), but memetics should not lead us to be skeptical about the justifications for all our ethical memes just by demonstrating that they exist as units of cultural transmission.

---

9 For an introduction to these complex issues and a defense of the view that explanatory reductions need not lead to the rejection of mental concepts, see (Putnam 1975).

10A glaring (yet all too familiar) example of how not to understand the relationship between our mental concepts and their memetic interpretation can be found in the opening lines of Anthony O'Hear's review of The Meme Machine by Susan Blackmore: 'If what The Meme Machine says is true, then there is no reason to believe it. For what it says is that things are believed only because they succeed in installing themselves in our brains through various tricks in order to get themselves reproduced.' (O'Hear 1999, my emphasis). The inflammatory use of the word 'only' creeps in here with a vengeance. O'Hear completely fails to grasp that memetics is an account of what it means to believe that something is true. More importantly, he unfairly assumes that our beliefs are necessarily unjustified if they are interpreted as units of cultural selection.
Consider a second example: the fact that sociobiology provides us with a genetic explanation of why we care so much for our children. Our understanding of the biological forces contributing to our feelings towards our children in no way undermines the normative influence of the feelings themselves. Of course, if we find that these genetic dispositions produce pathological side-effects in certain situations we may want to guard against such results, but the mere fact itself that our feelings towards our children are influenced by biological factors is not enough to undermine our commitment to these feelings. Hence, most parents who are familiar with sociobiology do not think, 'Well, I thought I cared deeply about my children, but I guess this feeling is really only a product of my genes.' Similarly, unless we find something specific in the cultural history of our moral intuitions that gives us reason to alter the confidence that we place in them, the mere fact that these intuitions exist as evolutionary products is not, in itself, enough to extinguish their normative authority.

Those unconvinced that memetic skepticism can be avoided in this way may want to argue that ethical claims constitute a special case and cannot be compared to our continued reliance on the experience of hunger or the intuition that we ought to care for our children. The difference, presumably, is that in these examples we have basic motivating desires impelling us to continue to respect the authority of our intuitions (or our phenomenological experience in the case of hunger), whereas in cases where ethical imperatives are at stake our self-interested desires routinely conflict with the authority that our moral memes possess. However, it is surely a mistake to think that memetics must live up to the expectation of being able to convince us to be moral in cases where our interests lie elsewhere. Why pick on memetics for not providing a metaphysical framework that can answer to the question, 'Why be moral', when this question presents an irresolvable dilemma for nearly any metaphysical framework?

The sense of disillusionment that results from discovering the memetic basis for our moral intuitions is therefore only created if we cling to unrealistic expectations about
what is required from the metaphysics of a legitimate source of moral authority. The question, 'Why should I respect the authority of my moral intuitions if these intuitions are only memes?' is predicated on the naive expectation that some other metaphysical framework exists where our moral intuitions are manifestly binding, and this expectation clearly demands too much in the way of meta-ethical foundations. If memetics makes it less likely that our moral intuitions are grounded in a metaphysical foundation capable of solving the compliance problem in meta-ethics (indeed, it cannot even prove that such a foundation is not coincidentally true), it still does not leave us any worse off in terms of investigating contemporary approaches to meta-ethics that have given up on the idea that such a foundation exists. Memetics is therefore no less meta-ethically sound than any other naturalistic (or even broadly mechanistic) metaphysical framework, even if the fact that it is ambitiously clear about how this naturalistic view of the world functions makes it more intimidating.

The important point is that memetics will not dissolve the prescriptive authority of our moral intuitions simply in virtue of the fact that it reveals the evolutionary character of these intuitions. Memetics provides an evolutionary articulation of the view that our ethical intuitions originate from metaphysically arbitrary origins, but this result adds little, if anything, to the existing debate over skepticism and the search for foundations in meta-ethics. The unique evolutionary histories of our ethical memes may give us reason to reevaluate their prescriptive status in certain circumstances, but the fact that they have evolutionary histories at all is not in itself enough to deny (or affirm) their normative authority. Once again, the revolutionary consequences foreseen for memetics and ethics seem to crumble under philosophical scrutiny.

III. Memes and Moral Realism

The reactions examined in section two foresee dramatic changes to the study of ethics, yet each proposal fails to convincingly demonstrate that it is capable of achieving its
intended result. The proposals touch on very deep philosophical problems which I don't pretend to have conclusively resolved. (This would be an enormous task that is well beyond the scope of the present inquiry.) Instead, I have suggested directions in which these dramatic applications of memetics to ethics ought to be resisted. In doing so, I have also hopefully demonstrated that memetics will not alter the existing philosophical research on the relationships between ethics and social engineering, determinism and moral skepticism in any meaningful way.

Before moving on to what I believe to be the most persuasive application of memetics to ethics, another proposal deserves consideration that is more plausible than those discussed in the previous section. This is the proposal that the study of memetics will support a naturalistic form of moral realism by clearly articulating the empirical facts of our relationships with one another as moral agents. If one takes the premise of moral realism seriously, detailed theories of cultural evolution could prove to be exceedingly useful because they will supply a rich level of content to the hypothesis that moral properties exist as empirical entities in our social environments that are open to scientific investigation. Hence, it is possible that memetics will make a contribution to ethics by mapping out the cultural connections of our social relationships and thereby specify the objective standards of what may be interpreted as the 'real' facts of our moral landscape.

The Resurrection of Moral Realism

For the greater part of the twentieth century, moral realism has remained a thoroughly unpopular meta-ethical theory, yet in the past two decades this theory has resurfaced in the work of many prominent philosophers and acquired a renewed level of academic respect. The basic hypothesis of moral realism remains unchanged - that moral claims represent objective, independent, truth-conditional facts about the world (or
universe). Yet new versions of the theory situate this hypothesis within a more plausible ontological framework than their soundly discredited predecessors. By appealing to contemporary developments in epistemology and the philosophy of mind, new theories of moral realism have adopted a *naturalistic* approach to the ontology of moral facts that has allowed significant progress towards explaining how these facts generate their unique action-guiding character without relying on the existence of entities that would be, as Mackie notes, "metaphysically queer". The result has been an explosion of interest in moral realism, and even those suspicious of the claim that new versions of the theory can overcome the problems that extinguished its older variations have been forced to recognize that moral realism must once again be taken seriously.

The leading contemporary advocates of moral realism are David Brink, Peter Railton and Richard Boyd, although several other noted theorists have recently proposed either explicitly realist theories or notably sympathetic points of view. Brink, Railton and Boyd each give their own unique articulation of moral realism, yet one finds a striking similarity in their methodologies, which reflect the deeply naturalistic character of moral realism's new outlook.

Brink's view combines a nonreductive naturalism with both a coherentist moral epistemology and an externalist moral psychology in order to produce an argument for the idea that moral realism is *supported* by the notion that moral claims derive their action-guiding character from widely shared psychological facts. (Brink notes that this is an unusual argument, since the action-guiding character of moral experience is usually

---

11 I take these three attributes to be those most commonly accepted as essential to moral realism, although Geoffrey Sayre-McCord has given a convincing argument for the claim that moral realism requires embracing just two theses: "(1) moral claims, when literally construed, are literally true or false (cognitivism), and (2) some are literally true." (Sayre-McCord 1986, p. 3) The second thesis is required to rule out cognitivist error theories such as the one defended by J. L. Mackie (1977). All other conditions, Sayre-McCord claims, arise only in the context of defending these two core principles.


13 See (Brink 1986, 1989), (Railton 1986) and (Boyd 1988). Further references to Railton and Boyd will refer to these works.

14 These include: (Dancy 1986), (Sturgeon 1985), (McDowell 1988) and (Norman 1997).
appealed to in arguments against moral realism.) Similarly, Railton defends a conception of moral realism that is based on the empirical details of what can be considered "instrumentally rational from a social point of view". (p. 200) Here, what is socially rational is grounded in (but because of interpersonal dynamics not directly reducible to) an idealized account of individual rationality, the normative status of which is ultimately traceable to the shared conditions of our human motivational system. Finally, Boyd argues for a close analogy between moral realism and a nonfoundationalist interpretation of scientific realism. On his view, moral terms are compared to natural kind terms and defined according to their reference to natural properties that are described as rigid "homeostatic property clusters" (p. 197). Despite its ostensible complexity, Boyd's interpretation of the meaning of moral terms is based upon the familiar epistemological notion of semantic reference being a matter of appropriate causal connections between natural facts. In the case of moral terms, Boyd argues that the natural facts in question are the shared conditions of our moral psychology and social context.

The most important similarity underlying these three theories is the turn towards a naturalistic basis for deriving the objective facts of morality - one that is contingent on the details of our moral psychology and our methods of social organization. Rather than postulating moral facts that exist independently of human beings entirely, new versions of moral realism postulate moral facts that are objective in the sense that they are independent of our individual subjective opinions at any particular point in time. Our moral beliefs may be objectively true or false, but moral realism now judges their truth conditional nature to be a matter of whether the beliefs in question cohere with the appropriate conditions of our collective moral psychology and social environments (rather than the appropriate metaphysical conditions of the universe). As Railton tells us:

A teacher of mine once remarked that the question of moral realism seemed to him to be the question whether the universe cares what we do. Since we have long since given up believing that the cosmos pays us any mind, he thought we should long since have given up moral realism. [...]

However, the account offered here gives us a way of understanding how moral values or imperatives might be objective without being cosmic. They need be grounded in nothing more transcendental than facts about man and his environment, facts about what sorts of things matter to us, and how the ways we live affect these things.

(Railton 1986, pp. 200-1)

Thus, the naturalistic outlook grounding new versions of moral realism offers the theory far more contemporary plausibility. In a certain respect, of course, older conceptions of moral realism (with the notable exception of Moore's) can still be described as naturalistic in that they argue for moral facts existing in the natural world with or without our presence in it. Yet, newer versions of moral realism are specifically naturalistic in that they rely on a naturalistic understanding of human thought and behaviour to generate the basis for objective moral claims. It is this more specific form of naturalism - one which situates naturalistic moral facts within the observable realm of human affairs - that has given moral realism the renewed sense of credibility it now enjoys.

Memetic Moral Realism: the Role of Social Theory

The switch to a naturalistic basis for moral realism is important not only because it makes the theory more plausible, but because it emphasizes the need for increased empirical investigation in the social sciences. If the objective facts of morality depend on the details of our evolved psychology and our methods of social organization, then we have reason to take up a more rigorous empirical approach to these facts about ourselves and our social arrangements. As Boyd admits of his view, "Like any naturalistic account, [the substantive conception of morals offered here] rests upon potentially controversial empirical claims about human psychology and about social theory." (p. 202) For this reason, theories of cultural evolution might finally begin to play an essential role within ethical theory. By precisely outlining the evolutionary forces contributing to our social impulses and cultural environments, a science of memetics can help us establish a better understanding of the natural facts that serve as the basis of our moral experience.
Consequently, memetics may become a powerful resource for moral realism to draw upon in order to clearly specify the empirical standards of what this theory interprets as our objective moral reality.

In this context, the precision promised by memetics finally suits the task for which it is intended. This stands in stark contrast to the earlier suggestion of using memetics to support ethical social engineering. In that case, appealing to memetics seemed fundamentally at odds with the normative character of ethical theory, and the combination of the two gave the impression of trying to fit the square peg of detailed evolutionary information into the round hole of prescriptive ethical debate. In the case of moral realism, however, the precision with which memetics professes to be able to supply us seems well tailored to the task of studying the naturalistic connections between our shared moral dispositions and our social relationships. Indeed, if the parallel with scientific realism and its commitment to empirical observation is taken seriously, then a rigorous account of our social relationships of at least *some* form will be required. In Boyd's words:

According to [my] homeostatic consequentialist conception of morals (indeed, according to any naturalistic conception) goodness is an ordinary natural property, and it would be odd indeed if observations didn't play the same role in the study of this property that they play in the study of all the others. According to the homeostatic consequentialist conception, goodness is a property quite similar to the other properties studied by psychologists, historians, and social scientists, and observations will play the same role in moral inquiry that they play in the other kinds of empirical inquiry about people.

(Boyd 1988, p. 206)

Considering the fact that an observational social theory will be required to fulfill the role in moral inquiry Boyd refers to - one which will presumably map out the cultural and psychological characteristics of our shared moral experience - then memetics appears to be a perfect candidate. As a vehicle for the scientific investigation of moral phenomena, one can hardly think of a better tool to allow us to discover and accurately define the
objective basis for our natural moral properties. Thus, if we are willing to accept the terms of moral realism, it seems reasonable to think that the empirical potential supplied by memetics will have a profound effect on the lucidity of our efforts to determine the content of our objective ethical values.\footnote{In fact, projects remarkably similar to this suggestion have already been proposed. In \textit{The Biology and Psychology of Moral Agency}, (1998) William Rottschaefer makes no reference to memetics or even cultural evolution, but he defends a naturalized view of moral agency that employs our sociobiological moral capacities, a behavioural account of moral learning and supervenient mental properties emerging from our neurophysiology in order to support an argument for a realist conception of the ontological status of moral values. Rottschaefer's project is strikingly parallel to the application of memetics to support a naturalistic conception of moral realism. (The less than convincing appeals to Skinnerian behavioural psychology in the book might therefore be profitably replaced with a memetic account of how we acquire culturally-specific moral variants.) Also, William Harms (1998) has proposed using a bio-semantic theory of signaling and the adaptive histories of our moral capacities to support a form of moral realism concerning normative human functions. Again, Harms makes no explicit use of memetics (on his account culture simply generates an important source of sociobiological variability and carries no direct role in adaptive explanations), but his appeal to an evolutionary naturalism to ground a realist conception of moral values is extremely similar to the proposal at hand.}

\textit{Why Memetic Moral Realism is Prescriptively Unhelpful}

I have separated this proposal from those presented earlier because it is not obviously deficient in the same way as the initial reactions discussed in section two. Yet it is far from clear that a memetic conception of moral realism will meaningfully contribute to the study of ethics. Memetics is at least plausibly well suited to the proposed application of studying the empirical evidence for natural moral facts. (Moreover, considering the rising popularity of naturalistic interpretations of moral realism, this proposal could gain considerable support if models of cultural transmission become widely accepted.) But I think careful reflection suggests that this application of memetics to ethics is not nearly as profitable as one might think.

In keeping with the heuristic device of a best-case scenario, the aim of my argument against a memetic conception of moral realism is to show that if we imagine the implications of a fully operational version of such a theory it will be inevitably pushed in one of two directions: either it will be morally untenable or it will be prescriptively
unhelpful. Either way, memetics will not end up providing the valuable contribution to ethics that the combination of a realist ontology with a memetic empirical social theory might lead us to expect. Again, the issue at stake touches on a deep problem in contemporary philosophy, and I do not intend to settle the general problem at stake in the context of memetics. (Resolving the debate between realism and anti-realism is beyond both my ambitions and abilities.) Rather, my objective here is to show that memetics will not advance the existing debate in any meaningful direction, and that the heightened sense of precision it offers will not lead to moral realism appearing any more plausible to those not previously convinced of its prescriptive merits.

To begin, imagine what we can expect from a best-case scenario in terms of using memetics to investigate the properties of our social environments that moral realism takes to be the empirical facts of moral experience. Results will vary depending on the particular conception of moral realism we employ, but a best-case will involve our having mapped out the appropriate natural facts of our moral psychology and social relationships to such an extent as to be able to potentially determine the truth or falsity of our moral claims. It is probably too much of a stretch for our imagination to contemplate what it would be like to determine the truth or falsity of all our moral claims (this would be less a best-case scenario than an extreme logical possibility - perhaps even a condition for being God), but imagine that memetics succeeds at providing enough empirical information for us to determine the truth of at least some moral claims.

What exactly would this look like? It is difficult to say for certain, as none of the principal defenders of naturalistic moral realism spell out in detail how the social functions supporting their empirical approach to moral facts will operate. (Brink is mostly interested in the general possibility of externalist moral realism, and both Boyd and Railton cautiously point out that they are only outlining programmatic forms of moral realism (Railton, p. 163, 204), in order to broadly "establish its plausibility and to offer a general framework within which further defenses of moral realism might be understood".
The result is a marked lack of concrete examples of how social theory will reveal the empirical details of our moral experience. However, we can speculate that by revealing the selection forces operating in our cultural systems, memetics will allow us to observe the kinds of moral memes of which we tend to approve (i.e. replicate and distribute to others) under ordinary conditions. Thus, memetics will uncover the cultural filters responsible for shaping our shared social norms, and it will presumably identify the specific social functions that moral realists interpret as observable natural entities, whether these be 'homeostatic clusters' or examples of 'social rationality'. In this sense, memetics can be described as explicating the natural properties, or 'real moral facts' of the social environments we inhabit.

For example, we might find that the cultural selection pressures acting in our social environment overwhelmingly favour the meme that it is wrong to lie to others. Suppose that this meme is almost always replicated without hesitation, and suppose that when it is not, its denial is almost always the result of some recognizable cultural distortion or temptation displayed on the part of particular agents in particular cases. In this case, one could plausibly assert that memetics has allowed us to observe the relevant natural properties of ourselves and our social surroundings, and that the most coherent explanation of these properties is that the principle that it is wrong to lie to others is an objectively true fact about our social existence.

The problem for combining memetics and moral realism in this way is best illustrated by asking the question of whether an investigation into the properties of our social systems can uncover other moral 'facts' like this one that are nonetheless morally unacceptable. This possibility represents a self-contradictory result, yet if moral realism genuinely presumes that moral properties are empirical entities that can be observed like any other natural properties, then there is no obvious reason why the results of our empirical investigations could not produce results that we would find unsatisfactory, perhaps even repugnant. This seems especially true if we view our social environments
as evolutionary systems. If the social facts of our surroundings are generated by an imperfect cultural evolutionary process, then there is no obvious reason to believe that the moral norms produced by this process will necessarily be optimal, or even satisfactory.

For example, what if our empirical observations of the existing cultural selection forces in our social surroundings reveal that the best interpretation of our shared moral norms include 'it is wrong for women to seek vocations other than rearing children' as an objectively true moral fact. This would clearly be a troubling outcome. (As would similar conclusions that in certain social systems, 'slavery is acceptable', 'torturing enemies is permissible' and 'disabled persons are less valuable' are all true moral facts.) Yet, recall that naturalistic forms of moral realism nonetheless postulate that "there are moral facts and true moral propositions whose existence and nature are independent of our beliefs about what is right and wrong." (Brink 1986, p. 24 my emphasis) If these facts are truly independent, observable properties of the structure of our social systems, why should we have any a priori guarantee that memetic empirical inquiry will discover virtuous results?

Memetic moral realism has, I think, two available options left open to it in order to respond to this problem. First, it can bite the bullet and concede that outcomes of this kind represent legitimate possibilities. This option remains admirably consistent with the project of grounding the objective facts proposed by moral realism on the details of our natural, contingent human moral capacities and unique social situations. It therefore fiercely lives up to the spirit of moral realism's updated ontology by refusing to cave in to the perfectionist elements present in older conceptions of the theory. Nevertheless, this option is clearly self-contradictory on any reasonable interpretation of what we ought to expect from what we want to describe as moral facts. Hence, it is prescriptively
unacceptable, and its inadequacy is self evident to most proponents of naturalistic moral realism.\textsuperscript{16}

The second option available to memetic moral realism is to deny that counter-intuitive outcomes can be candidates for objective moral facts. This option salvages the prescriptive credibility of real moral facts, but it requires an explanation of why it is that the relevant properties generated from our evolved moral psychology and the circumstances of our social arrangements cannot turn out to be unsatisfactory. Unless we subscribe to a naive functionalism with respect to the biological and cultural evolution of natural properties, how can we be sure that what emerges from this naturalistic process will necessarily match up with our informed expectations about what ought to refer to 'true' moral facts? Thus, this second option owes us an explanation of how empirically observable moral entities will always prove to be acceptable without begging the question of the kind social evidence we can identify as discernibly 'moral'.

A provisional explanation can be found in the version of moral realism defended by Railton. Though he is obviously not concerned with using memetics as an empirical social theory, Railton is well aware of the danger of having moral realism depend on a 'complacent functionalism' (p. 199) by placing our faith in the mechanisms of (biological) evolution. Railton's theory avoids the charge of complacent functionalism, however, because he bases his definition of moral facts on a hypothetical conception of what is socially rational. This conception is described as an idealization of social rationality, and is defined as "what would be rationally approved of were the interests of all potentially affected individuals counted equally under circumstances of full and vivid information." (p. 190) In this respect, what is socially rational (and thus an objectively true moral fact) is still arguably an empirical question. But because Railton's conception of social

\textsuperscript{16} No so for everyone, however. Peter Danielson has vigorously defended something like this first option (personal communication) by claiming that whatever we can 'grow' within evolutionary stable moral ecologies is enough to be reasonably interpreted as a justifiable moral principle. Similarly, the position outlined by William Harms (1998) presumably implies that whatever biofunctional signaling mechanisms happen to evolve as adaptive human social capacities ought to be considered objectively true.
rationality includes hypothetical conditions regarding full information and the inclusion of all affected individuals, the empirical outcomes it identifies as true moral facts are virtually guaranteed to be satisfactory. A memetic version of Railton's moral realism could therefore adopt these hypothetical conditions in its criteria of the relevant coevolutionary social forces to observe and thereby avoid the accusation that it will identify prescriptively unacceptable empirical outcomes as morally true.

Though it is obviously more appealing than legitimizing deplorable social circumstances as real moral facts, the problem with this second option is that it turns what is morally true into a hypothetical moving target - one that, despite its empirical pretensions, cannot be measured in such a way as to give our social observations any increased credibility in terms of the prescriptive authority of the moral 'facts' these observations reveal. The fact that we can use Railton's idealized conception of social rationality to doubt the authority of our existing social properties is an asset if we are worried about identifying unacceptable cultural circumstances as moral facts. But if we give ourselves the capacity to doubt the results of our empirical investigations in this way, then we can always deliberate further about what would be socially rational under more idealized conditions. If this is the case, then the increased precision of memetics as an empirical social theory is rendered impotent by the fact that its results can always be overturned by appeals to more authentic empirical conditions.

In short, if the conditions required to identify objective moral properties are sufficiently idealized to ensure that what is true and what is prescriptively justifiable cannot come apart, then it is unclear how we can ever come to know when we have access to genuinely objective moral facts. Because this second option effectively admits that even our best empirical estimations of observed social convergence are not reliable enough to secure the truth of moral claims (without simultaneously admitting objectionable social traditions), separating the truth of moral claims from our best empirical estimations leaves moral realism hanging without a clear decision-making
recourse to determine the prescriptive status of putative moral facts. Stripped of the ability to appeal to the natural properties of existing social systems in order to provide determinate moral judgments, naturalistic moral realism loses a considerable amount of its prescriptive impact. Thus, by dodging a bullet few are willing to bite - accepting disturbing social situations as candidates for objective moral facts - the second option available to memetic moral realism opens itself up to the complaint that it can always revert back to a more idealized conception of the 'true' natural moral properties of a situation. making the notion of moral truth a slippery enough target to be prescriptively unhelpful.

*The Open Question Argument Rises Again*

Those familiar with the existing literature on moral realism may recognize the preceding argument as a variation on G. E. Moore's infamous *Open Question Argument* which pretty much annihilated earlier naturalistic conceptions of moral value at the turn of the century. Moore's argument specifically targeted claims to *analytic* ethical naturalism. His aim was to demonstrate that the predicate 'good' (a quality Moore considered to be the fundamental subject matter of ethics) referred to a simple, unanalyzable quality - one which cannot be defined in terms of any other natural characteristics. The Open Question Argument states that the predicate 'good' cannot be defined because one can always ask whether the subject matter of the definition offered is itself good without this question being insignificant. Hence, if the good is defined, for example, as 'X', then the question, 'Is X itself good?' is not a meaningless question, and this ensures that the predicate 'good' cannot be analytically defined.17

My intention is certainly not to follow Moore in concluding that the good is a real but non-natural property. Rather, I simply want to point out that the argument presented

---

17See (Moore 1994), chapter one.
above against memetic moral realism can be interpreted as being *in the spirit of* Moore's argument, because it states that one can always ask of what is presented as an empirically determined natural moral fact whether this fact has been observed under sufficiently idealized conditions to be genuinely considered justified. If memetic moral realism were to claim that 'Natural property X is an objective moral fact because empirical evidence tells us that this fact accurately reflects the details of our social situation', one will always be able to ask 'Is X true?' without the question being insignificant, because it will remain an open question whether this property is the product of sufficiently idealized conditions to command prescriptive authority as an authentic moral fact. This updated version of Moore's argument does not seek to demonstrate that moral facts exist within a separate metaphysical realm. (We are all happy naturalists here...) Instead, it points out that what we observe to be natural moral properties are not determinate enough, without decidedly non-empirical reflection about appropriate conditions, for moral realism to help us deliberate about issues of prescriptive justification, no matter how much empirical evidence about our social environment we may acquire.\(^\text{18}\)

It is important to note, however, that this argument does not prove that memetic moral realism or more general articulations of naturalistic moral realism are false.\(^\text{19}\) What I hope to have shown is that the prospect of using memetics to fill in the empirical content of moral realism will not help the recent movement towards naturalistic conceptions of the theory to achieve an increased level of authoritative precision with respect to identifying prescriptively justified, 'true' moral claims. How damaging one finds this conclusion depends on one's interest in the issue of moral realism. If one is primarily interested in moral epistemology, then memetic moral realism may nonetheless

\(^\text{18}\) This is a poetic result in that a more empirically robust version of the resurrected naturalistic theories of moral realism currently in fashion is shown to be defeated by a resurrected naturalistic interpretation of the argument that originally brought down older conceptions of ethical naturalism in the first place.

\(^\text{19}\) For more general arguments against recent versions of moral realism, see: (Copp 1991), (Rosenberg 1990) and (Horgan and Timmons 1991).
provide an interesting perspective on what it means to deliberate over the merits of justifiable moral facts. Moral realism offers a compelling descriptive interpretation of how objective ethical properties can be reconciled within a naturalistic framework, and it is therefore well suited to providing an account of what we can know about our ethical claims within a naturalistic ontology.

Yet if one's interests lie in questions of prescriptive justification (as mine do), then the precision foreseen for memetics as an empirical social theory is wasted, for this increased precision will not lead to a greater capacity for moral realism to provide determinate prescriptive recommendations. Although memetics may be able to supply us with valuable information about the shared moral norms that flourish in our social systems, it will not have an impact on what we ultimately sanction as prescriptively justified because the concept of truth provided by moral realism is sufficiently idealized to depend less on matters of direct empirical investigation than on questions of hypothetical reflection on empirical matters. Once again, we find that the substantial potential promised by memetics is unable to supply a meaningful contribution to the existing study of prescriptive ethical theory.

IV. A New Genealogical Resource

It seems at this point as if ethics will emerge from a confrontation with memetics completely unscathed. The initial reactions to combining memetics and ethics and the prospect of using memetics as an empirical social theory for moral realism both appear to have no substantial impact for prescriptive issues in the study of ethics. This might be enough to suggest that the fusion of memetics and ethics will not alter our ordinary ethical judgments, but perhaps we have been asking for too much by looking only at extreme possibilities. In this final section of chapter three, I propose that memetics does in fact have something substantial to contribute to ethics, but that this contribution is more subtle than the options discussed so far. Instead of seeking a way for memetics to
change the entire structure of ethics, I believe more profitable results can be obtained by situating the intricate empirical information supplied by memetics within the scope of our existing efforts to systematically organize our justified moral intuitions.

The proposal I have in mind consists of using the unique genealogical information provided by memetics as a resource for reevaluating the confidence we place in the authority of our moral judgments. By exposing the cultural evolutionary histories of our moral intuitions, memetics will allow us to identify and potentially reevaluate those aspects of our intuitions that have evolved because of structurally anomalous cultural selection pressures - pressures that promote arbitrary moral intuitions beyond their justifiable authority. As we have seen in the first section of this chapter, memetics will not lead us to discard the authority of ethical claims simply because this knowledge reveals that the origins of these claims and our moral intuitions are *metaphysically* arbitrary. However, it is nevertheless possible to use memetics to weed out certain memes that are *especially arbitrary* from inside what we acknowledge to be a naturalistic ethical framework. It would be foolish to eliminate the entire basis for ethical justification because it appears unsuitably contingent on a cultural evolutionary system. Yet it might make perfect sense to see certain evolutionary origins as more eccentric than others and to therefore attempt to purge ourselves of these memes within the limits imposed by an admittedly contingent (and constantly evolving) ethical perspective.

Thus, memetics will contribute to ethics by helping us single out particularly suspicious memes that propagate because of unusual (and often unperceived) selection pressures. The ability to identify these memes and reveal how they distort our moral intuitions will allow us to systematize our ethical judgments in a more consistent manner, since it provides us with the opportunity to avoid situations where the source of a meme's cultural fitness has become detached from the selection forces we normally associate with the reasons for its ethical justification.
The strategy behind this proposal should be a familiar one. Indeed, it is directly parallel to the strategy outlined in chapter one with respect to current versions of Evolutionary Ethics. There it was proposed that we use the empirical information provided by sociobiology to either reinforce or resist our genetic social dispositions depending on whether the genealogies of these dispositions demonstrate that they have been selected for reasons that can be reconciled with contemporary moral standards. In this chapter, the proposal is that we use the information provided by theories of cultural evolution for a virtually identical purpose: to either uphold or eliminate moral memes depending on whether their genealogies demonstrate that the (cultural) selection pressures responsible for their existence are consistent with the normative authority they otherwise seem to possess. Although the substrates differ, in both cases the objective is to rid ourselves of unwanted biases in our moral intuitions by appealing to facts about the evolutionary origins of these intuitions.20

Weeding Out Cultural Distortions

The key to understanding why this application of evolutionary theory to ethics will be profitable lies in the realization that selection forces, both cultural and biological, do not necessarily lead to optimal results. This fact is a central tenet of modern articulations of Darwinism, yet the point deserves constant re-emphasis because it is so easy to tacitly assume otherwise. Many evolutionary adaptations appear as if they have been designed for a specific function (the intricate construction of the heart being a classic example), and the fact that these adaptations serve our best interests more often than not can lure us into believing that this will always be the case. But we must bear in mind that this is a seductive illusion. Chronic back problems, our unhealthy preferences

20 Although it has been noted earlier (chapter 2), it is worth reiterating that these two substrates are deeply interconnected. They are treated as distinct processes here for the sake of clarity about what we can learn from cultural selection forces, but obviously these forces will be greatly influenced by biological concerns and vice versa.
for sugar and fat, and the fact that the appendix seems to have no function whatsoever anymore all serve as instructive reminders that biological selection forces cannot be trusted to secure our well-being in contemporary circumstances. Similarly, chronic political problems, our unhealthy preferences for simplicity and peer approval, and the fact that so many cultural traditions seem to have no function whatsoever anymore all serve as reminders that cultural selection forces cannot be trusted to secure our welfare anymore than their biological counterparts.

The heightened sense of genealogy afforded by memetics will therefore allow us to weed out instances where cultural selection has for some reason generated undesirably arbitrary results. These instances might include endemic cultural biases, leftovers from past social environments, or perhaps unperceived side effects from otherwise acceptable patterns of thought. Take, for example, the fact that as cultural beings we are incredibly impressionable, particularly as children. This characteristic is generally advantageous, since it allows us to learn vast amounts of information about our environment very quickly. However, in a contemporary technological setting this characteristic can produce a deluge of cultural side effects because, as any parent can verify, children are exposed to so much material that is not in their best interest to internalize. In this respect, our otherwise beneficial capacity to accumulate cultural information can easily be hijacked by societal cultural selection pressures that promote harmful cultural variants.

Furthermore, consider an example where a more specific disposition is hijacked by the circumstances of a new cultural environment: the tendency to vigorously stand one's ground when engaged in disputes with others. In a small cultural system one might rarely find oneself in heated disputes, and if such disputes did occur the issue at stake might often be important enough to make a disposition to stand one's ground a worthwhile strategy. If so, this underlying disposition would likely produce - and be highly reinforced by - such memes as "stand up for what you believe" or "trust yourself" (or simply "take pleasure in being bull-headed"). Yet if these memes persist in
environments where disputes become more frequent (perhaps because of increased interaction with unfamiliars), they can become highly unproductive, if not outright dangerous, depending on the risks associated with both parties refusing to back down.21

Similarly, many other memes can become deleterious if they persist in situations where the underlying factors contributing to their cultural fitness are no longer relevant. Even such seemingly trustworthy memes as "emulate those you admire", "respect authority" and "uphold the standards of your community" can lead to disastrous results if their influence extends to situations for which they are badly suited. (E.g. the "respect authority" meme is only as reputable as the authority in question.) The challenge, then, is to distinguish deleterious situations from those where a meme's cultural fitness is still linked to forces we would continue to accept if we were fully aware of their influence upon us.

Memetics provides us with a means of making these distinctions by giving us intricate genealogical information about the evolutionary history of our ideas that can be used to critically access their trustworthiness. In the context of ethics, memetics will potentially allow us to distinguish cases where the cultural influences acting on our moral intuitions are distorted from situations where we would continue to collectively embrace these influences after they have been identified and their consequences evaluated. By rendering the cultural selection forces acting on our intuitions transparent, memetics will enable us to evaluate the cultural selection pressures shaping our moral intuitions and eliminate moral memes whose cultural fitness depends on factors that have become historically obsolete or radically disengaged from their advertised ethical function. Just as we attempt to suppress genetic dispositions that are selected for purposes that are no

---

21 Many will recognize this as a cultural articulation of the hawk/dove scenario in game theory. In the dispute over a resource, hawks will escalate the conflict until injured or their opponent retreats, whereas doves will initially contest the claim but back down if further conflict ensues. (See Maynard Smith 1982, pp. 11-20) As many philosophers have pointed out, in a contemporary environment where conflicts involve the possibility of nuclear weapons, the situation where two hawks meet can be extremely hazardous.
longer consistent with contemporary moral standards (e.g. dispositions for aggression, sexual jealousy, etc.), memetics will give us the power to wean ourselves from memes that propagate because of cultural influences that are inconsistent with what we perceive to be morally valuable.

**Recognizing Distorted Intuitions**

The application of memetics just described is, I hope, appealing, but the following question immediately presents itself: how exactly will we recognize cases where the cultural influences on ethical memes are inconsistent with the selection pressures we expect to lead to morally valuable results? In particular, how will we recognize cases where the cultural fitness of a meme is sufficiently detached from its advertised ethical function to indicate that the origins of this meme are arbitrary and that we ought to suppress its normative authority? The problem is complex because the ethical memes we collectively endorse are often supported, at least in part, by memetic forces that cannot withstand critical scrutiny of this kind. If we were to clear away every seemingly arbitrary cultural influence acting on our moral intuitions, we might not be left with very much! The question is how eccentric the genealogy of a moral meme has to be in order to demonstrate that its prescriptive influence ought to be limited in our cultural environment.

This is, in fact, a very difficult question. (Difficult enough that we will return to it in more detail in chapters five and six.) At one end of the spectrum we have presumably flawless memes such as "equality", "fairness", and "justice". These memes may have questionable cultural origins if we push their genealogies back far enough, but they are so fundamentally intuitive and essential to social organization that they seem irreplaceable. Just as we would not want to rid ourselves of some of our unique physical attributes if we discovered that their original biological functions were no longer strictly adaptive in a contemporary setting, it is unlikely we would want to dispose of our memetic dispositions
for justice and fairness once we uncover the genealogy of their cultural origins. Yet we cannot become complacent about the cultural origins of our moral intuitions, because at the other end of the spectrum there are ethical memes we would almost certainly want to discard if we realized the anomalous character of their genealogical histories.

Consider the following familiar example: the meme 'what is natural is necessarily good'. This meme somehow manages to flourish in contemporary moral discussions, yet its genealogy would almost certainly reveal that its origins and continued existence in our social environment involve matters of prudential reasoning that are unrelated to the ethical situations to which it is applied. The result is that convictions based on this meme (e.g. that homosexuality is wrong, or that withholding readily available medical care is acceptable while 'actively' terminating a patient's life is not) become uncritically grounded on a moral intuition that survives in our cultural infosystem because of selection forces not directly related to the justificatory role this meme is thought to possess. This is the sense in which the normative authority of the meme becomes disengaged from the underlying cultural influences that have given it such a strong cultural fitness. The "natural is good" meme is a useful example because there is not much of a mystery concerning the underlying cultural influences in question: our experiments tampering with the natural world often end in disaster, giving us good reason to rely on the general principle that what is natural is usually a safer option.22 Problems arise, however, when this general principle is internalized and then applied to new circumstances without retaining a firm connection to the original basis for its cultural selection. There may be excellent prudential reasons for valuing what is natural in many cases, but even a cursory look at the genealogy of a meme stating unequivocally that

22 While the fitness the 'natural is good' meme currently possesses is probably related to ecological and personal health concerns, there may be other social factors contributing to the meme's popularity that are much older and deeply entrenched in our cultural systems. These factors (that are still equally immaterial to questions of homosexuality and end of life decision making) could include tendencies like an inherent conservatism towards deviations from existing norms - an advantageous trait for social learning because it enables us to more effectively emulate behaviour that is adaptive in hazardous environments.
what is natural is necessarily good shows that this meme in no way deserves our uncritical faith in complex moral situations.

How many other examples are there of memes functioning within ethics whose cultural fitnesses are sufficiently disengaged from their advertised moral functions for us to want to limit the scope of their authority? It is here that the level of precision promised by memetics will finally become an asset if it enables us to identify the otherwise unperceived existence of arbitrary memes in our ethical thinking and the cultural selection pressures that allow these memes to continue to propagate. The 'natural is good' meme is an easy target, but other distorted moral intuitions may be more difficult to pick out as evolutionary products of corrupt selection forces. Thus, if memetics succeeds at cataloging the mechanisms contributing to cultural fitness, we might discover that these selection forces help disseminate a great many ethical concepts according to subtle evolutionary rules rather than our discerning intentions. The unprecedented level of accuracy supplied by memetics can therefore be applied to a problem that requires this much accuracy about the cultural forces acting on our ideas in order to be effectively resolved.

How valuable the contribution memetics will provide ethics is thus an empirical question which depends on the level of precision memetics achieves and just how eccentric the selection pressures associated with cultural evolution turn out to be. Again, we find that the explanatory independence of cultural evolution is a key factor in determining whether mixing evolutionary theory and ethics will be worthwhile. If sociobiology is correct in postulating that cultural evolution is tightly bound to biological evolution via epigenetic rules, there will be little room for cultural selection pressures to carve out rules of cultural fitness that deviate substantially from the underlying rules of genetic fitness. The result would be that the increased precision of memetics would be wasted identifying arbitrary selection forces that can already be picked out by sociobiological predictions. However, if cultural evolution is capable of breaking free
from biology and acting as an independent (although intricately connected) system of inheritance, then the potential for memetics to tell us something interesting about the cultural origins of our moral intuitions becomes far more exciting. In this case, memetics can potentially unlock the curious evolutionary logic of our cultural environments and provide us with an incredible new source of empirical information about the potentially defective cultural forces supporting our moral ideas. In my opinion, it is this possibility, combined with the possibility of memetics emerging as a fully operational science, that presents the most intriguing prospect for the union of ethics and evolutionary theory.

**Conclusion**

After sketching the various ways in which theories of cultural evolution might affect the study of ethics, chapter three has endeavored to demonstrate that, interestingly enough, the least revolutionary proposal suggested holds the most promise in terms of providing a worthwhile contribution to the study of ethics. The chapter has been premised on a 'best-case scenario' approach in the sense that memetics - the theory used here to represent the promise of new theories of cultural evolution - is still in its early stages of development and requires us to speculate on the details it will uncover once it establishes itself as an operational science. Using this approach, three initial options for memetics and ethics were discussed: the prospect of using memetics to engineer more effective ethical memes, the possibility that memetics will lead to a form of determinism which renders ethics obsolete, and the prospect of memetics leading to general moral skepticism. All of these proposals were shown to be highly exaggerated initial reactions to the combination of ethics and memetics, and each one was shown to have a negligible impact with respect to questions of prescriptive justification. Next, the possibility of using memetics to elucidate the standards of a naturalistic form of moral realism was discussed. Although this option is more viable, in the end it too was shown to contribute very little to our ability to prescriptively justify our moral claims.
After examining these potential combinations of memetics and ethics, the chapter closed by suggesting that memetics will make a substantial contribution to the study of ethics by implementing the empirical information provided by memetics to improve our existing efforts to systematize our moral intuitions. This contribution is an exceptionally precise account of the genealogy of our moral intuitions, which can be used to identify ethical memes that survive because of selection forces that have drifted beyond the perceived causes for their cultural fitness. If we trace the cultural evolutionary history of our moral memes and find that they continue to propagate because of properties only vaguely related to those we normally acknowledge as the reasons for their continued existence, we will have good reason to be more restrictive about the normative authority we assign to these memes in contemporary ethical debate. If we can regulate the dynamic mechanisms of cultural evolution in this manner, I believe that memetics could have a profound effect on our ability to consistently justify the authority of our moral intuitions. It is this option that we will examine in further detail in chapter four.
Chapter 4

The Genealogy of Moral Memes

Introduction

In the final lines of *The Selfish Gene*, Richard Dawkins passionately urges his audience to "rebel against the tyranny of the selfish replicators" (Dawkins 1976, p. 201). This call to arms carries a considerable sense of urgency, as Dawkins has just finished providing a convincing argument for the claim that human behaviour is profoundly influenced by the peculiarities of genetic replication. Moreover, Dawkins has also suggested that a parallel type of replicator exists in our cultural environment. If we accept the fact that our intuitions are deeply affected by the eccentric logic of both genetic and cultural replicators, it seems reasonable to follow Dawkins' advice and seek to liberate ourselves from the apparently insidious power of these replicating entities.

However, it is not at all clear what kind of "tyranny" Dawkins is referring to, nor what kind of rebellion he has in mind. Dawkins wants us to believe that replicators (be they biological or cultural) cannot be trusted to secure our well-being, and that we ought to fight to regain control over our destinies instead of allowing blind evolutionary forces to direct our actions. But what exactly will this rebellion involve? Like most of Dawkins' writing, the prose is eloquent and inspiring, and his final rallying cry evokes a powerful response by encouraging the reader to take possession of his or her own life. But because the replicators in question form an essential part of the lives of which we are to take control, it is not at all clear just what Dawkins means when he recommends that we rebel against the authority of all replicating units.

In this chapter, I defend a means of systematizing our moral intuitions that I take to be the most coherent interpretation of Dawkins' suggestion that we rebel against the tyranny imposed on us by selfish replicators. The idea of rebelling against something that forms an integral part of oneself may seem problematic, but if Dawkins' recommendation
is stripped of its exaggerated rhetorical overtones I believe it can be realized in a meaningful and important way. The implementation I have in mind is a continuation of the methodology developed in chapter one and the end of chapter three: using the information provided by evolutionary theory (interpreted as a general theory of both biological and cultural evolution) as a valuable genealogical resource for the aim of systematizing the authority that we attribute to our moral intuitions.

In chapter one this methodology involved examining our genetic social dispositions in order to determine whether these dispositions ought to be resisted or promoted. In chapter three, the focus turned towards cultural evolution and the consequences memetics might hold for ethics. In chapter four, the proposal introduced at the end of chapter three will be discussed in greater detail. This proposal involves using the empirical evidence supplied by memetics to construct a supplementary basis for evaluating the prescriptive authority that our moral intuitions deserve. By investigating the underlying cultural selection pressures acting on our thoughts and moral institutions, I believe we can develop a new source of guidance for the esteem we ascribe to (or withhold from) our ethical intuitions. In doing so, we can remove some of the arbitrary characteristics present in the landscape of our moral thinking, and we can improve our ability to consistently justify our prescriptive ethical judgments.

The chapter is structured in four parts. Part one addresses the apparent paradox of rebelling against an essential part of ourselves when we take up a critical standpoint towards the memes that constitute our moral intuitions. It may seem dubious that this critical standpoint can possess the privileged status required to critically assess our intuitions, since it is itself merely composed of some other assemblage of memes. I argue, however, that this circularity is not vicious, and that we can make as much sense of a memetic critical standpoint as the non-memetic critical standpoint we regularly appeal to when we evaluate the prescriptive merits of our common-sense intuitions. After
settling this meta-ethical issue, part two moves on to the primary task of the chapter and outlines several different ways in which an understanding of memetics could change our views about the moral intuitions we ought to prescriptively endorse. Part three pursues one of these options in more detail and explains the basis for what I consider to be the most promising area for cultural evolution's ability to have an impact on the authority we attribute to our intuitions: identifying the propensity for cultural selection to favour indirect indicators of fitness and the tendency for these indicators to become favoured beyond the point where they ought to be considered reliable. Finally, part four examines the prospects for this application of evolutionary theory in an explicitly ethical context and provides examples of how the genealogy of moral memes will affect our ethical judgments in tough cases where our intuitions conflict with the recommendations of ethical theory.

I. Memetics and the Archimedean Point

There is something suspicious about speaking as if memes exist as separate entities that can be observed from an external standpoint. It is certainly tempting to describe memes as foreign objects that somehow manage to infect our otherwise independent minds. But if the implications of memetics are considered carefully it seems clear that our minds cannot provide us with an outlook that transcends the realm of memes, since our minds are themselves composed of memes and are therefore inescapably constrained by the forces driving cultural replication and transmission. In fact, when pushed to its natural conclusion, memetics is as much a theory of consciousness as it is a theory of cultural change. As Daniel Dennett (and more recently Susan Blackmore) has pointed out, memetics implies that consciousness simply is the collection of memes that occupy our brains.1 What we call the mind results because human brains form unique 'nests' that are

1 This is most clear in (Dennett 1991) chapter seven, especially p. 210. See also (Blackmore 1999) chapter seventeen.
especially conducive to being occupied by cultural replicators. This supposition, that our minds are nothing more than convenient hosts for replicating memes, may not be an especially appealing suggestion, but it is the most consistent interpretation of memetics and one which graphically demonstrates the contradiction involved in attempting to remove oneself from the substrate of memes in order to judge them from an impartial standpoint.

Thus there is something misleading, if not incoherent, about the prospect of rebelling against what Dawkins calls "the selfish memes of our indoctrination" (Dawkins 1976, p. 200) if this rebellion presupposes a completely independent point of view. After all, *we are our memes!* If memetics is taken seriously, our identity and most cherished values become exactly the kind of selfish replicators that Dawkins recommends we oppose.² The goal of resisting the tyranny of all selfish replicators is an impossible task because everything we are, right down to the standards we wish to promote for our own benefit, must be composed of the selfish memes from which we are told to liberate ourselves.

This paradoxical catch for the task of rebelling against our own memes seems to pose a problem for the methodology outlined in chapter three. There, it was suggested that we single out memes that propagate because of unusual cultural selection pressures and then critically reevaluate these memes in order to remove arbitrary characteristics from our moral thinking. However, critical reevaluation may be just as problematic as outright rebellion if it presupposes that we can achieve a completely meme-independent point of view. How can we cast suspicion on the authority of some particular set of memes if we can only do so by appealing to more memes and their own potentially suspicious cultural histories? How can we single out certain memes as being arbitrary when the concept 'arbitrary' is itself only a meme that is especially adept at filtering out

² In the words of Walt Kelley's famous *Pogo* aphorism, "We have met the enemy, and he is us."
its cultural competitors? It looks as if the circularity attributed to Dawkins' call for insurrection presents itself all over again: if we accept the fact that our ideas are composed of memes then consistency suggests that all our ideas are ultimately memes, and this leaves us with no room to step back from the forces of cultural evolution to evaluate their trustworthiness from a privileged standpoint. In short, if human cognition is a memetic construct then it is a construct all the way through, leaving us with no impartial territory to occupy in order to launch a fully external critique of the evolutionary forces acting on our thoughts.

An Internal Critical Standpoint

It may seem as if the preceding argument creates an insoluble problem for the task of critically reassessing the memes that constitute our moral intuitions. However, it is important to separate the fact that we cannot escape the confines of a memetic substrate from any further conclusions we may draw about how this fact affects questions of ethical justification and our ability to critically evaluate our thoughts. In particular, it is entirely possible to accept the fact that an independent, non-memetic perspective is beyond our reach without abandoning the project of critically reevaluating our moral intuitions. The reason this remains possible is that although all our ideas must ultimately be composed of memes, this fact does not necessarily entail that all memes must possess the same authoritative standing. We may not have access to an external standard by which to judge the authority of memes and the cultural forces that promote them, but we can nonetheless appeal to compelling internal standards in order to construct a basis for accepting the trustworthiness of some of our intuitions over others from inside a memedependent point of view.

Hence, the mistake is to believe that because we must ultimately work within a metaphysics of memes we must also be committed to judging all memes as equally valuable. Not only does this conclusion presuppose its own dubious impartial standpoint,
but as an objection it succeeds only in undermining the methodology of casting doubt on our intuitions just because we have identified them as being 'merely' memes. Such a methodology would indeed involve a contradiction, for it would imply that intuitions not called into doubt (and the thoughts responsible for doubting itself) are composed of something other than memes. But clearly this is not what is being proposed.\(^3\) What is being proposed is that we use the information provided by an evolutionary perspective to more effectively identify intuitions that seem suspicious from within the perspective of what is admittedly a meme-constructed point of view. Particular intuitions are therefore scrutinized not simply because they are discovered to be memes, but because they are discovered to be memes with genealogies that are suspect when they are viewed from the perspective of (the memes that constitute) our capacity for critical thinking. Rather than attempting to adopt an unattainable independent standpoint to critically access our memes, the aim is to use the resources available to us to better systematize the authority of these memes from inside the network they construct for us.

*Rawls and Reflective Equilibrium*

Those familiar with contemporary issues in ethics will recognize that the point at stake here is analogous to the debate concerning the type of privileged status (if any) that theoretical principles ought to exert over our more particular moral intuitions. Critics of the view that theoretical judgments ought to possess an authoritative standing over our ordinary common-sense intuitions often point out the impossibility of taking up a completely impartial theoretical standpoint towards our ethical intuitions, i.e. a standpoint removed from all pre-theoretical human moral intuitions.\(^4\) Critics refer to this illusory standpoint as an Archimedian point. Archimedes claimed that if he was given a single, immovable point outside the Earth to act as a fulcrum, then given a long enough lever he

\(^3\) In fact, this is precisely what was rejected in last part of section one in chapter three.

\(^4\) See, for example, (Williams 1985, chap. 6).
would be able to move the entire world. Similarly, critics of moral theory often argue that the primacy of theory over our intuitions requires a completely-secure starting point on which to base the lever of impartial ethical principles, and that the prospect of achieving a point of view so abstracted from absolutely all human biases (if such a point of view is even coherent) is virtually impossible. Just as there is something dubious about attempting to evaluate memes from a wholly external standpoint, there seems to be something problematic about building up ethical theory from a completely impartial foundation.

Ethical theorists have attempted to resolve the problem of requiring an Archimedian point in a variety of ways, but the most common antidote to this dilemma is the move to a fluid system of critical evaluation which begins from provisional starting points - a process John Rawls has termed "reflective equilibrium". (Rawls 1971, p. 20) As Rawls describes it, reflective equilibrium involves building an initial theoretical construct from widely shared assumptions and then continually reevaluating whether this construct can accommodate our considered moral judgments. In this respect, reflective equilibrium follows a traditional approach to ethical theory by seeking to develop an initial theoretical standpoint against which particular intuitions can be tested. However, Rawls introduces an important feature that distinguishes his view from conventional interpretations of the relationship between theory and judgment. This feature is that the form of critical reflection involved in reflective equilibrium allows either our ordinary intuitions or our theoretical principles to be revised in the event that the two conflict. Thus, if the theoretical standpoint we construct from our shared assumptions conflicts with the recommendations of our particular judgments, Rawls suggests that in some cases

---

5 This argument is, of course, not limited to the domain of ethics. Critics of the priority that foundationalist outlooks place on a detached theoretical perspective have raised the issue in all areas of philosophy, most notably in epistemology, the philosophy of science, and the philosophy of mind. See, for example, (Quine 1969), (Rorty 1979) or (Taylor 1995).
our particular judgments must be thrown out while in other cases the principles derived from our theoretical standpoint must be revised.

The idea that ethical justification ought to work from both ends of the division between theory and intuition may seem self-evident given the way we actually proceed with deliberating over the merits of ethical theories (e.g. using the intuitions generated from particular examples and counter-examples). But Rawls' articulation of reflective equilibrium is nonetheless significant because it represents a unique instance where the task of constructing rigorous ethical theory is taken extremely seriously without this theory requiring an independent Archimedean starting point. The most interesting aspect of Rawls' account of reflective equilibrium is therefore the fact that the assumptions used to begin the process of theory building - convictions in which we are said to have "the greatest confidence" - are described as "provisional fixed points" and are acknowledged to be potentially revisable. (p. 20) These convictions carry the authority to cause us to abandon our particular judgments if need be, yet they are not put forth as "necessary truths or derivable from such truths" (p. 21). They are instead presented as judgments derived from circumstances favourable enough that they can be considered authoritative from within an admittedly fallible point of view.

**Memes and Reflective Equilibrium**

Rawls' account of reflective equilibrium provides a useful model to appeal to in order to demonstrate how the task of critical reflection can operate within a memetic perspective. Rather than presupposing the existence of a completely meme-independent standpoint, critical reflection can begin with a set of memes in which we have the utmost confidence but we nonetheless take as provisional fixed points within the scope of what is metaphysically a meme-dependent point of view. Using these provisional memes as a starting point, we can develop a critical perspective to evaluate other memes in which we have less confidence in order to determine whether they ought to be revised or perhaps
discarded. The provisional nature of the framework allows for the possibility that the memes grounding our theoretical principles can be revised if necessary. Yet the fact that the framework provides the capacity to initially commit ourselves to the greater confidence we attribute to such fixed point memes (or meme clusters) as rationality, consistency, and fairness ensures that we can set the machinery of critical reflection in motion from inside the memetic perspective we inevitably inhabit. Thus, we may not have access to an Archimedean point that will allow us to evaluate our memes from somewhere outside of ourselves, but we have access to enough of an internal Archimedean standpoint to allow critical reflection to proceed in a nonetheless meaningful way.

Of course, the prospect of a memetic version of reflective equilibrium is not without difficulties. Sorting out exactly how the ontology of an internally coherent memetic justificatory system is structured leads us very quickly into some of the most difficult and controversial areas in contemporary philosophy. Depending on how one conceives of the nature of objectivity and the intricate epistemological problems surrounding questions of truth and justification, there are many different ways to interpret how the process of reflective equilibrium would (or would not) function in a memetic context. Moreover, Rawls himself is rather elusive when it comes to details concerning how this process operates in an ordinary ethical context (especially in terms of how the process determines when we ought to revise our principles versus our particular judgments). Thus, depending on one's underlying philosophical commitments and the way in which one interprets Rawls' project, the appeal to reflective equilibrium as a means of generating a critical standpoint within a memetic context will carry varying degrees of plausibility.

But the issue at hand is not whether a memetic implementation of Rawls' theory constitutes an infallibly adequate means of critical reflection. Much more would need to be said to demonstrate this claim, and I am not offering anything like a thorough defense
of such an implementation. What I hope to have made clear is that a memetic implementation of reflective equilibrium is at least a viable option, and that the move to an interpretation of the process that uses the terminology of memes rather than judgments will at least not introduce any additional problems for the difficult task of sorting out contemporary issues of truth and justification. Hence, one may not be fully convinced that critical reflection can operate within a memetic perspective without an Archimedian standpoint, but I submit that if this is so then the target of one's objections has little to do with memetics and much more to do with the controversial nature of objectivity and our expectations about meaningful ethical justification. While these latter concepts are important, they are sufficiently controversial that one can hardly hold memetics responsible for their not being resolved. Thus, the question of how we achieve enough of an independent perspective to critically assess our intuitions remains a long-standing philosophical puzzle, but it is a puzzle that we are at least no worse off addressing from the language of memes rather than 'considered moral judgments'.

II. The Value of Genealogy

If there is no significant difference between memetic and non-memetic attempts to solve the difficulties surrounding critical reflection and our ability to achieve an authoritative justificatory standpoint, what exactly do we gain by switching to a language of memetics? The move towards the terminology of memes rather than judgments may not leave us any worse off when it comes to sorting out the meta-ethical issues associated with building an internally coherent ethical system, but nor does it seem to leave us any better off in this respect. Can a memetic perspective therefore provide us with potential advantages that would make the transition to the terminology of memes worthwhile?

As I have suggested in chapter three, I believe that a memetic outlook will indeed provide us with an advantage in virtue of the fact that it will supply us with a novel source of empirical information about the evolutionary history of our moral intuitions.
By familiarizing ourselves with this information, we can use a memetic outlook to evaluate the prescriptive authority we assign to our moral intuitions in a new and interesting way. This advantage over a non-memetic perspective may not constitute a revolutionary breakthrough for *meta*-ethics, but the presumption that anything less than such a breakthrough must be uninteresting demands far too much. Rather than attempting to change the entire groundwork on which ethics is based (and virtually guaranteeing that memetics will come up short), introducing an empirical resource with intriguing consequences for the way we regard our moral intuitions and their justification is, I think, a far more interesting and worthwhile project.

What exactly will this project entail? I have suggested that information about the evolutionary genealogy of our moral intuitions will allow us to more effectively systematize the authority we attribute to the intuitions we consider to be worthy of prescriptive justification. However, this proposal requires further elaboration. In the remainder of this chapter I will therefore describe in more detail how the genealogy of moral memes can supply us with an additional method of ascertaining the trustworthiness of our intuitions. Moreover, in doing so I hope to demonstrate that paying attention to the evolutionary forces acting on our intuitions will allow us to learn something interesting about the ways in which our ethical judgments are susceptible to distortion. In this section of the chapter, three possible applications of this empirical information provided by memetics will be broadly outlined. By sketching the range of these possibilities the aim is to situate what is being proposed in a clear conceptual framework. The remaining sections of the chapter explore the most profitable of these three options in further detail. Using examples from cultural transmission models and hard cases in ethics, sections three and four present an extended argument for memetics' capacity to influence the authority of our moral intuitions in a unique and instructive manner.

---

6 For an example of exactly this kind of 'all or nothing' approach to memetics, see (Kingwell 1999).
Potential Applications: Runners-Up

The three potential applications of memetic genealogical information I have in mind differ according to how startling the results they produce appear, considering the ethical contexts to which they are applied. Each one of these possibilities has something interesting to contribute to our understanding of ethics, but the magnitude of this contribution depends on how eccentric the cultural selection pressures they reveal turn out to be. (This relationship is not necessarily linear though, as we shall see below.) By considering the ways in which an increased awareness of the memetic forces acting on our thoughts will expose factors that are either more or less strange and unforeseen, I think we can acquire a better sense of how dramatic an effect memetics will have on the confidence we place in our ethical judgments.

The first application of memetics as an empirical resource involves cases where an exploration of the evolutionary forces acting on our moral intuitions does not uncover cultural mechanisms eccentric enough to cause us to modify the esteem we have for these intuitions. In such cases we might discover, for instance, that our intuitions have been shaped by forces that weed out exactly the kinds of memes we expect to be eliminated by a well-functioning ethical system: those that appear irrational, self-contradictory, imprudent, or just plain abhorrent. (E.g. memetics reveals only that cultural evolution selects against intuitions like 'murder is fun' or 'theft is acceptable', etc.) We might, of course, find evidence of cultural forces that allow certain undesirable memes to pass through these filters because of imperfections in the system. (E.g. 'it's OK to drink and drive' or 'stealing office supplies is permissible'.) And we might find that memetics provides indefinite results if applied to controversial topics. (E.g. selection favours both the 'abortion is permissible' meme and the 'abortion is impermissible' meme in a stable equilibrium.) But we would nonetheless find nothing surprising about the exceptions revealed by a memetic investigation of how our moral intuitions have evolved. 'In effect.
this possibility represents the hypothetical scenario where the empirical information we uncover fails to supply us with any new evidence regarding the potentially dysfunctional nature of our moral thinking.

This scenario may seem like an odd possibility to cite as a case where memetics will have an impact on ethics. In fact, chapters two and three both explicitly argued for the claim that the implications of memetics only become interesting when we examine the exceptions to our ordinary expectations about what an evolutionary account of culture might produce. How will an account that produces no such exceptions be of any interest to ethics? The answer is that although these cases are certainly not as interesting as those producing unexpected results, it is easy to ignore the fact that there would still be an important explanatory benefit associated with understanding the evolutionary forces responsible for shaping our intuitions in cases where this understanding confirms our existing judgments. Thus, even in situations where the information supplied by memetics will not lead us to rethink those intuitions we consider to be prescriptively justified, this information may still teach us a great deal about the underlying forces that cause us to draw these justificatory conclusions. (Moreover, these results would help confirm the integrity of our moral judgments by providing additional reasons for believing that our intuitions constitute part of an orderly, internally coherent system.) The extra benefits provided by this evidence may not be immense, but these explanatory gains are nonetheless valuable enough to be worth noting since they are too often ignored completely in the rush to find novel applications for memetics' seemingly revolutionary potential.

The second application of memetics as an empirical resource involves instances of the opposite extreme: examples where the information provided by memetics uncovers a shocking degree of unperceived or dramatically underestimated cultural selection pressures acting on our moral intuitions. In these cases, an investigation into the evolutionary genealogy of our intuitions reveals that some of these intuitions possess
gratuitously arbitrary cultural histories despite the fact that they continue to flourish in our minds and social institutions. By 'arbitrary', I mean that the memes in question have been promoted by cultural evolution because of selection pressures that would appear plainly capricious if exposed and examined from a critical standpoint. Thus, memes with arbitrary memetic genealogies are those that survive in our cultural environment for reasons that seem completely unrelated to their being rational, helpful, prudent, aesthetically pleasing or generally beneficial in any significant respect. The reasons why these memes continue to exist in our cultural environment may be complex and rather difficult to grasp without an understanding of the peculiar evolutionary forces responsible for their propagation. It may be the case that our minds have been constructed in such a way that we are especially susceptible to certain cultural forces that are otherwise incongruous with the selection pressures we expect from a well-functioning ethical system. A memetic explanation of these details could therefore potentially unravel the tangled cultural genealogy behind these intuitions and explain how anomalous ethical memes manage to survive within a system that ought to allow its subjects to perceive them as unfavourable.

This application of memetic empirical information seems as if it should be the most promising option possible in terms of memetics being able to have a productive impact on ethics. What could be more helpful than identifying (and enabling us to rid ourselves of) the most extreme and unexpectedly arbitrary elements present in our moral thinking? Using the evidence given by memetics for this purpose - to weed out the most glaring examples of cultural evolution's peculiar, non-optimizing structure - seems like an extremely valuable project and precisely in the spirit of what was proposed at the end of chapter three. Indeed, I think this option has considerable promise. However, there is a problem standing in the way of this application making a substantial contribution to our existing understanding of ethics.
The problem is that although it may be appealing to spot uncommonly arbitrary moral memes in our cultural environment, it is unlikely that we would need to rely on the empirical information provided by memetics to accomplish this goal if the genealogies of these intuitions are so obviously incongruous with those of the rest of our cultural system. While it may be a worthwhile project to try to identify and eliminate arbitrary patterns of thought from our intuitions, in cases where the genealogies of the memes in question are dramatically arbitrary we would likely not need memetics to successfully meet this objective. Consider the following examples: blind faith, male-chauvinism, scapegoating, and the thirst for vengeance. If we examined the genealogy of each of these memes we would likely find contingent factors in their cultural evolutionary histories that would make them seem unsuitably arbitrary from the perspective of a carefully considered critical standpoint. A memetic approach would therefore allow us to single out these memes and explain why the peculiar circumstances of their cultural genealogies have allowed them to flourish. However, it also seems clear that we could just as easily identify these concepts as being unsuitably arbitrary without the help of memetics.

The prospect of using memetics to rid ourselves of exceedingly arbitrary memes is therefore either superfluous or a highly unlikely possibility. It would be useful for memetics to help us recognize that intuitions persist in our heads because of dysfunctional selection forces, but the likelihood that memetics will allow us to pick out cases where the dysfunctional nature of these intuitions is extreme but unforeseen is rather slim, because we will usually be able to recognize concepts with glaringly arbitrary genealogical histories without having to study them as memes. Again, there may still be an important explanatory benefit associated with understanding the longevity of these intuitions from a memetic perspective, but if this is the only benefit derived from these cases then they reduce to a variant of those discussed above - instances where memetics helps to confirm what we already suspect about our intuitions. It is certainly possible that memetics could uncover completely unexpected results about the arbitrary genealogies of
some of our intuitions. There may be memes existing in our cultural environment that seem entirely ordinary but are actually promoted by evolutionary forces that would appear completely bizarre if brought to our attention. The problem is that, virtually by definition, we cannot predict what these memes look like or how many there might be in existence. As a result, the benefits of this application of memetics' empirical powers remain somewhat mysterious and diminished by the fact that moral intuitions with extremely arbitrary genealogies can usually be identified without the aid of memetic explanations.

*Potential Applications: The Winner*

The last application of the empirical information provided by memetics to be examined in this section falls in between the two possibilities discussed so far. The suggestion in cases of this third type is that genealogical information about the transmission of *ordinarily praiseworthy* memes can reveal particular instances where these memes are promoted by evolutionary mechanisms that have become disengaged somehow from the selection pressures we normally associate with their continued existence and credibility in our cultural environment. In situations of this type, an ordinarily acceptable meme (or at least one not glaringly arbitrary) can continue to be favourably selected in circumstances where the qualities that make the meme a trustworthy cultural replicator no longer apply. The evolutionary history of these memes might be compatible with the selection forces that a critical standpoint would approve of (e.g. selection forces favouring rationality, consistency, beneficence, etc.) when taken on the whole, but in particular circumstances our intuitions may propagate because of unperceived cultural side-effects that produce selection forces we would not approve of - forces that are unsuitably arbitrary (and perhaps even hazardous) despite the fact that the intuitions in question retain the same sense of normative authority they have acquired from situations where their genealogy is not problematic.
In these hybrid cases, the information supplied by memetics is not being used for a purely explanatory purpose, nor is it being used for the potentially superfluous task of identifying intuitions with exceptionally (and in most cases recognizably) arbitrary genealogies. Instead, this information is being used to fine tune the approval we attach to normally meritorious intuitions in situations where unique selection pressures arise that promote these intuitions beyond the point where we have good reason to trust their authority. It is within this middle range between the two previous applications of memetics' empirical results that I think an awareness of the evolutionary genealogy of our moral intuitions will be most profitable.

What exactly will instances of this third application look like? What kind of selection pressures exist that can cause trustworthy intuitions to drift away from the reasons we have esteem for them? The next two sections of the chapter attempt to answer these questions in further detail. The conclusion drawn is that an awareness of the genealogy of memes in these cases provides us with a particularly interesting source of evidence to appeal to in order to determine the trustworthiness of our moral intuitions. This evidence may not be decisive in and of itself, but it may be very useful in cases where our intuitions come into conflict with one another and we are forced to find a way to choose between them. In such situations, an investigation of the peculiar structure driving cases where cultural evolution favours estimable intuitions beyond the point where we ought to respect their authority can tell us something important about how to resolve conflicts between our moral intuitions in a way that we would not be able to discover without the help of memetics' precise empirical capacities.

III. Cultural Runaway Effects

The hypothesis that information about the genealogy of our moral intuitions will enable us to more effectively evaluate the authority these intuitions ought to possess depends critically on the idea that cultural evolution operates according to its own unusual
evolutionary logic. If cultural evolution always produced the results we expect from a system operating according to reason, consistency and our general best interests, then there would be little or no benefit (except an explanatory one) associated with investigating the genealogy of memes for the purpose of improving ethics. It is only when memetics uncovers examples where cultural evolution fails to promote the memes we expect to be favoured by a well-functioning system that we have reason to believe that the confidence we place in our ethical judgments may be undeserved. As we have seen in chapter two, however, the structure of cultural evolution gives us good reason to think that this is in fact the case. Because cultural evolution promotes memes that happen to be good at replicating themselves (by whatever means necessary), situations arise where cultural evolution generates surprising results by favouring efficient replicators that are unfortunately not rational, prudent or beneficial.

*Indirect Indicators*

If we investigate the structure of cultural evolution we find that it arrives at these unexpected outcomes in certain recognizable ways. The most notable of these (discovered so far) involve cases where ordinarily acceptable cultural patterns are exaggerated to the point where they become disengaged from, and perhaps even detrimental to, our best interests. As we have seen, we will likely already be able to identify cases where cultural patterns are dramatically out of character with our best interests without the help of memetics. The cases we are most interested in are cases where the details of cultural transmission allow us to recognize mechanisms that furtively promote ordinarily benign memes beyond the point where we ought to accept the authority of their intuitive power. The most compelling of these mechanisms involves situations where a particular behaviour or property is taken up as an *indirect indicator* of cultural fitness by agents seeking to acquire the most adaptive cultural variant in their environment while at the same time attempting to minimize the risks associated with
exposing themselves to the perils of individual trial and error learning. In these situations, individuals begin to rely on cultural transmission as a means of avoiding the costs of deciphering the hazards in their environment on their own, but they also attempt to economize on the costs of testing the options made available to them via cultural transmission. In a sense, then, these individuals attempt to cut corners on what is already a savings in the risks associated with learning the most adaptive behaviour in a particular setting. To do so, they use indirect indicators of success to more efficiently determine the fitness of the variations presented to them by their (cultural) parents and peers.\textsuperscript{7}

This process gives individuals a more immediate route to the patterns that appear to be fitness maximizing. However, it also leaves these individuals susceptible to the problems associated with relying on second-hand information. If the indicators chosen to represent fitness begin to stray from what is actually adaptive in their local environment, individuals can find themselves at the mercy of what Robert Boyd and Peter Richerson have aptly labeled a cultural "runaway process". This is a situation where the perceived advantage of imitating an indicator trait creates its own escalating pressure towards the selection of more and more dramatic indicators, despite the fact that the trait in question has long since become a poor indicator of fitness and perhaps even an option that is thoroughly maladaptive.\textsuperscript{8}

If there is a significant risk that indicator traits can drift away from the fitness enhancing properties they are supposed to represent and then potentially lead to maladaptive outcomes, how is it possible that this mechanism continues to exist without natural selection strongly working against it? The reason indirect transmission mechanisms continue to play an active role within cultural evolution is that despite the

\textsuperscript{7}One may remember that this is the same cultural mechanism that we saw in chapter two, where it was used as an example of the fact that cultural transmission produces outcomes that are explanatorily independent from the predictions of sociobiology.

\textsuperscript{8} See (Boyd and Richerson 1985, p. 259-60). The term 'runaway process' is taken from Fisher's original work on the biological equivalent of this phenomenon - a process that explains anomalous results in population genetics. I will return to the parallel between these two processes shortly.
potential for maladaptive outcomes, the benefits associated with economizing on learning costs are simply immense. In the same way that the initial jump from individuals learning about their environment using trial and error to their acquiring information via cultural means is prone to mistakes but massively beneficial when taken on the whole, the jump from testing various cultural options that have been presented to trusting one observable indicator as a mark of fitness may be disastrous in certain cases but generally constitute a huge savings in time and risk. Thus, relying on cultural transmission to narrow down the appropriate behavioural options displayed by one's peers and then appealing to an indirect means of selecting the most profitable of these options is hardly foolproof, but it is an advantageous enough strategy to remain an enduring mechanism within cultural evolution.

*Dramatically Suboptimal Outcomes*

How exactly do situations arise where the appeal to indirect indicators of fitness leads to maladaptive outcomes? As described above, this process begins with naive individuals actively seeking out the most adaptive cultural variants in their local environment. To economize on information costs these individuals fixate on characteristics that are assumed to be reliable indicators of the fitness exhibited by more experienced individuals in the environment. Yet, these chosen characteristics may only be coincidentally related to the fact that the cultural models exhibiting them are reproductively or culturally successful. Naive agents can look to such characteristics as social status, hunting techniques, methods of grasping tools or even styles of dress in order to emulate the behaviour of the most successful agents in their environment. Some of these characteristics might genuinely be indicative of what is necessary to achieve the level of fitness displayed by the cultural models possessing them, but there is no guarantee that this will be the case. Once the indicator is established in a social group as
a desirable trait to acquire, however, it often becomes admired and imitated regardless of whether or not it is actually related to the success it supposedly represents.

Thus, the most important aspect of cases where indirect indicators of fitness are appealed to in a population is that once a substantial number of individuals conclude that these characteristics are reliable indicators of fitness, selection pressures begin to act on the indicators themselves so that the drive towards acquiring the traits is reinforced (and perhaps even amplified) whether they are genuinely indicative of the fitness of those possessing them or not. In effect, once the mechanism is in place to allow individuals the opportunity to commit themselves to indirect cultural shortcuts, the process takes on a life of its own as agents adopt indicator traits in more extreme and elaborate ways without stopping to question the utility of these characteristics in the changing circumstances of a fluid cultural context.

The process is analogous to the mechanisms proposed to explain the emergence of exaggerated male traits in cases of biological sex selection. As R. A. Fisher's work on the subject has helped to demonstrate, situations can arise where females tend to prefer males with slightly disadvantageous behavioural or morphological traits (such as the excessively visible tails of male peacocks) to such an extent that the adaptive benefits associated with sexually desirable males can begin to outweigh the ordinary selective pressures acting against males possessing this maladaptive trait.9 The result is a runaway effect where the reproductive fitness of males with the disadvantageous characteristic exceeds the fitness of other males (with whom they are in mating competition) despite the fact that each member of the population would obviously be more likely to survive if they did not carry this sexy but maladaptive attribute.

For example, in the case of peacocks, males with smaller, less visible tails will be more likely to escape predators. However, because these peacocks will not be selected as

---

9 See (Fisher, R. A. 1930).
suitable mates by females preferring larger tails, their overall reproductive fitness will nonetheless be inferior to males with excessively large, colourful (and strikingly impractical) plumage. The reasons why females initially fixate on these harmful characteristics as indicators of sexual compatibility are usually obscure, but once the process is set in motion this initial relationship becomes irrelevant to the strength of the force driving the population to favour more and more exaggerated instances of the indicator trait. As the process continues, natural selection continues to favour the reproductive benefits of the chosen indicator trait. The process therefore spirals into a situation where the population encourages characteristics that are more and more maladaptive with each successive generation until a constraint of some kind eventually keeps the process from escalating any further.¹⁰

In cultural examples of the runaway process, a parallel situation occurs where the selection forces acting on culturally maladaptive indicator traits are not strong enough to counteract the forces pushing individuals to adopt these characteristics. Without an understanding of how a process of this kind has originated one would expect such disadvantageous cultural patterns to be strongly selected against. However, because the pressure to economize on information costs is great enough to create powerful selection forces in favour of relying on indirect methods of determining cultural fitness, the structure of cultural evolution allows for the possibility of a runaway process where irrational, imprudent or otherwise harmful behaviour is favoured despite its presumably disadvantageous character.

The best example offered by Boyd and Richerson of a cultural runaway process involves a small island where the prestige of the men is considered to be a function of their ability to contribute very large yams to periodic feasts given by the chiefs of each

¹⁰ In one of the more famous instances of runaway sex selection, the constraint imposed on the size of Irish Elk antlers was the massive amount of weight that left sexy males virtually unable to lift their heads.
At these feasts the head of each farmstead provides a special prize yam that is evaluated as a symbol of the contributor's generosity and skill as a farmer. (Moreover, the chiefs promote the contributors of large yams to more distinguished positions in the community.) The most interesting aspect of this case is that the ability to grow massive yams has been exaggerated beyond the point where it might reasonably serve as an indicator of the farmer's capacity to contribute worthwhile nutritional sustenance to the feast. Instead, the growing of the yams clearly has become a cultural practice that may once have been an indicator of farming skill but is now purely a mark of prestige that has been exaggerated to bizarre extremes. This conclusion is confirmed not only by the fact that the yams sometimes reach a staggering nine feet in length, but by the observation that the nutritional value of other vital foodstuffs (including even the combined volume of multiple yams) is not taken into consideration to measure the farmer's deserved status. In fact, the yams used by families for food are much smaller and do not require the elaborate and extremely laborious techniques required to grow monster yams. The division between what is nutritionally valuable and the exaggerated prestige attached to the prized yams is so great that apparently families often "go hungry at home when they have large yams in their farms ready for harvest". Thus, the drive to produce prized yams is an example where a maladaptive cultural pattern continues to be promoted because the forces favoring the trait as an indirect indicator of fitness have overpowered the selective forces that would otherwise remedy this unproductive cultural practice.

The farming of monster yams on a small Micronesian island may seem a distant example, but the cultural forces supporting a runaway process also operate much closer to home. In fact, when we consider the various instances of prestige-based cultural practices that continue to exist in modern societies, yam farming may seem fairly

---

11 The example is described in (Boyd and Richerson 1985, p. 269-70), but the original information is gathered from (Bascom 1948).
12 (Bascom 1948, p. 212)
reasonable by comparison. If we reflect on such practices as teenage drag racing, body piercing, high-heeled shoes (let alone foot binding), circumcision (both male and female), the propensity to choose stylish over practical clothes in cold climates, drinking contests, breast implants, street gang initiations and Spain's yearly 'running of the bulls', it ought to seem painfully obvious that indirect indicators of prestige in modern industrialized cultures work in mysterious, exaggerated and often injurious ways. Furthermore, it is interesting to speculate whether some of these practices may have started out with less extreme forms of prestigious behaviour. By analyzing these social phenomena as runaway processes and by paying attention to the evolutionary pressures acting on the indicator traits in question, we may be able to explain how such seemingly drastic behaviour begins from innocuous origins. Some of these practices may have initially started from very mild preferences for bravery, fashion or sex appeal, but once the indicator traits become valued for their own sake these mild preferences can get locked into a runaway process where they are pushed towards excessive proportions. The most disturbing characteristic of this process is therefore that it is impressively easy to get cultural runaway effects going despite their often hazardous outcomes. As a result, these mechanisms within cultural evolution may be far more common that we would like to believe.\(^\text{13}\)

\(^{13}\) In fact, not only are runaway processes surprisingly easy to get going once selection forces begin to operate on initially harmless indicator traits, but more elaborate cultural forces can also contribute to these traits being promoted to exaggerated proportions. As Boyd and Richerson point out in a subsequent paper (Boyd and Richerson 1992), punishment mechanisms within cultural evolution can work to reinforce the continuation of maladaptive behaviour within social groups if being punished by the group is sufficiently costly compared to the individual benefits of conforming to the behaviour required to avoid such punishment. Boyd and Richerson reach this result in the course of investigating the conditions necessary for the evolution of cooperation in large sized groups, but, as they and Hirshleifer and Rasmusen (1989) carefully point out, the fact that punishment can sustain individually maladaptive behaviour is in no way dependent on this behaviour producing group benefits. Thus, if an effective punishment mechanism is established within a cultural system it can reinforce the promotion of any behaviour, no matter how arbitrary or potentially disadvantageous. As a result, punishment mechanisms greatly enhance the existing chances that runaway processes can entrench dysfunctional patterns of behaviour into the fabric of our cultural environment.
IV. Ethical Runaway Effects

If the structure of cultural transmission leaves our social institutions susceptible to runaway effects, can we expect a similar process to exist in the more specific context of our ethical concepts and intuitions? If so, we would presumably have an even more pressing interest in investigating cases where this occurs, since the consequences of allowing such distortions to go unnoticed in our ethical thinking could be more serious than a failure to notice exaggerated fashion trends or other marks of prestige in our surroundings. Yet, given the general structure of the runaway effects discussed in section three, it seems entirely plausible that similar mechanisms operate within the landscape of our moral intuitions. In fact, if we consider how prevalent our reliance on indirect cultural indicators tends to be, it would be surprising if there were not instances where our moral intuitions fall prey to runaway-like evolutionary forces. As we noted in section two, many of these instances may be identifiable without the help of a rigorous memetic analysis of our ethical beliefs. But there may also be cases where our moral intuitions have become indicators of other ethical values but we remain unaware of circumstances where these indicators have drifted away from their underlying ethical targets. In cases such as these - where moral memes have unknowingly become dramatically reinforced or exaggerated in the course of promoting other moral values - a detailed investigation of the memetic forces acting on our intuitions could prove to be extremely valuable.

*Complexity and Satisficing*

The evolutionary forces leading to a runaway process are impressively general, but they do rely on the existence of strong selection pressures acting to promote the use of indirect indicators in order for these indicators to become excessively promoted (e.g. 14 This might not be the case if the marks of prestige in question are as serious as female circumcision or the running of the bulls. Nevertheless, staggering as it is that such practices still exist, the dubious nature of their authority certainly does not go unnoticed.)
pressures giving individuals compelling reasons to prefer indicators over more direct assessments of their environment). In the examples of cultural runaway effects discussed above, the considerable benefits associated with looking to prestige as a means of cutting down on learning costs served to justify the fact that strong selection pressures would indeed favour a preference for indirect markers of prestige. Is there then a parallel force acting on our moral intuitions that could lead to particular memes being preferred as indirect ethical markers? I believe there is. Just as the general landscape of cultural options made available to us is complicated enough to favour appealing to indirect means of determining the most profitable behaviour to emulate given our limited experience and potentially hazardous surroundings, the landscape of moral options available to us as agents with limited knowledge and pressing time constraints is complicated enough to favour indirect means of determining the most appropriate behaviour to adopt given our complex moral circumstances and abstract guiding principles. Even if we assume that our moral objectives are consistent and legitimate, the sheer complexity of the choices we are forced to make in order to carry out the requirements of what morality demands in the real world makes it extremely plausible that selection forces will favour indirect means of guiding our intuitions to economize on the costs associated with real-time moral decisions.

The idea is analogous to Herbert Simon's influential observations of the nature of rational decision making in economics (Simon 1959). According to Simon, the classical theory of instrumental reason used in models of economic decision making is inadequate because the theory assumes too much about the cognitive capacities of rational agents (i.e. it assumes that these agents have perfect knowledge about the statistical probability and expected utility of the options available to them). In its place, Simon proposes a theory of "bounded rationality". On this view, agents indirectly maximize utility by appealing to proximate goals instead of attempting to calculate the expected utility of all available outcomes. The idea is that in situations of substantial uncertainty and
complexity, agents grappling with the realities of time constraints and their own limited cognitive capacities are better off relying on tangible heuristics in order to secure certain fixed, proximate goals instead of seeking to directly maximize their overall interests. This is described as "satisficing", since the theory claims that the best description of rational behaviour under conditions of uncertainty is one where "In-principle, unattainable optimization is sacrificed for in-practice, attainable satisfaction."\(^{15}\) (Simon 1979, p. 499)

Similarly, I think it is reasonable to expect that once a complex set of ethical imperatives has been established in a social system (or, rather, a simple set of imperatives is established in exceedingly complex circumstances), cultural evolution will begin to favour intuitions that serve as proximate targets in order to direct agents towards the attainable satisfaction of their ethical objectives under conditions of difficult and uncertain moral decision making. It is important to note, however, that neither Simon's theory of rational decision making or cultural evolution's tendency to promote indirect indicators of ethical value imply anything about individuals being aware of the satisficing principles guiding their decisions. In the case of bounded rationality, individuals may be aware of the fact that they are sacrificing direct maximization for the certainty of more tangible heuristics, but this awareness is not necessary for the model to give accurate predictions of economic behaviour. In the case of cultural evolution, an awareness of the satisficing principles responsible for economizing on moral decision making costs by creating proximate moral intuitions is also not necessary for the intuitions to have emerged from the dynamics of a complex social system. Hence, the claim that cultural selection pressures will lead us towards indirect ethical indicators is not a claim about whether individuals will knowingly direct themselves to adopt more tangible ethical

\(^{15}\) It is perhaps worth emphasizing, however, that the underlying goal of maximization is not given up. Instead, satisficing is adopted as the best possible means of achieving maximization in the face of extreme complexity and uncertainty. A helpful example is the chess player: she wishes to win as efficiently as possible, but since she is unable to compute every possible combination of future moves, she must adopt proximate strategies in order to best achieve the goal of winning, considering the complexity of the game.
strategies. Instead, the claim is that evolutionary forces will drive cultural selection to favour the promotion of shortcuts within our moral intuitions in order to make better use of our limited cognitive abilities considering the complex ethical situations we encounter.

*Ethical Indicators and Insulation*

If a general pressure exists to accept concrete heuristics as acceptable moral shortcuts, then the possibility arises that conditions occur where these shortcuts are reinforced or exaggerated beyond the point where they reliably indicate behaviour that would be in the best interests of our primary ethical objectives. This can occur because the mechanisms leading to a cultural runaway process are initially engaged by the pressure to economize on the information costs associated with moral decision making, and once this process is set in motion the selection pressures working against memes that do not advance our ethical interests may be overpowered by the underlying drive to find the most efficient means of choosing appropriate actions within complex moral environments.

How exactly can moral shortcuts become reinforced or exaggerated beyond the point where they ought to be acting as indicators of our moral interests? In the representative example of a cultural runaway process listed above, the farming of monster yams provided a clear illustration of a cultural pattern that had become exaggerated to the point where it no longer reliably indicated its original objective. Yet, in the case of our moral intuitions it is not as clear how the exaggerated character of a runaway process will manifest itself. The answer, I believe, is that the content of our moral intuitions will not so much be exaggerated as they will be reinforced to the point where they become insulated from our ability to critically reflect on their merits. Hence, the cultural selection forces favouring indirect means of achieving our moral objectives will tend to forcefully entrench ethical indicator memes within our psyches so that we will be generally unwilling to act against the recommendations of these memes no matter what
the circumstances. In short, the pressure to economize on the costs of moral decision making will favour the strong \textit{internalization} of moral shortcuts, so that the relevant moral indicator memes will continue to exert a powerful normative pull on our choices regardless of the details of our particular situations.

Within reasonable limits this strategy is advantageous, because acting on the guidance of internalized moral norms allows us to avoid the pitfalls of having to continually negotiate the complexity of our immediate moral circumstances. In the case of a runaway process, however, the insulation of moral shortcuts can be reinforced to the point where these shortcuts become so deeply internalized that their normative influence resists even the most compelling situation-specific reasons for acting contrary to their recommendations. Hence the exaggerated character of an ethical runaway process consists in the internalization of a proximate moral intuition to such a degree that this intuition continues to hold a strong sense of normative authority over us in situations where the goal this cultural shortcut is designed to promote is clearly better served by some other course of action. It is in this respect that the unusual structure of the cultural mechanisms operating on our moral intuitions allows the confidence we place in our moral intuitions to potentially become distorted. By promoting the deep internalization of routinely praiseworthy memes, cultural evolution opens up the possibility that these moral shortcuts will be internalized to such exaggerated proportions that the intuitive pull they exert on us becomes disengaged from the characteristics that make the memes praiseworthy as indicators of appropriate moral conduct.

The idea of disengagement is important, because strong internalization alone is not necessarily a sign that our intuitions have been produced by an illicit runaway process. For example, consider the broad intuition that it is wrong to kill others. This intuition is usually (one would hope) internalized by individuals in a profound way. Yet, when faced with situations where agents are forced to kill in clear self-defense, the
authority carried by the intuition not to kill drains away without much difficulty.\textsuperscript{16} Thus, the existence of runaway effects in our moral intuitions is not alarming solely in virtue of the fact that these effects lead to the strong internalization of moral prescriptions. The existence of ethical runaway effects is disturbing because the shortcuts produced may be internalized in such a way that the advantages of relying on their guidance become disengaged from the power of the memetic selection forces favouring the shortcut, leaving us unable to recognize situations where we ought to abandon the authority these shortcuts ordinarily possess.

The idea that moral shortcut memes are susceptible to being internalized in this way demonstrates how helpful an awareness of the evolutionary forces acting on our moral thinking might be. If we are ordinarily unable to notice cases where the strength of an intuition has been generated by the insulating mechanisms of a cultural runaway process, then gaining access to details about the intuition's evolutionary history will provide us with an additional means of evaluating its legitimacy. This extra empirical information may not be required if we have no cause to doubt the authority of our internalized intuitions (or we have already found good reason to reject their authority). But in cases where our intuitions are in conflict and we find that we must choose between the competing merits of otherwise compelling intuitions, our understanding the genealogy of these memes might allow us to spot instances where generally praiseworthy intuitions continue to exert a normative pull on us in contexts where they no longer serve the objectives we would endorse if fully informed of their cultural evolutionary history.

\textsuperscript{16} This is not to say that anyone finds it "easy" to kill another person, even in self-defense. The point is that the \textit{prescriptive authority} in question drains away, so that in clear cases of self-defense we do not have any difficulty arriving at the judgment that one is legitimately entitled to disregard the deeply internalized intuition not to kill.
Human Life is Sacred?

Consider the following examples. First, suppose that we could trace the evolutionary history of the meme that all human life is necessarily sacred. The genealogy of this meme would likely reveal that it has been able to flourish in our cultural environment because it serves as an important proximate indicator for numerous other valuable moral ends. The idea that human life is sacred acts as an efficient justification for the prohibitions we want to uphold against actions we sometimes find ourselves tempted to commit. Such actions include disrespecting persons with disabilities, sacrificing individuals for what appear to be greater purposes, and harming (or perhaps failing to rescue) persons we resent or to whom we have no significant emotional attachment. Because the 'human life is sacred' meme is so adept at cutting through the suspicious rationales we tend to give for unacceptable behaviour, it has probably been strongly selected for by the cultural forces driving our ethical institutions. Moreover, the meme is concise and efficient enough to have been internalized by vast numbers of people with secular belief systems who have long since abandoned the religious connotations that may have originally given the idea an additional sense of intuitive force.17

If the genealogy of the intuition that all human life is sacred reveals that this meme possesses an evolutionary history like the one described above, then I think we would have good reason to describe this meme as a proximate moral shortcut - a useful way of safeguarding appropriate moral decisions in situations that might otherwise be complicated by distractions and the seductive logic of less virtuous options. Furthermore, in this case the shortcut seems like a very reasonable one to accept: as a general rule the intuition that all human life is sacred is tremendously appealing. However, when a moral

---

17 For a detailed articulation of the idea that human life is inviolably sacred, see (Dworkin 1994), especially pp. 81-101.
meme is internalized as powerfully as this one has been, it becomes possible that cultural runaway effects may be produced as by-products.

Consider some of the issues addressed in bioethics concerning tough cases at the fringes of both the beginning and end of life. With respect to the former, there are many occasions where the technology of modern medicine leaves us with difficult choices to make concerning the fate of unborn human life. Prenatal testing provides us with the opportunity to abort pregnancies where the fetus is confirmed to be carrying highly deleterious abnormalities, such as trisomy 13, Tay Sachs disease and anencephaly. Yet many parents choose to continue with these pregnancies despite the fact that the children born are destined to live extremely short and painful lives. It seems many parents feel that even in these extreme circumstances there is something morally questionable about terminating a form of human life, no matter how decisive the reasons for doing so may be. In these cases, we might recognize that there are compelling reasons to end a pregnancy, yet we may also find it extremely difficult to set aside our intuition that there is something necessarily sacred about the form of human life at stake. These situations are likely generated by cultural runaway processes, since the authority our intuitions exert over us appears to be the result of cultural selection forces that have internalized the 'human life is sacred' meme beyond the point where it best represents the underlying ethical objectives it has been promoted to safeguard.

Other cases concerning the treatment of prenatal human life can bring the difficulty associated with this meme into even sharper focus. Consider the fact that embryonic neural tissue is used for medical research and has now been successfully used in an experimental treatment to help patients with Parkinson's disease regain some degree of normal functioning. A surprisingly large number of people, many of whom have no preexisting objection to the abortion of embryos at similar stages of gestation, feel very uncomfortable with the use of embryonic tissue for this purpose. Moreover, the aversion most people have to using embryonic tissue for medical treatment is not connected to
concrete claims about the potential harms that such a practice might produce. Instead, the claim is largely inexpressible. For many people there just seems to be something significantly wrong with treating a form of human life in this manner. Thus, it seems likely that in this case the intuition that all human life is sacred has been internalized in such a way that the selection pressures responsible for the meme's intuitive power have become detached from the forces we would want to endorse from an informed, critical perspective. The most compelling evidence that this is true is that individuals find it difficult to escape from the grasp of this meme's intuitive appeal despite the fact that they find themselves at a loss to explain why or to articulate any further justification for the intuition that might cause us to believe that it ought to outweigh the benefits of medical research.

Plus, there are potential candidates for ethical runaway effects when this meme is applied to moral dilemmas at the opposite end of human life. In cases where people have fallen into irreversible comas or have reached the final stages of painfully debilitating diseases, we may have compelling reasons for ending their lives. Yet we may nonetheless feel strong resistance to ending even the most hopeless of these cases because the intuition that such action violates the general sanctity of human life continues to keep a firm grip on us. Once again, in many of these situations we find that our best critical assessment of the situation indicates that terminating the person's life is the most ethical action possible considering the circumstances, but the intuitive pull of the meme that human life is sacred may continue to keep us in a distressing state of moral conflict.

If we had access to the evolutionary genealogy of the intuition that all human life is sacred, resolving ethical dilemmas like those listed above might be a somewhat more tractable exercise. Specifically, if the genealogical information about the meme allowed us to see that the intuitive power it possesses in these unique cases is derived from the meme's ability to serve as an indirect indicator of other moral objectives that are not relevant to the circumstances at hand, then this information would give us good reason to
reconsider the authority we might otherwise attribute to the meme's seemingly powerful appeal. In effect, the genealogy of moral memes allows us to examine the cultural mechanisms supporting memes such as the intuition that all human life is sacred, so that we can determine whether we would continue to accept the authority of these memes' apparent credibility if we knew exactly why they seemed so inexplicably compelling. Without this resource, we risk allowing the influence of runaway effects created by the indirect cultural fitness of these memes to disrupt our ability to make moral decisions according to what we would rationally endorse from a critical standpoint.

*Pornography is Harmful to Women?*

As a second example, consider the meme that pornography is necessarily harmful to women. A vast amount of literature has been written on the subject of pornography and the potential harm it causes women. Much of the literature is focused on whether or not this harm is sufficient to justify censoring violent pornography or pornography that is clearly degrading to women (including both the women involved in its production and women in general, who are victimized by the attitudes it promotes in those who consume it). However, an ongoing debate also exists over the question of whether consensual, nonviolent sex can be depicted in pornography without somehow harming women in less dramatic but nonetheless significant ways.

Many feminists believe pornography *inherently* harms women no matter how innocuous its content may appear. By promoting subtle forms of sexual domination and making the existing inequalities between men and women seem sexy and natural, pornography is said to be an inevitably oppressive medium by which patriarchal beliefs are legitimized. The two most notable proponents of this view are Catharine MacKinnon and Andrea Dworkin. MacKinnon famously argues that pornography constructs a definition of female sexuality as "lust for self-annihilation". (MacKinnon 1989, p. 212) and Dworkin asserts that pornography is designed "to control, dehumanize, humiliate,
injure, and subordinate". (Dworkin 1997, p. 35) Both MacKinnon and Dworkin are especially concerned with eliminating the violence against women they believe pornography incites, but to do so they indict all forms of pornography as complicit elements of an inevitably injurious industry. Hence, they and other antipornography feminists believe pornography necessarily victimizes women and further entrenches their subordinate status.

On the other hand, some women argue that an unqualified condemnation of pornography leads to a repressive social atmosphere in which many of the changes feminists seek will be inhibited. These women are uncomfortable with the alliances struck between conservatives and antipornography feminists, and they consider the limiting of free expression to be a more dangerous option than allowing the exchange of sexually explicit material. Many believe pornography is in fact harmful to women. (They simply take the view that censorship is a worse alternative.) However, some anticensorship feminists refuse to accept even the initial assumption that pornography is always harmful. Instead, they view it as a potentially valuable form of expression - one that can be used to promote the sexual liberation of women if it is hijacked from the exclusive control of men and produced from a non-misogynistic perspective. In this respect, a significant number of feminists tend to resist the categorical claim that pornography is an unavoidably injurious enterprise.

In the debate between these two dissenting views within feminism, I think the idea that pornography is necessarily harmful can be plausibly interpreted as the product of a cultural runaway process. Specifically, there are surely strong cultural selection forces acting to promote this meme as a means of satisficing over the complexity involved in accounting for the harms pornography inflicts on women. Pornography affects women in a myriad of very subtle ways and need not involve overt violence or degradation to produce damaging results. It can have a profoundly suggestive effect on the way men and women perceive their sexual identities and gender roles, and seemingly
consensual acts that involve unequal power dynamics can perpetuate harmful assumptions about women's sexual desires, or even their broader attitudes towards their subservient societal status.

In fact, in many cases it is precisely the consensual character of the sex being depicted that is detrimental, because pornography of this kind encourages women to become willing participants in their own exploitation. This can occur through coerced sex that is made to seem consensual. But it can also be the result of women willingly accepting sexual roles that diminish the quality of their lives after their own interests and desires have been shaped by a cultural system that is disrespectful of their personal dignity and autonomy. Within such a system, women are left unable to freely choose autonomous conceptions of the good life, and they are subsequently caught in a self reinforcing cultural cycle that leads them to embrace misogynist societal norms - a result that undermines (among other things) the possibility of genuinely consensual pornography. Thus, even the most willing female participants may be victims of false consciousness who are unaware of the harm being done to them and to the women influenced by their seemingly enthusiastic involvement in sexually explicit material.

The question of false consciousness underscores the fact that consensual pornography is often harmful in very subtle ways that are not easily measured or even identified. Sex is an exceedingly complicated, and in many ways peculiar, human activity. The full extent of its emotional and psychological impact on us is thus extremely difficult to account for, and the harmful effects of its presentation in pornography are easily overlooked. For this reason it is understandable that the

---

18 See for example Linda Lovelace's harrowing account of being coerced into sex on the set of the now infamous film Deep Throat. (Lovelace and McGrady 1980)

19 For an interesting defense of the idea that liberals can look to Mill for a justification of censoring pornography in order to protect the autonomy of women because of potential harms of this kind, see (Dyzenhaus 1992).

20 These easily overlooked harmful effects are of course not confined to issues of false consciousness. They include perpetuating myths about women having insatiable sexual urges, setting sexual expectations
intuition that pornography is necessarily harmful seems so persuasive. The intuition serves as an efficient cultural shortcut that relieves us from having to catalogue the innumerable ways in which pornography generates harmful effects that are not always immediately obvious.

Yet, persuasive as this intuition may be, it is clearly a meme that in some cases has become reinforced beyond the point where its prescriptive authority ought to be respected. It is simply not believable that all forms of pornography are necessarily harmful to women. Gay and lesbian pornography, for example, represent clear instances where sexually explicit material can be used as an empowering means of self-expression for persons whose sexual preferences are otherwise unfairly discouraged. Moreover, there is no reason why depictions of heterosexual sex are necessarily excluded from having a similarly empowering potential. There may be a multitude of subtle ways in which such depictions can be harmful, but it is unlikely that every instance of consensual pornography is an example of false consciousness or a presentation of misogynist ideals. In fact, it would be condescending to claim that every woman who supports or is actively involved in explicit material that displays female sexuality in a positive light is a victim of false consciousness. In the past twenty years a growing number of women have begun to reclaim their sexuality through art or entertainment that could be considered pornographic. These women often show commendable self-understanding and a keen awareness of the options available to them, and there is no categorical reason to think that they will be unsuccessful in employing sexual images to further their autonomous aims.

Under special conditions, then, the meme that pornography is necessarily harmful is surely unacceptable. Yet it is difficult to shake off the authority of this intuition because it is commonly internalized so deeply as an effective moral shortcut. That it has

that can undermine women's self-confidence, and leading men to generally view women as sexual objects to the extent that men are less likely to take women seriously in professional roles.
been internalized in this way is not surprising considering the depth of suffering women have experienced because of pornography. But if we could trace the genealogy of this intuition and uncover its cultural role as a proximate indicator of the harms pornography inflicts on women, then we would be in a much better position to reevaluate its prescriptive authority in situations where we suspect it is inapplicable to the circumstances at hand. As useful as this meme may be as a general moral shortcut, we do not want to fall victim to a runaway process where our intuitions are reinforced to the point where we are unable to escape from the meme's intuitive grasp in cases where honouring its authority is counter-productive to women's interests.

The issue is important because our critically reflective attitudes towards consensual pornography affect how we structure an effective feminist response to the harms that pornography generates. If we trust the intuition that pornography is necessarily harmful, then we should resist attempts to ameliorate the content of pornography and instead redirect our efforts to fighting it in its entirety from a more detached, legalistic standpoint. However, if we can identify the intuition that pornography is necessarily harmful as a potentially untrustworthy runaway effect, then the possibility arises that we can co-opt the medium of pornography to promote non-patriarchal sexual ideals.

At the very least, investigating the genealogy of this intuition invites us to pay attention to the detailed reasons why the idea replicates so successfully in our cultural environment, and these reasons ought to help reconcile the antagonism between antipornography and anticensorship feminists. By uncovering the fact that absolute condemnations of pornography serve as a means of satisficing over the unacknowledged harms inflicted on women, the absolutist position can be viewed as sensible rather than fanatical. By demonstrating that the absolutist position is in some cases the product of a

---

21 See for example the lengthy and extremely disturbing testimony presented at the Minneapolis civil rights hearings, compiled in (MacKinnon and Dworkin, 1997).
cultural runaway process where our intuitions have been detached from their ethical objectives, antiscensorship feminists can be viewed as activists in the reclamation of female sexuality rather than apologists for misogyny or victims of false consciousness. Thus, in addition to helping to identify the best way to fight the harms of pornography, examining the genealogy of our intuitions will help us promote better communication and break down the rhetorical walls that add confusion to what is already a complex issue.

_Ecosystems are Inherently Valuable?_

Finally, as a third example consider the meme that ecosystems possess their own inherent moral value. In the past two decades, the sub-discipline of Environmental Ethics has witnessed the noteworthy emergence of _ecoholism_ as a theory of environmental value. Ecoholism (also referred to as ecocentrism) attributes moral standing to holistic natural systems and claims that our primary obligations are to protect the ongoing vitality of these systems. Rather than claiming that natural systems derive their value from their instrumental ability to promote the welfare of their individual components (e.g. humans, non-human animals or even individual trees), the value of natural systems is said to depend upon the proper functioning of the systems themselves. Hence, the theory recommends that we restructure our normative priorities to reflect a holistic environmental perspective in which the primary locus of value is considered to be the proper functioning of broad ecological habitats. In the lasting words of Aldo Leopold, whose writings serve as the principal inspiration for the move towards ecoholism:

A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise.

(Leopold 1949, p.262)

Beyond this clearly articulated criterion of rightness, Leopold's own writing falls short of offering a fully developed environmental philosophy. However, others have been strongly influenced by his work and have endeavored to construct more robust
environmental theories based on the pivotal principle that natural ecosystems possess inherent moral value.\textsuperscript{22}

The move to ecolohicist ethics is a radical step, yet the underlying idea that ecosystems ought to be respected as sources of intrinsic value is in many ways completely understandable. There are a multitude of potential harms that can result from tampering with the integrity of complex natural systems, and many of these harms are (like the harms of seemingly consensual pornogrophy) potentially devastating but impressively difficult to recognize in the moment of ordinary moral decision making. Thus, if we try to account for our obligations to the integrity of natural ecosystems from an individualistic perspective, we may overlook considerations that are vital to the functioning of the natural systems in which we inhabit. Individualistic attempts to account for the complicated relationships that bind interdependent systems simply tend to be myopic in the face of such profoundly intricate structures. For this reason, individualistic theories often appear unable to provide sufficient consideration to the depth and complexity of our feelings towards our natural environment.

It is not surprising, then, that cultural evolution favours the intuition that ecosystems possess their own separate intrinsic value. Much like the intuition that what is natural is necessarily good (discussed in chapter three), the meme that ecosystems are inherently valuable functions as a shortcut to compensate for our inability to recognize the complexity involved in maintaining our natural environment. It is a \textit{satisficing} shortcut that is selected as a proximate indicator of the imperspicuous benefits that natural ecosystems provide their individual components. The meme has therefore gained acceptance as a reaction to the shortsighted individualism exhibited by our species in attempting to manage such ecosystems.

\textsuperscript{22} The most outspoken of Leopold's disciples is J. Baird Callicott. See (Callicott 1989).
Unfortunately, though, the intuition that natural ecosystems possess their own inherent value is a shortcut with unfavorable side-effects. In particular, this meme can fall victim to a runaway process if it is internalized to the point where it leads to the promotion of natural systems at the dramatic expense of the individual components of these systems. If we take Leopold's original ecoholist maxim seriously as a criterion of rightness, it clearly states that our primary ethical duty is to the biotic community as a whole. Thus, it strongly suggests that this duty ought to take precedence over our responsibilities to individual entities within the biotic community. If this suggestion is adopted, however, ecoholist ethics leads to plainly contemptible outcomes. For example, if we must choose between saving the life of a starving child and saving a rare and beautiful flower, our duty to the integrity and beauty of the biotic community as a whole requires that we choose the latter. (One more human is a liability to its ecosystem, whereas a rare type of flower is a valuable asset.) Moreover, if we had the opportunity to eradicate a deadly disease like malaria or AIDS, an ecoholist outlook ought to prevent us from doing so. Deadly diseases undoubtedly contribute to the integrity of our existing ecosystems, and a consistently holistic approach leaves us with no way to account for the overwhelming benefits that eradicating these diseases would bring to the individual members of the ecosystems they inhabit.

The tendency for ecoholism to sacrifice individuals for the sake of the greater environmental whole has therefore led to sharp criticism. L.W. Sumner, for instance, describes the idea that the biosphere has a good of its own that is distinct from the goods of its individual members as "not just nonsense, but dangerous nonsense".\(^\text{23}\) and Tom Regan has gone as far as to label ecoholism "environmental fascism".\(^\text{24}\) This is tough talk, coming from philosophers, but these critics are surely justified in objecting to the offensive outcomes implied by ecoholism. To protect the integrity of ecosystems at the

\(^{23}\text{See (Sumner 1986, p. 81).}\)
\(^{24}\text{See (Regan 1983, pp. 361-62).}\)
dramatic expense of the very individuals who are able to benefit from preserving this integrity seems like dangerous nonsense indeed. Yet if we interpret the rising popularity of ecoholism as the product of a runaway process, we can see why ideas of this kind could arise from the overzealous adoption of an otherwise profitable moral shortcut.

Thankfully, few proponents of ecoholism, if any, advocate the kinds of extreme implementations of holistic ethics that would lead to repugnant results. Most ecoholists instead seek to modify the doctrine in order to ensure that it produces more moderate recommendations. But when we address tough issues in environmental ethics, even moderate interpretations of ecoholism confuse intelligible debate and often, ironically enough, act as a hindrance to furthering the goals of the biotic community. By serving as a shortcut for the complicated reasons why stable ecosystems benefit their individual constituents and then lending itself to absurd interpretations, the idea that ecosystems are inherently valuable actually undermines the goal of protecting the healthy functioning of natural ecosystems, because in many cases it encourages skeptics to dismiss the shortcut in its entirety without investigating the subtleties associated with the value of our natural environment. We can see, then, how a genealogical analysis of the intuitive appeal ecoholism has achieved will be helpful. Instead of accepting or dismissing the credibility of this meme in its entirety, if we trace its cultural genealogy and investigate the reasons why it manages to persist in our ethical landscape, we can gain an understanding of why we ought to respect the immense complexity of natural ecosystems without allowing this intuition to dangerously cloud our judgment.

---

25 See, for example, (Moline 1986) and (Marietta 1988).
26 In this respect moral shortcuts are inherently risky, for they rely upon a kind of 'Maginot Line' strategy by investing their intuitive power in the ability of just one meme. If this single meme is rejected, all the complexities built into the shortcut are discarded in one swift act of cultural selection, as opposed to the underlying complexities having to be dealt with one by one.
Conclusion

As I hope the preceding examples demonstrate, when we are faced with moral dilemmas that involve strong intuitions that seem to be inconsistent with the rest of our carefully worked out ethical judgments, appealing to the cultural genealogy of our intuitions can serve as an additional source of guidance about the trustworthiness of these suspicious intuitions and the authority we think they ought to possess. This source of guidance is distinctive not only because it enables us to identify the aberrant structure of runaway processes within our ethical thinking. It also satisfies an important explanatory function by providing a plausible account of why these intuitions manage to replicate so successfully in our cultural environment. The empirical information offered by memetics is therefore unique because it serves both as a means of evaluating the trustworthiness of our moral intuitions and as a means of explaining why we sometimes find ourselves unable to give up on the authority of our intuitions in cases where they clearly conflict with our primary ethical objectives.

This, I submit, is the most coherent interpretation of what Dawkins asks of us when he recommends that we rebel against the tyranny of the selfish replicators. Memetics tells us that cultural evolution promotes the longevity of whatever patterns of information most effectively replicate and distribute copies of themselves in our cultural environment. Plus, it reveals that the transmission mechanisms responsible for directing our shared thoughts and institutions are imperfectly designed. Hence, unless we actively guard against the possibility of our memes replicating according to their own peculiar logic (as opposed to what we can at least minimally agree is in our best interests), then in an important sense we will be allowing ourselves to be manipulated by the blind evolutionary design of selfish replicators instead of our own willful intentions.

This chapter has therefore defended two claims. First, I argued that it is not incoherent to rebel against the memes of which we are composed, because we can take up a critical perspective towards some of these memes from within the provisional context of
other memes in which we have a more substantial level of confidence. By appealing to the Rawlsian conception of reflective equilibrium, I hope to have demonstrated that approaching the question of critical reflection in ethics will be no more arduous using the terminology of memetics than the terminology of considered moral judgments. Second, I argued that the most promising way for memetics to have an impact on ethics will be to appeal to this reflective standpoint in order to identify distortions in our moral intuitions that can occur because of transmission anomalies like a cultural runaway process. By allowing us to identify cases where our moral intuitions have been influenced by selection forces that favour indirect indicators of ethical value, the information provided by memetics can help us prevent instances where our intuitions are insulated to the point where they no longer represent reliable shortcuts to our general ethical objectives. If memetics can improve our ability to remove these arbitrary elements from our moral thinking, I believe it will have a remarkably valuable effect on our ability to systematically justify the authority we attribute to our prescriptive ethical judgments.
Chapter 5

*A Critical Setback: The Corrosive Effects of Memetics*

**Introduction**

At this point it is worth summarizing what has been presented so far. We have seen in chapter one that knowledge of our genetic social dispositions cannot serve as a direct source of prescriptive value, but that as a source of empirical information this knowledge can help us to systematize the authority we assign to our moral intuitions by uncovering the otherwise obscure circumstances of their sociobiological history. We have also seen, in chapter two, that the emerging science of memetics can serve as an important source of empirical information about the evolutionary forces guiding cultural transmission. If memetics succeeds at mapping out the evolutionary dynamics of our thoughts and cultural institutions, it will surely provide us with a valuable source of insight into the cultural genealogy of our moral judgments. Chapter three argued that appealing to this genealogical information in order to more effectively assess the trustworthiness of our ethical intuitions will be the most productive application of memetics to ethics. Chapter four developed this proposal in more detail by describing how the genealogy of moral memes will enable us to identify cases where the peculiar structure of cultural selection reinforces our intuitions beyond the point where they reliably indicate our best interests. Chapter four also defended the claim that although all our ideas are ultimately composed of memes, we can nonetheless take up a critical standpoint towards the memes that constitute our common-sense moral intuitions.

The conclusion reached at this point suggests that memetics will have a significant impact on the authority we attribute to our moral intuitions. By alerting us to potential distortions that arise in the cultural transmission of our moral attitudes, investigating the genealogy of our moral memes provides us with an opportunity to eliminate undesirable side-effects from our moral thinking and thus to more effectively
systematize the authority we ascribe to our prescriptive moral judgments. Considering how dependent we are on indirect indicators of cultural value and how often distortions in our intuitions occur when our moral indicators are subverted by anomalous selection pressures like the cultural runaway process, the benefits of applying memetics to ethics seem extremely promising.

However, in chapter five an unsettling problem is introduced for the project of using memetics to systematize ethics. Although the genealogical information supplied by memetics can help us regulate the authority we place on our moral intuitions, the level of critical reflection necessary to accomplish this goal carries considerable risks. Specifically, the process of isolating a particular ethical meme and examining the merits of its genealogy is a very delicate exercise. It is clearly beneficial to identify the evolutionary pressures that allow dysfunctional moral attitudes to propagate, but the process of singling out the cultural histories of our moral intuitions can nonetheless be a surprisingly hazardous undertaking.

The general difficulty for this methodology is that taking up a critical perspective towards the genealogy of our moral intuitions invites a temptation to dismiss the authority of more than just dysfunctional memes. Once we begin questioning the prescriptive merits of our moral intuitions, we cannot limit the scope of our inquiry to just those memes that have been hijacked by the effects of a runaway process. Instead, a Pandora's Box is opened for all our intuitions, and it becomes alarmingly easy to begin scrutinizing (and potentially rejecting) the authority of memes that serve valuable moral ends - memes that are not disengaged from their moral objectives but may nevertheless appear to be unsuitably arbitrary when examined from the detached perspective of a critical standpoint. Thus, the danger in appealing to the genealogy of our intuitions from a critical standpoint is that by seeking to undermine the normative control that a
particular set of memes exert over us, we may end up freeing ourselves from much more than we originally bargained for.

Two ways this can occur are addressed in this chapter. Section one looks at how projects seeking to explain the significance of moral norms in terms of more basic components can lead to the dissolution of praiseworthy intuitions whose subtle advantages are not easily integrated into the context of an instrumental framework. Section two examines an even more serious instance where justifiable moral memes may be dismantled by an investigation into cultural genealogy: cases where an ethical intuition is specifically designed to operate in situations where critical reflection brings about self-defeating outcomes. Both sections demonstrate that the otherwise useful methodology of adopting a critical perspective to examine the cultural origins of our intuitions can have a corrosive effect on the authority of many legitimate moral norms.

I. The Trouble with Genealogy

If the conceptual machinery used to weed out runaway memes is functioning properly, then critical reflection ought to expose the genealogy of intuitions that possess a sense of normative authority in situations where this authority no longer promotes our best interests. What exactly constitutes our moral best interests is left indeterminate in this investigation. Depending on the moral framework one accepts, the goals and corresponding cultural selection pressures considered to be ethically appropriate will vary significantly. Yet, whatever framework one appeals to, our reliance on indirect ethical indicators and the general structure of cultural runaway processes ensures that our moral intuitions can sometimes be internalized to the point where their prescriptive control outstrips their aptness. Thus, appealing to the genealogy of our moral memes provides us with a very specific capacity to identify cases where our intuitions are promoted by obviously anomalous cultural selection forces (i.e. forces that have become disengaged
from the selection pressures that normally lead us to whatever it is we consider to be reasonable and appropriate\(^1\).

Unfortunately, allowing ourselves the capacity to reflect on our moral intuitions opens up a problem for the prescriptive authority of many legitimate ethical norms. The problem is that each time we adopt a critical standpoint towards our moral intuitions we have to be able to clearly recognize the genealogical value of the intuition under investigation. If we are somehow unable to discern the ethical merits of an intuition it may appear to be unsuitably arbitrary and we may subsequently strip it of its normative status. In itself, this may not seem like a problem. Weeding out memes we believe to have been selected by imperfect cultural forces is exactly how the process of appealing to the empirical information provided by memetics is supposed to function. Yet this methodology sets a high standard for our critical abilities to execute if we want to retain the normative control that many useful moral precepts exert over us. By giving ourselves the onerous responsibility of evaluating the cultural genealogies of our intuitions, we rely on the assumption that we are capable of recognizing the kinds of complex cultural interactions that promote valuable moral ends. For if we fail to properly appreciate the evolutionary forces that produce justifiable intuitions, we risk prematurely dismissing the authority of memes that serve valuable social roles but for some reason possess obscure genealogical histories.

Subtlety and the Recognition Problem

The worry, then, is that many legitimate moral intuitions have been constructed by cultural evolution in complex and potentially indecipherable ways. Because our

---

\(^1\)Recall how Daniel Dennett expresses the distinction between pernicious memetic forces and those that lead to what we view as being reasonable and appropriate by identifying the latter category as the set cultural forces that produce just what we would expect in terms of such general properties as "good", "truth" and "beauty". (Dennett 1995, 361-69) Dennett's treatment of the distinction is helpful, as he appropriately emphasizes the fact that it is the production of unexpected deviations from what is good, true or beautiful that is intriguing in memetics, rather than the conventional difficulty that these properties are not at all easily agreed upon.
ethical memes have been shaped by an elaborate and exceedingly complex process of cultural selection, the requirement that we be able to identify the complicated series of evolutionary pathways that have implanted these memes in our collective consciousness is a demanding one. The example of the runaway process is presumably a fairly cut and dry case where cultural selection forces have ceased to function in an appropriate manner. But if we set for ourselves the task of eliminating cases where selection forces appear defective, we risk becoming overconfident about our ability to recognize the many ways in which cultural selection can function 'normally'. Cultural evolution works in mysterious and very subtle ways. Moreover, because selection forces optimize without foresight they often produce surprisingly sophisticated results out of long, convoluted genealogical histories. If we are at all careless or short-sighted when we evaluate the data supplied by memetics, we will likely fail to appreciate the value of many subtly constructed but justifiable moral memes. The process of dissecting the ethical merits of the cultural genealogies of our intuitions is thus a rather forbidding exercise. If we are unable to clearly recognize the value of a meme's evolutionary history we will perceive the meme as a product of arbitrary selection forces and attempt to purge ourselves of its prescriptive authority, even if we discover later that the meme's subtle consequences are sorely missed.

Consider, for example, the peculiar virtue of modesty. Recent literature has grappled unsuccessfully with the task of defining exactly what modesty is, yet most would agree (and do agree since the literature almost always seeks to vindicate modesty's ethical merits) that we would be made worse off if we stripped this meme of its intuitive authority. Indeed, modesty seems to both serve an important social function and constitute a praiseworthy personal characteristic. However, the reasons why this is the case remain mysterious. This mystery begins with the following dilemma: either the modest agent is aware of her talents and merely presenting a reduced estimate of herself to others (suggesting deception), or she is genuinely unaware of the extent of her talents
and subscribes to an inaccurately low opinion of herself (suggesting ignorance). With
decception on one side and ignorance on the other, it is difficult to see how either of these
interpretations can account for modesty's virtuous character. Hence, many authors have
engaged in spectacular forms of philosophical gymnastics in order to reconcile the
enigmatic quality of modesty with what we perceive to be its obvious moral worth.2

No conclusive solution has yet been produced by these hypotheses. In fact, if
anything, recent investigations of modesty only dramatically reveal that much more is at
work beneath the surface of this virtue than the brief Oxford English Dictionary
definition ("having or expressing a low estimate of one's own merits or achievements")
might suggest. The debate over modesty instead points out that the value of this is virtue
derived from a complicated array of social considerations: one's interest in minimizing
jealousy or feelings of inferiority in others, the propensity to carefully avoid
overestimating the merits of one's own accomplishments, possessing an appropriate
attitude towards unhealthy forms of ranking and competition, and perhaps the realization
that one's achievements are not nearly as significant as we believe they are when
evaluated in a broader context (e.g. according to what matters most in life or the grand
design of the universe). The many different aspects of modesty's value indicates that the
moral significance of this virtue is the culmination of an elaborate and obscure fusion of
cultural components.

With this in mind, consider what might occur if we were to examine the
genealogy of modesty from a critical standpoint. Faced with this perplexing set of social
factors contributing to our intuitions, we would likely be unable to distinguish a clear
evolutionary pathway for modesty that has been forged by recognizably justifiable

---

2See Julia Driver's "The Virtues of Ignorance" (Driver 1989), Owen Flanagan's "Virtue and Ignorance"
(Flanagan 1990) and G.F. Schueler's "Why Modesty Is a Virtue" (Schueler 1997). The most gymnastic of
these accounts belongs to Schueler, who gives an account of modesty so immune from objections based on
duplicity and ignorance that he ends up with something that looks nothing like modesty at all: "the
disposition to not care at all whether people are impressed with [one] for [one's] accomplishments". (p.
840-1)
cultural selection forces. Thus, the cultural genealogy of modesty could be so convoluted that we abandon all hope for deciphering its genealogical merits. If so, the promotion of truthfulness and accurate self-knowledge could be strong enough considerations to make us want to dissolve the intuitive authority that modesty currently possesses. Without evidence to the contrary, it will seem as if modesty is simply a peculiar social side-effect we are better off without. (Moreover, a case could be made that the normative status modesty currently enjoys is specifically a cultural runaway effect - an offshoot from Christian accounts of humility which postulate that one ought not take credit for one's accomplishments since these accomplishments are derived not from one's own talents but from the gifts that God has chosen to implant within us.\(^3\)) From a critical standpoint, unless we can identify permissible cultural forces acting to promote the intuition that modesty is valuable then the meme's authority ought to be eliminated.

This, however, would be a shame, as would eliminating the authority of many other moral memes with equally obscure cultural genealogies. Modesty therefore serves as a reminder of the fact that the value of our moral intuitions may be exceedingly subtle and difficult to measure from a critical standpoint. It demonstrates that singling out individual memes and subjecting them to critical scrutiny places a dangerous burden of proof on our ability to recognize the fact that our moral intuitions have been selected by cultural forces that we consider to be legitimate. In many cases, this difficulty can be avoided because the distinction between what is defective and what has been selected by cultural pressures favouring rationality and our shared moral interests will be straightforward. Yet, by critically examining our intuitions according to cultural genealogy, we still risk penalizing intuitions with more subtle ethical merits, since our inability to correctly decipher the forces that sustain their prescriptive merits can lead us to prematurely dismiss their legitimate normative status.

\(^3\)See (Richards 1992).
Temptation and Self Interest

The problem discussed so far, hereafter referred to as the 'recognition problem', is made even more acute by the fact that we want to be able to apply the machinery of critical reflection to particular situations where the authority of an ethical intuition is called into question.\(^4\) In chapter four I described cases where ethical memes could be generally beneficial yet harmful or misguided in certain specific contexts. This occurred when the authority of these memes had been internalized to such an extent that our intuitions were unable to adapt to circumstances where compliance no longer served what we would reflectively endorse as being in our best interest. To prevent runaway effects from hijacking the authority of generally acceptable intuitions, it was suggested that we adopt a critical perspective towards the genealogy of these intuitions in specific cases where our continuing to honour their normative authority could be the result of a localized runaway effect (e.g. situations where the otherwise praiseworthy intuition that 'all human life is sacred' sanctions questionable moral norms in specific bioethical dilemmas).

Allowing ourselves the capacity to scrutinize moral memes in special situations helps us to combat these runaway effects, but it also compounds the recognition problem by creating situations where we will be tempted to interpret particular circumstances as grounds for taking exception to the authority of our legitimate intuitions. If we know that there are specific cases where even the most exemplary moral memes ought to have their prescriptive authority scaled back, we may be seduced into thinking that we experience these cases far more often than they actually occur. Considering how powerful self-interest can be at causing us to see only what we want to see, the temptation to interpret

---

\(^4\)Note that the 'recognition problem' is not so labeled only because there are difficulties involved in recognizing the complex genealogies of our moral intuitions. The problem is also that we can fail to recognize the moral significance of genealogies that we may correctly discern. Thus, I am using the term to denote a broad range of errors of interpretation that may result because of our limited abilities to analyze complex information about the ethical significance of what we examine from a critical standpoint.
the details of immediate moral situations as grounds for excusing ourselves from our ethical obligations will be lamentably compelling. Enabling ourselves to critically reflect on the genealogy of our intuitions in particular instances will therefore have a corrosive effect on the authority of legitimate moral norms, because it will facilitate our propensity to improperly construe the empirical data provided by memetics as grounds for exempting ourselves from the prescriptive demands of justifiable moral memes.

In short, the ability to evaluate our intuitions at any particular moment plays into our existing propensity to believe that we can opt out of our moral commitments whenever they seem unclear or especially burdensome. Playing into this existing propensity is worrisome because every individual instance where moral rules ought to be obeyed often feels as if it is an exceptional case and that the authority of the rule in question is not necessarily binding. For example, each time someone lies to avoid social embarrassment, runs a yellow light because they think no one is entering the intersection, or copies software because they think 'everyone does it' is sufficient justification for their action, they usually feel that the authority of the rule being broken is absolved momentarily because of the unique details of their particular situation. Even when the rules being broken are more severe, at the moment of choice people often feel as if their specific circumstances immunize them from the authority of ordinarily sanctioned moral norms. (The following excuses ought to sound sadly familiar: "Yes, but no one was looking.", "Yes, but there was a lot of money at stake!", "True, but no one got hurt, OK?" or even "Yes, but he/she was asking for it!") Consider, for example, the all too common view that stealing from any large, lucrative corporation (especially one as vilified as Microsoft) is somehow uniquely justified. No one holding such a view would want to deny the authority of the prohibition against stealing. The point is that we are often lulled

---

5Note that these are not examples of Akrasia, where one knows exactly what one ought to do yet suffers from an overwhelming weakness of the will that leaves this person unable to follow through on his or her beliefs. In the cases cited above one does not experience weakness of the will so much as one is tempted to believe that circumstances are exceptional enough that one is temporarily relieved of one's otherwise binding moral duty.
into thinking that the circumstances related to the wealth and anonymity of the company somehow relieve us from our ordinarily binding moral obligations.

Thus, by opening up the possibility of examining the genealogical merits of our intuitions in particular contexts, we tap into a very seductive pre-existing vulnerability within ourselves to twist the details of our moral situations to our own advantage. With the knowledge that runaway effects can cause generally trustworthy memes to carry too rigid a sense of prescriptive authority in certain instances, this temptation will potentially cloud our judgment and cause us to turn a blind eye towards important justificatory details within the evolutionary histories of our moral memes. The recognition problem is in this way made that much more corrosive by the power of self interest. This corrosive effect is mitigated somewhat by the fact that the genealogical critique of our intuitions will often take place in a social, rather than individual, context. But the tools to implement this kind of critique will eventually be left to individuals seeking the best course of action in their day to day lives. When this occurs, critical reflection into the genealogy of our moral memes in special cases will make us more likely to cry 'exception' because of what appear to be extenuating circumstances.

*Moral Instrumentalism and Moral Minimalism*

It is worth noting that neither of the two problems raised in this section present wholly unavoidable problems for the task of critically reflecting on the authority of our moral intuitions. If we meticulously examine the complex cultural genealogies of our moral intuitions we can avoid situations where we prematurely dismiss the legitimacy of memes that subtly promote our best interests. Similarly, if we can resist the temptation to misinterpret the details of our particular circumstances as grounds for evading our moral obligations, then we can avoid situations where we dismiss the authority of a generally sanctioned moral intuition because self-interest has led us to believe it has fallen into the grip of a localized runaway process. Neither of these two issues constitute insuperable
obstacles for gaining an understanding of the evolutionary forces that promote our moral intuitions, yet both introduce serious difficulties for the methodology proposed in chapter four simply because we happen to be so especially vulnerable to their respective pitfalls. It is impressively difficult to be both meticulous and impartial enough to prevent the corrosive consequences that short-sightedness and self-interest create for an investigation into the properties sustaining our moral intuitions. Even if these consequences are not necessarily unavoidable, they are prevalent enough to be virtually inseparable from attempts to account for the underlying components of morality from a detached, critical standpoint.

Thus, the process of breaking down our moral intuitions into the basic evolutionary forces responsible for their existence tends to have a powerfully corrosive effect on the authority of these intuitions. In fact, the strength of this tendency is so significant that I believe a non-memetic parallel of it is largely responsible for the more general opinion that instrumentalist theories in ethics cannot possibly capture the full extent of what we ought to value when we reflect on the content of our moral judgments. By 'instrumentalist' theories I mean approaches in ethics that seek to provide explanations for the our moral beliefs in terms of more basic, instrumental components. In this investigation, a memetic view of ethics qualifies as what I will refer to as an instrumentalist approach because it attempts to explain our moral intuitions in terms of the cultural selection forces by which they have been promoted. However, other projects in ethics can similarly be broadly construed as examples of moral instrumentalism. The attempt to explain instances of presumed intrinsic value in terms of more basic sources of instrumental value is an obvious example. Other examples include utilitarian

---

6 Many of these approaches involve reductions of moral claims into more basic elements. But identity relations of the kind required for such reductions need not be attempted for an approach to seek to explain the significance of moral claims in terms of more basic elements. Hence, I use the term 'instrumental' in order to capture the broader category of these approaches.
constructions of rights and virtuous character traits, interpretations of political rights attributed to groups in terms of aggregations of individual rights, and attempts to provide game theoretic interpretations of moral obligation in terms of rational self-interest. All of these projects share the characteristic that they attempt to explain one or more ethical concepts or intuitions in terms of another set of seemingly more basic properties.

Not surprisingly, each of these projects has been met with resistance from critics claiming that such projects inevitably fail to account for the total significance of the concepts in question. Yet many, if not all, instrumentalist projects in ethics fully intend to ascribe the same level of prescriptive authority to their instrumental constructions as is held by the ethical concepts they seek to explain using more basic components. Utilitarian accounts of rights seek to establish the same prima facie inviolability as any reasonable deontologist account would provide. Utilitarian constructions of virtue attempt to justify the cultivation of genuinely virtuous dispositions among moral agents. Defenders of group rights derived from individual rights nonetheless argue that group rights ought to carry legitimate political power. Game theorists seek to justify our moral obligations via self-interest, not dismantle them. Hence, the aim of these instrumental conceptions is not to explain away the significance of the ethical properties in question. Rather, the aim is simply to explain how these properties ought to be valued as aggregations of more fundamental components.

The tension between the fact that instrumentalist theories in ethics aspire to retain undiminished levels of prescriptive authority and the fact that critics are passionate about the inability of these theories to do so can be sensibly accounted for by the risks outlined earlier in the chapter. Despite all the best intentions, our limited capacities for

---

7 See, for example, (Sumner 1987) for a consequentialist account of rights. See (Railton 1988), (Slote 1992), and (Sumner 1998) for discussion of utilitarian accounts of virtue.
8 See (Kymlicka 1989).
9 See (Gauthier 1986).
recognizing the complexity of instrumental aggregations and our vulnerability to self-interest ensure that straightforward implementations of instrumentalist projects will tend to erode the prescriptive authority assigned to our moral concepts and intuitions.

The problem is rooted in a very general phenomenon: by placing a burden of proof on our ability to identify the complicated interactions necessary for the basic components of our moral institutions to account for more elaborate ethical constructions, we invite a situation where all mistakes lead to a diminishment of authority of these constructions. The implementation of critical reflection required to examine the underlying components of our moral beliefs is therefore highly corrosive, since its every mishap unintentionally leads to a weakening of the authority of our existing moral norms. As a result, moral instrumentalism has a tendency to degenerate into moral minimalism by attributing prescriptive authority to only those moral norms that can be easily recognized as suitable aggregations of more basic components.

Listening to the Critics of Instrumentalism

This purely methodological inertia driving instrumentalist theories towards moral minimalism is rarely cited by critics as an objection against such theories being able to generate adequate levels of authority. Yet I think that this inertia is greatly responsible for the belief that there is always 'something more' that instrumentalist accounts necessarily fail to capture. Despite protest to the contrary, the sense that there must always be something valuable omitted by instrumental constructions is, I think, deeply intertwined with the suspicion that even if such constructions could theoretically succeed at representing the full extent of their objectives, they would be too complex to allow ourselves to be trusted with their implementation. (In fact, the belief in some critical moral remainder that necessarily exists above and beyond any instrumentalist theory is likely a cultural shortcut that has evolved to safeguard against the tendency for instrumentalist theories to drop into moral minimalism. Proponents of the view that
moral instrumentalism is always insufficient will undoubtedly be infuriated by this instrumentalist interpretation of their own objection to instrumentalism. However, their position rests on exactly the kind of inexpressible intuition that looks suspiciously as if it has been internalized as a means of 'satisficing' over the details necessary to overcome temptations that lead us to adopt narrow conceptions of prescriptive moral authority.) In any event, the methodological obstacles that lead to a diminishment of our moral norms confirm the alarming conclusion drawn by critics that straightforward applications of instrumentalist projects in ethics tend to come up short on prescriptive authority if left unchecked.

Thus, without conceding the point that moral instrumentalism can, in theory, attach sufficient levels of authority to our moral norms, adherents of instrumentalism have sound methodological reasons for paying close attention to their critics. And there has certainly been no shortage of critics. Within contemporary philosophy, numerous scholars have expressed a deep sense of dissatisfaction with what I broadly describe as moral instrumentalism. These scholars include such leading figures as Jonathan Dancy, Ronald Dworkin, Bernard Williams, Thomas Nagel, Charles Taylor, Philippa Foot, and Alasdair MacIntyre. Each one of these philosophers has launched his or her own unique arguments against the perils of instrumentalism within the respective contexts in which he or she is interested, yet they all share the conviction that their instrumentalist targets fail to capture something essential about the depth and plurality of what ethics represents.

Dworkin, for example, has voiced strong misgivings about instrumental attempts to account for intuitions such as the concept of rights, our intuitions about the intrinsic value of species and our (secular) belief in the sanctity of life. Williams, on the other hand, has...

---

10 See (Dancy 1993), (Dworkin 1977, 1994), (Williams 1973, 1981 a, 1981b, 1985); (Nagel 1986), (Taylor 1982, 1991, 1992, 1995), (Foot 1985) and (MacIntyre 1984). This long list is not even meant to be exhaustive; the critics of explaining ethical concepts in terms of instrumental components are many indeed.
hand, is deeply suspicious of the entire basis for theory-driven, systematic approaches to philosophy. For him, the attempt to dissect the structure of our moral intuitions inevitably distorts and trivializes the richness of our shared ethical experience. Any attempt to make this experience fit the constraints of moral theory is, from Williams’ point of view, a move that ignores the important but largely indistinct aspects of what it means to lead a meaningful life.\textsuperscript{11} Taylor too shares this suspicion about the role of theory in philosophy (or the role of what he refers to as 'epistemology') and its effects on the integrity of persons. Moreover, Taylor, Dancy and Foot share Williams’ particular distrust for utilitarianism (a theory Williams backhandedly credits as being "the most ambitious ethical theory") and its procedural aim of reducing all ethical judgments to a single, quantitative dimension.\textsuperscript{12}

All of these critics maintain that the versions of moral instrumentalism they attack cannot possibly address the full extent of our moral experience. Thus, each one attempts to provide \textit{categorical} arguments for dismissing the robustness of instrumental moral explanations. Yet in doing so, many of their arguments manage to pinpoint the distinguishing features of the \textit{contingent} methodological problems associated with the recognition problem and our ability to investigate the underlying elements of our moral intuitions. Consider the following indicative passages from instrumentalism's critics:

Many people think that great paintings, for example, are intrinsically valuable. They are valuable, and must be respected and protected, because of their inherent quality as art, and not because people happen to enjoy looking at them or find instruction or some pleasurable aesthetic experience standing before them. [...] The thought of [one of Rembrandt's self-portraits] being destroyed horrifies us -

\textsuperscript{11}What \textit{exactly} these aspects are that get left out by moral theory is rarely discussed in specific detail. (To give a detailed outline of their content would perhaps be to distort or misunderstand them... ) They are often referred to only as "thick" ethical concepts, such as \textit{courage, promise, or brutality} and contrasted with "thin" moral concepts directly related to obligation, such as \textit{duty, rights, or utility}. However, in the context of his critique of utilitarianism, Williams gives the closest thing to a concrete answer to the question by claiming that utilitarianism (and all other agent-neutral moral theories) invariably alienates us from what is important to the integrity of our life projects and unity of our personal character - an issue that will be discussed later in the chapter. See both (Williams 1973) and (Williams 1981a)

\textsuperscript{12}See (Williams 1985, p. 118), (Taylor 1982), (Dancy 1993) and (Foot 1985).
seems to us a terrible desecration - but this is not just because or even if that would cheat us of experiences we desire to have. We are horrified even if we have only a very small chance of ever seeing the painting...
(Dworkin 1994, p. 72; my emphasis)

Theory looks characteristically for considerations that are very general and have as little distinctive content as possible, because it is trying to systematize and because it wants to represent as many reasons as possible as applications of other reasons. [...] Theory typically uses the assumption that we have too many ethical ideas, some of which may well turn out to be mere prejudices. Our major problem now is actually that we have not too many but too few, and we need to cherish as many as we can.
(Williams 1985, pp. 116-7; my emphasis)

A consequentialist theory, even one which had gone beyond utilitarianism, would still be a Procrustes bed. It would once again make it impossible for us to get all the facets of our moral and political thinking in focus. And it might induce us to think that we could ignore certain demands because they fail to fit into our favoured mode of calculation. A meta-ethics of this kind stultifies thought.
(Taylor 1982, p. 144; my emphasis)

The passage quoted from Dworkin is revealing because even though he argues that it is impossible for instrumental considerations to account for the perceived intrinsic value of certain objects, he is careful to address an example where we still have a small chance of obtaining the instrumental benefits derived from such objects. Moreover, in a footnote Dworkin defends this decision by claiming that the abstract case where there is absolutely no possible chance that any instrumental benefits could be derived (e.g. if all human beings were somehow eliminated but Rembrandts remained) is not pertinent to his discussion.

However, it seems obvious that the abstract case where instrumental influences are wholly eliminated is crucially pertinent to the discussion. (This seems obvious, at least, to Moore, who takes on the more ambitious task of constructing elaborate thought experiments, as part of his "method of absolute isolation" in order to isolate potentially intrinsic goods from their instrumental consequences.13) Otherwise, we still have strong reason to suspect that the concerns grounding our intuitions could be related to doubts

---

13See (Moore 1994), chapter six.
about prematurely cutting ourselves off from the possibility of acquiring subtle forms of instrumental value.

Williams' passage is interesting because he so astutely perceives the tendency for critical theorizing to diminish our ethical ideas. Not only does he express the fear that systematizing our moral experience will have corrosive consequences, but he seems keenly aware of the fact that this fear is based on the burden of proof created by our tendency to accommodate only those ideas we clearly recognize as having ethical value (and dismiss whatever is left as mere prejudice). This fear seems suspiciously analogous to a methodological concern, since Williams cannot be objecting to the mere assumption that we have some prejudices among our ethical ideas. (This assumption seems inevitably true.) The objection must be, rather, that the burden of proof associated with this assumption leads to excessively minimalist conclusions about the kinds of ideas we allow to play a role within normative debate.

Finally, the selection taken from Taylor suggests that he has a similar methodological warning in mind. He initially claims that it is impossible for (consequentialist) theory to capture all of the different facets of our moral thinking. However, he then goes on to warn against the possibility that we "might" be led to ignore certain ethical demands because they fail to fit into the terms of our favoured theoretical calculus. This second point is an interesting extension of the initial claim that ethical theory is necessarily a Procrustes bed, since the second point suggests that once the project of systematization is implemented it might lead us to prematurely dismiss certain ethical ideas rather than investigating their (potentially subtle) merits in further detail. The second point is not necessarily in conflict with the first, but it is noticeably different than the initial absolute objection leveled against instrumentalism's capacity to account for all facets of our ethical thinking. In fact, Taylor's second claim is virtually identical to the recognition problem discussed earlier. Instead of being an in-principle prohibition, it
is a general methodological worry about the fact that instrumental projects are vulnerable to corrosive side-effects if they are not administered extremely carefully.

We might draw a number of conclusions from the recurring theme found in these selections. One conclusion is that these critics may be overstating their case by confusing the claim that instrumentalism could never account for the full extent of our experience with the separate methodological concern that such projects often fail to live up to the scope of their objectives. While this may be true, I will not argue for this controversial hypothesis here. Instead, I want to suggest that if we set aside the issue of whether moral instrumentalism is viable in the abstract we can interpret the critics of instrumentalist projects as having identified something genuinely problematic about the implementation of such projects. This is the fact that the kind of critical reflection necessary to explain our moral experience in terms of more basic components is impressively difficult to maintain without incurring losses to the normative status of our more subtle moral beliefs. This point is often ignored in the rush to argue against the more extravagant claims made by critics arguing that instrumentalism is categorically inadequate. However, I think that when the arguments of these critics are examined closely, they strongly suggest that even the most hard-headed proponents of moral instrumentalism ought to take notice of the issues they raise - even if one interprets these arguments as methodological cautions rather than absolute theoretical obstructions.

II. The Methodology that Knew Too Much

So far, we have seen that critical reflection aimed at the genealogy of our moral intuitions will likely have a corrosive effect on the authority of legitimate ethical norms because: a) our limited ability to understand the intricate workings of cultural selection will prevent us from recognizing the value of norms selected by forces that very subtly promote our best interests, and b) the temptation to interpret our immediate situations as instances where generally authoritative intuitions have been distorted by localized memetic
anomalies will augment our existing vulnerability to believing that special circumstances relieve us from our moral obligations. Though both of these cases are not unavoidable (they can, in principle, be avoided by careful, impartial investigation), their contingent effects are strong enough that the project of systematizing the genealogy of our moral memes can easily drop into a form of moral minimalism.

In this section, however, an even deeper problem is addressed. As we have seen, cultural evolution is capable of promoting exceedingly complex moral memes that have subtle social consequences. However, not only is it possible for cultural evolution to promote memes with genealogical histories that are too complex for us to recognize: it is also possible for cultural evolution to promote moral memes that operate precisely because we are not directly aware of the underlying causes that allow them to function. In these instances, critical reflection can be carried out with the best intentions and perfect execution, but examining the genealogies of our moral intuitions will still end up diminishing the authority of very useful moral memes. Thus, the problem addressed here is not one of deficient realization with respect to a memetic investigation of our moral memes. In these cases, it is precisely the successful realization of our desire to understand the cultural history of our moral intuitions that brings about self-defeating outcomes.

Collectively Self-Defeating Outcomes

Within each of us lies a multitude of internalized social norms that allow us to more effectively navigate our respective cultural landscapes. These norms have been passed on to us, for the most part, through the intricate workings of our early socialization. We can certainly continue to learn and adapt to new social norms after this initial process is complete (e.g. most adults can switch to driving on the left-hand side of the road when vacationing abroad). But the vast majority of customs and social cues that enable us to interact with the rest of society without obstruction are learned early and then
deeply internalized. The result is that many of the social rules we follow have been ingrained within our attitudes to such an extent that their execution is virtually automatic. At the moment of action we rarely contemplate the hypothetical merits of upholding these conventions - we simply uphold them because cultural transmission has implanted their significance in our minds as direct reasons for action.

In this respect, there is a fundamentally normative component to human social action. This normative component consists of the guiding principles that provide us with direct reasons for obeying the perceived rules of our cultural context. By way of contrast, this normative facet of our behaviour inhibits us from engaging in the ordinarily instrumental deliberation used to determine what we ought to pursue as our most profitable courses of action. Social norms can certainly be obeyed in some cases because of instrumental concerns (e.g. I may choose to dress a certain way because I fear being mocked by my peers), yet they are often followed strictly because they have become second nature and directly motivate us to action without regard to hypothetical consequences (e.g. when I agonize over the math required to tip sufficiently in restaurants where I will surely never return). This capacity to integrate social norms into our direct reasons for action is unique, and perhaps somewhat alarming, in that it cuts us off from the benefits of instrumental reason. Yet, in certain situations we are actually better off with this cord severed.

Useful as instrumental reason is at securing our desires more often than not, the literature of game theory is saturated with coordination problems and practical dilemmas that reveal its unfavorable pitfalls. The most famous of these is the Prisoner's Dilemma - a situation where the combined logic of instrumental reason displayed by two players curiously results in suboptimal outcomes for both.14 Examples like the Prisoner's

---

14The original statement of the Prisoner's Dilemma is as follows. Two criminals are caught and questioned separately by the police. If both remain silent, they will each be charged with something less serious than the crime for which they are guilty. However, if only one of the criminals provides testimony against his accomplice, he will be set free and his accomplice will take the full extent of the blame for the crime. If they both testify against one another, they will share the penalty for the more serious offense. The situation
Dilemma have been useful in illustrating that circumstances exist where the successful use of individual instrumental reason leads to collectively self-defeating results. Yet, the heuristics of decision theory only highlight a problem that has been recognized at least as far back as Hobbes. The problem is that if individuals are left to exercise their own capacity for practical reason without some external means of ensuring that their interests coincide, the result is a lack of solidarity in our collective social actions that causes suboptimal outcomes and potentially leads to the "war of every man against every man". (Hobbes 1968, p. 188) Hence, the underlying problem game theory is so adept at clarifying is that the benefits of instrumental reason are tenuously dependent on some separate means of coordinating our otherwise disparate individual desires.

Not surprisingly, the existence of social norms in human culture fits this coordination role nicely. By guiding our behaviour towards certain pre-defined conventions, the ability to follow rules prescribed by our socialization enables us to avoid the coordination problems and tricky social dilemmas that otherwise plague the use of instrumental reason in unstable public contexts. These norms are not perfect, but they certainly seem to help raise human collective action above the level of a full-blown state of nature. The key to their ability to accomplish this task without Leviathan-like coercion is precisely the fact that they remain opaque to direct instrumental analysis. Hobbes himself noted that it is the individual capacity for "private" reason that must be overcome in order to ensure public stability, and this is exactly the service human socialization provides. Social norms constitute an efficient means of harmonizing collective action

---

is a dilemma for instrumental reason, since it is clearly rational for each of the criminals to testify against the other, yet when both do so they secure a worse outcome for themselves than if they had both kept quiet. The situation is interesting because the general structure of the dilemma can be shown to apply to a broad range of social situations, including everything from environmental pollution to traffic jams to nuclear deterrence. See William Poundstone's Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb (Poundstone 1992).

15For an elegant and exceedingly clear discussion of the advantage of normative social action over straightforward instrumental reason, see (Heath 1998). For a general discussion of the ways in which convention can evolve as an evolutionary response to social dilemmas, see (Skyrms 1996).
precisely because they hold the specific ability to bypass the hazards linked to our individual reliance on instrumental rationality.

How cultural evolution has managed to achieve this favourable result is another issue. The evolution, cultural or otherwise, of group-advantageous traits that are not individually profitable is a complex and highly controversial area in evolutionary game theory and the philosophy of biology. The problem for such traits is that no matter how profitable their collective benefits may be, the fact that they are individually disadvantageous (compared to the value of not contributing to their creation but continuing to reap the benefit of their public rewards) invites a free rider problem that erodes the adaptive advantage of group-advantageous traits from the inside.\(^\text{16}\) Moreover, the standard means of avoiding this result (i.e. via kin selection and conditional reciprocity) apply only to restricted circumstances where individuals either share a similar genetic heritage or interact in exceedingly small groups - conditions that do not hold in many cultural systems.\(^\text{17}\) The fact remains, however, that we continue to observe these traits in human culture. Plus, there have recently been several promising attempts at building evolutionary models of how cooperative social norms might evolve and survive in larger interacting groups.\(^\text{18}\) It therefore seems likely that, despite their mysterious origins, social norms have somehow been promoted by cultural selection because they so elegantly prevent us from being 'too clever by half' and bringing down a heap of social dilemmas on ourselves.

\(^{16}\text{For further information about the difficulties and evolutionary merits of group-advantageous traits and their selection, see: (Wilson 1983), (Dugatkin and Reeve 1994) and the massive collection of commentary that follows (Sober and Wilson 1994).}\)

\(^{17}\text{See (Axelrod and Hamilton 1981) and (Boyd and Richerson 1988).}\)

\(^{18}\text{As was mentioned in chapter four, recent work on punishment mechanisms within cultural systems may be opening new territory for evolutionary explanations of the existence of group-advantageous traits in large groups of unrelated individuals.}\)
Fragile Social Norms

This result comes as no surprise to scholars of sociology who are by now familiar with the move to defend normative explanations for social action. An impressive tradition of theorists, including Weber, Durkheim, Parsons and Geertz, have argued that an essentially normative component is required to successfully explain human social behaviour. A survey of the material in this tradition is well beyond the scope of this chapter. However, the important characteristic in the work of each of the authors in this tradition is that they appropriately emphasize the irreducibly normative quality of social action. Whether it is faith in Weber's charismatic leader, the unifying power of Durkheim's supervenient collective consciousness, the necessity of Parsons' "voluntaristic" action or the motivating character of Geertz's symbolically encoded cultural values, the essential feature these explanations recognize is that the norm-related reasons for action that individuals give themselves cannot be explained by instrumental considerations without distorting the nature of our social experience. Thus, by looking at what it is that directly motivates us to action, these authors pinpoint the unique quality necessary for social norms to help us escape the collective pitfalls of individual reason.

But if the effectiveness of social norms is critically dependent on their not being understood instrumentally, then investigating the genealogy of our moral memes carries the potential for damaging results. Many of our ethical intuitions are dependent on some set of culturally transmitted social norms. If we give ourselves the capacity to critically reflect on the origins and instrumental value of these norms, we risk allowing our capacity for individual reason to corrupt the immediately motivating quality that these norms possess. If this occurs, we may irreversibly lose the advantage of being provided with reasons for action that are immune to the social dilemmas associated with the individual use of instrumental reason. Thus, by seeking to understand the cultural history

19See (Parsons 1949) for both a summary and lengthy analysis of the sociological turn towards normative explanations for human behaviour.
of those norms that underlie our moral intuitions (and therefore the instrumental reasons why they have been promoted by cultural selection), we may inadvertently release a genie from the bottle of our shared memetic heritage and unintentionally unravel the intricate moral socialization that keeps us free from collectively self-defeating outcomes.20

One may wonder how this genie can be released if the normative aspect of our behaviour is truly irreducible to instrumental reason. But it is important to realize that the irreducibility in question refers only to the fact that an explanation of what motivates us to action when we comply with social norms cannot be redescribed in terms of instrumental reasons without distorting the nature of our social experience. This irreducibility is therefore specifically related to the task of describing the immediate causes for human action - it is not a judgment about whether social norms possess metaphysically dubious holistic properties. As a result, it is possible to uncover instrumental explanations for the existence of social norms in terms of their cultural fitness and evolutionary history, while simultaneously acknowledging that the subjective reasons why agents act in accordance with these norms cannot similarly be so reduced.

Unfortunately, though, a corrosive potential is released for the status of our social norms once we uncover the instrumental explanation for why they survive in our cultural landscape. The genie is released by the possibility that once we recognize the forces responsible for the cultural fitness of our social norms we will probably be unable to go back to allowing it to directly motivate our behaviour. This will not present a problem for all social norms with ethical consequences (e.g. I recognize perfectly well what the instrumental reasons are for driving on a designated side of the road, but this does not lead me to alter my compliance.) However, in cases where a social norm specifically

---

20To be specific, our socialized norms keep us from the individual pursuit of self-interest, which is what Derek Parfit describes as a "directly collectively self-defeating" theory. Technically, the theory that gives each person the aim of choosing outcomes that would be best for himself is directly collectively self-defeating if it is certain that, if we all successfully follow this theory, we will cause the aims of each to be worse than they would have been if none had successfully followed the theory. See (Parfit 1984, pp. 53-6).
allows us to overcome social dilemmas because it prevents the exercise of individual reason, critical reflection on the evolutionary origins of this norm will cause us to view its merits instrumentally and potentially dissolve its social function.

The problem is most obvious in situations exhibiting the structure of a multiple player Prisoner's Dilemma. If a social norm evolves that automatically directs agents toward cooperative behaviour, then revealing the norm's memetic genealogy will succeed only in alerting individuals to the structure of the social dilemma in question and the fact that they are each better off by maximizing according to individual reason. (E.g. if one had never examined the structure of the norm for carpooling, an analysis of highway congestion as a social dilemma will succeed only in revealing that traffic will be the same each morning whether one drives at one's own leisure or chooses to submit to the inconvenient schedules of one's coworkers.) By giving us the blessing of being able to systematically justify our obedience to social norms, the capacity for critical reflection carries the added curse of encouraging us to reason instrumentally about these norms - a move that precipitates the temptation to disregard their otherwise instinctive authority.

Consider, for example, the fact that economics students have been shown to be more likely to defect (i.e. choose straightforward maximization) compared to students sampled from other disciplines in experiments where individuals are asked to play Prisoner's Dilemma styled games.\footnote{See (Frank 1988), pp. 226-239. The results Frank cites are taken from (Marwell and Ames 1981). Frank also notes results from a similar experiment where economics students were shown to be more likely to maximize rather than act according to norms of fairness in a 'Divide the Dollar' game. (Kahneman et al. 1986).} There are numerous possible explanations for this behaviour (e.g. that these students are disproportionately influenced by economic theory to think that they ought to always maximize profit, or that individuals who choose to study economics have a selfish personality type to begin with). But the fact that these students are trained in the field of rational choice whereas other students are (usually) not suggests at least one plausible explanation: that a formal understanding of the forces at work in a
social dilemma makes one more likely to disregard the socialized norms that otherwise lead to cooperative collective action. But this should hardly be surprising: it is more rational for each individual to disregard their social norms and to choose to defect in a Prisoner's Dilemma. That is the dilemma! What is interesting is that an awareness of this fact is not such a good thing, and that those who have not had the benefit of studying the intricacies of rational choice are somehow able to resist the obvious logic of instrumental reason and instead choose to promote collective rewards.

Unfortunately, gaining access to the genealogy of our moral intuitions will be much like teaching economics students the rules of rational choice. In the hope of systematizing our intuitions from a critical standpoint, the methodology proposed in chapter four will expose exactly what allows collective social norms to function in order to assign prescriptive authority to our intuitions according to their instrumental ability to further our interests. However, an awareness of the way our social norms function is often exactly what leads to the temptation to ignore their intuitive power. The result is that each agent attempting to invoke this methodology will be more likely to see that complying with what our norms demand of us is not individually profitable. Consequently, those memes related to the principle that 'everyone ought to do their fair share' (the general feature of norms designed to solve situations resembling multiple player Prisoner's Dilemmas) will be placed in jeopardy. Everything from cleaning up after one's dog to recycling, carpooling, and giving a percentage of one's income to charity will no longer seem automatically expected. Instead, these norms and others designed to solicit contributions to a public good will be judged on the merits of their instrumental value.

This is not to say that an understanding of the genealogy of these norms will undoubtedly lead to their dissolution. It is certainly possible to perceive the individual reasons for defecting in a social dilemma and still choose to cooperate (presumably because other ethical considerations motivate us to 'do our fair share'). The point is that
this understanding opens up an *additional source of temptation* with potentially corrosive consequences - one that is not created if we instead choose not to investigate the genealogy of our moral intuitions. It is also worth noting that this temptation is distinct from the temptation outlined earlier in section one. There, it was shown that critical reflection leads to a temptation to misinterpret the significance of our intuitions' genealogies. Here, there is no problem with our ability to recognize the information supplied by memetics or its significance. The problem is precisely that if we *successfully* understand the significance of the evolutionary forces supporting our social norms, we will be exposed to strong instrumental reasons for discarding their motivational authority. If we succumb to this temptation and pursue individual goals instead of acting on the authority of our internalized cultural conventions, we will unfortunately discover that the sound execution of each individual's rational aims brings about worse outcomes for everyone than would have been brought about if we had not discharged our capacity for instrumental reason. In short, we will end up in exactly the kind of collectively self-defeating situations that our social norms have evolved to help us overcome.

*Individually Self-Defeating Outcomes*

Despite the inertia pushing us towards temptations with corrosive moral consequences, to this point we have not seen anything that poses a completely insurmountable problem for investigating the genealogy of our moral memes. If we imagine highly idealized conditions and the possibility of added cooperative norms, the kind of critical reflection required to support our methodology could still conceivably avoid the pitfalls related to the recognition problem and collective social dilemmas. Nevertheless, in this section we shall see that there are circumstances where even ideal critical reflection and proactive ethical principles cannot escape producing self-defeating results.
The problem stems from the fact that certain moral goods cannot be pursued in a straightforward manner without sacrificing what it is that makes them commendable. The most common examples of such goods are virtuous dispositions and character traits. In particular, dispositions that allow us to achieve fulfilling personal relationships seem especially susceptible to being corrupted by the attempt to systematize their value or the basis for their moral justification. The source of this susceptibility is the fact that these dispositions are only virtuous if they lead to action that is carried out for its own sake or the sake of those involved in our relationships. If these actions are not pursued for these reasons - if, for example, they are pursued explicitly because they are thought to be virtuous acts - then they curiously forfeit the quality that gives them their praiseworthy character. Thus, dispositions for such esteemed goods as friendship, parenthood and romantic love are only valuable in so far as they motivate us to act for the sake of our friends, children and lovers. If these dispositions motivate us to act strictly because morality judges that friendship, parenthood and romantic love are worth promoting, then we somehow appear to have lost sight of what it is that makes these characteristics worthwhile.

This particularity in the way in which certain moral goods must be pursued creates a dilemma for theory-driven approaches to ethics. If we promote only those goods that critical reflection sanctions as being ethically justified, then our reasons for respecting their authority become dependent on their impartial value and justification. If this is the case, our motivations are no longer exclusively related to the features responsible for generating the valued character of moral goods like those derived from personal relationships. Therein lies the dilemma: one cannot allow theoretical considerations to guide our interest in promoting virtuous goods, because in doing so we neglect to appreciate the source of their moral value and ensure the failure of our own objectives. Strange as it may seem, seeking to achieve the best possible ethical results can in some cases guarantee our inability to successfully do so.
Moreover, the paradox here applies to virtues other than those associated with personal relationships. Consider such virtuous character traits as courage, generosity, sympathy or benevolence. If we praise someone for their benevolent attributes and they respond by stating that they do their best to live up to this description purely because they feel it is morally required, then although we may continue to view their actions as helpful we will likely be disappointed by the fact that they care for others because of impartial considerations and not because of an immediate concern for agents themselves. To be virtuously benevolent is to be directly interested in the well-being of others - it is not a quality that can be forced or even willfully adopted.\(^2\) (In this respect, virtuous character traits represent the motivational antithesis to the Kantian requirement that acts must be performed from a sense of duty in order to be morally praiseworthy.) Thus, a theoretical approach to determining the best way to promote moral goods will be unable to directly capture what is valuable about authentically virtuous dispositions.

_Fragile Moral Virtues_

This dilemma poses a serious difficulty for the aim of investigating the cultural genealogy of our moral intuitions. The difficulty is that the methodology required to systematize our moral memes and eliminate runaway effects is exactly the kind of critical reflection that leads to a failure to authentically promote virtuous dispositions. Because these dispositions must be pursued _unreflectively_, taking up a critical standpoint towards our intuitions and justifying them according to their genealogical merits will produce unsatisfactory results, since the process tends to give us excessively detached reasons for acting virtuously. By attempting to optimize the authority of our moral memes, we may

\(^2\)For further discussion on this point see (Williams 1985), especially pp. 7-10.
succeed only in undermining the legitimacy of those memes responsible for our personal relationships and virtuous character traits.\textsuperscript{23}

Critics of theoretical approaches to ethics have described the subversion of authenticity described here as a kind of "moral schizophrenia" or alienation from oneself and the projects that give meaning to our lives.\textsuperscript{24} The concern is that by imposing an impartial system of justification on our reasons for acting morally, theory-driven approaches to ethics inevitably drive a wedge between our reasons for performing our moral duties and the important personal motives that shape our broader ethical experience. To separate the justificatory basis for moral action from the features that make our lives worthwhile, it is argued, is to alienate ourselves from our life-projects and to dangerously bifurcate the integrity of our moral psychology.

Much of the criticism leveled against the alienating character of moral theory has been directed specifically at utilitarianism because of its unmistakably impartial maximizing structure. However, this criticism can be brought to bear against any systematic approach to ethics requiring that priority be given to impartial moral concerns. Whether it be utility maximization or the deontological rules of a Kantian framework, the obligations created by an impartial conception of what morality demands will routinely appear out of step with the integrity and personally rewarding aspects of our life projects. For the sake of simplicity, one can presuppose a consequentialist outlook operating in the examples that follow, but it should be noted that objections related to alienation apply equally to all modern ethical theories, despite the fact that consequentialism's provocative clarity makes it an obvious target.\textsuperscript{25}

\textsuperscript{23} Note that this dilemma is no less true if there is a substantial genetic contribution to the dispositions required for virtuous character traits, as opposed to such traits being composed entirely of aggregations of memes. Promoting these dispositions \textit{reflectively} will undermine their value as \textit{virtues} even if they are influenced by underlying biological factors that shape the structure of our primary character traits.\textsuperscript{24}

\textsuperscript{24}See (Stocker 1976) and (Williams 1973) respectively.\textsuperscript{25}

\textsuperscript{25}In fact, Michael Slote has recently noted that even virtue ethics cannot be pursued straightforwardly without ensuring its own failure. (Slote forthcoming) This is a highly ironic result considering that most of
Setting aside, for now, the question of whether impartial moral theories have resources available to overcome the objection that they are inherently alienating\textsuperscript{26}, it is worth investigating the sense in which this objection correctly identifies something defective about pursuing ethical goods from a critically reflective standpoint. There are at least three ways in which promoting ethical goods such as virtuous traits or personal relationships (I will use the latter from this point on for the sake of brevity) from an explicitly critical perspective can lead to a failure to adequately capture the value of these goods. First, if one attempts to promote the goods sanctioned by an impartial perspective to the best of one's ability, the result may be that one is alienated from one's life projects because one has little or no time and energy left for genuine personal relationships after the multitude of other, seemingly more important, impartial obligations have been secured. For instance, if Joe is choosing between two competing intuitions - one being that he ought to plan a special event for a close friend, the other being that he ought to spend the time helping less fortunate strangers - an impartial critical analysis will likely determine that the inner voice telling him to plan the event for his friend is a relatively unimportant side-effect of his (otherwise valuable) friendship and that he ought instead to spend the time helping more needy strangers. The logic of the decision seems difficult to deny, but if Joe's sense of critical reflection leads him to draw this conclusion at every point at which he faces such a decision, Joe's friendships will surely wither away and leave him alienated from an important part of his life. Moreover, Joe will probably be much less caring and less motivated to help others without the emotional support of these friendships. Thus, not only will he feel alienated, but he will be less effective at promoting his ethical goals than if critical reflection had not directed him to pursue the best agent-neutral choices whenever possible.

\textsuperscript{26}This question will be addressed in chapter six. However, for further discussion on this point see (Railton 1984), (Kapur 1991), (Cocking and Oakley 1995) and (Mason 1998).
The second way in which critics have accurately noted that moral theory can undermine our personal projects is that even if critical reflection leaves us with enough space to engage in personal relationships, these relationships will be conditional on the contingent fact that they produce enough agent-neutral value to be justifiable. If, for example, Wayne's sense of critical reflection attributes enough agent-neutral value to his friendships to keep himself from constantly diverting his attention from their importance, his friendships will nonetheless be potentially overridable by the possibility of encountering more pressing agent-neutral concerns. This may be alienating to both Wayne and his friends, since the threat that critical reflection will reveal some unforeseen moral obligation and cause him to drop his commitments to his relationships seems to undermine the trust and security required for authentic instances of friendship and other life projects. Moreover, even if the threat itself is not alienating (the idea that friendship must be completely immune to being overridden by less personal moral duties is a romantic fiction), the fact that the threat arises because one's relationships are conceived and evaluated from an impartial perspective seems to taint the integrity of what it means to be committed to another person. By having to conceptualize the value and limits of one's relationships from a critical standpoint, this perspective seems to trivialize the highly personal nature of these important ethical goods.

Which brings us to the third way in which critical reflection leads to a failure to appropriately value goods such as personal relationships. In this case, an argument can be made that no matter how remote the chance that one will have to give up certain aspects of one's relationships, the process of justifying these relationships from an impartial perspective is itself inherently alienating. Thus, Paul, for example, may never see the day where he would have to sacrifice even the smallest aspect of his friendships. Nonetheless, critics of moral theory argue that the mere fact that such friendships are pursued under a systematic framework unavoidably undermines the integrity and authenticity of his friendships. For critics like Michael Stocker, the problem arises
because *any* separation between our motives for acting directly out of concern for our friends and our reasons for acting in the interest of morality results in an unacceptable lack of harmony within our moral psychology. Similarly, Bernard Williams interprets this lack of harmony as one related to the agent being the prime motivator of her moral choices. In cases such as Paul's, Williams states that

... *once he is prepared to look at it [in terms of utility], the argument in any serious case is over anyway.* The point is that he is identified with his actions as flowing from projects and attitudes which in some cases he takes seriously at the deepest level, as what his life is about. It is absurd to demand of such a man, when the sums come in from the utility network which the projects of others have in part determined, that he should just step aside from his own project and decision and acknowledge the decision which utilitarian calculation requires. *It is to alienate him in a real sense from his actions and the source of his action in his own convictions.*

(Williams 1972, p. 116; my emphasis)

Though Williams objects to the possibility that utilitarianism can force agents to step aside from their personal projects, he is also arguing here that no matter how the numbers work out there is something alienating about moral theory giving agents separate reasons for action other than their own convictions. To do so is, for Williams, to deprive agents of acting in a fully autonomous manner.

While we may not go as far as Williams in believing that morality can rob us of our individual autonomy, there is a legitimate point here that if we receive our reasons for acting from an impartial ethical source, then these reasons no longer seem to be the *right kind of reasons* for achieving ethical goods as personal as our relationships and virtuous characteristics. A critically reflective standpoint is therefore ill-suited to providing us with an appropriate rationale for promoting certain moral goods, despite the fact that this perspective is clearly capable of both appreciating the agent-neutral value of these goods and recommending their promotion.

What we find in these three pitfalls for critical reflection is that directly pursuing our virtues and personal relationships out of respect for their ethical value is curiously
self-defeating. In fact, acting out of an awareness of the merits of these dispositions is individually self-defeating, whereas gaining an understanding of the merits of our social norms proved to be only collectively self-defeating. In the latter case, we can appeal to further ethical principles in order to overcome the suboptimal social outcomes produced by individual self-interest. In the case of virtuous dispositions, however, the self-defeating quality of acting with an explicit awareness of the value of these dispositions occurs within the limits of our own moral psychology. Thus, no further public ethical principles can help to alleviate the problem. (In fact, adding more ethical principles will likely make matters worse!) Straightforward implementations of what is recommended by critical reflection will therefore unavoidably fail to effectively promote those qualities that must be unreflectively pursued. In the words of Michael Stocker, "to the extent that you live the theory directly, to that extent you will fail to achieve its goods". (1976, p. 461)

The methodology of sifting through the genealogy of our moral intuitions in order to systematize their prescriptive authority is therefore at high risk for undermining the authenticity of memes that must be pursued in an unreflective manner. If we attempt to directly promote these memes after investigating their cultural history and precisely outlining the instrumental reasons for their justification, we will surely alienate ourselves from the special qualities that make these memes so uniquely valuable. In doing so, we end up foolishly outsmarting ourselves by unleashing the corrosive possibilities that follow from attempting to justify each individual component of our moral experience. Even if we proceed with the best intentions and flawlessly reveal the cultural forces that make our dispositions for virtue and personal relationships so important, we will find that we cannot directly promote these dispositions from a critical standpoint without undermining the success of our own efforts.

---

27 Again, this terminology is drawn from (Parfit 1984).
Conclusion

In this chapter we have seen a serious setback introduced for the aim of using memetics to refine the authority we ascribe to our moral intuitions. The setback revolves around the fact that attempting to account for the instrumental value of our intuitions from a critically reflective standpoint invariably leads to the prescriptive diminishment of many justified moral memes. I have outlined two different ways in which this can occur. First, the instrumental means by which our moral intuitions promote our best interests are in many cases exceedingly complex, and we will often fail to recognize the ethical merits of these intuitions either because temptation keeps us from doing so or because we are genuinely unable to grasp the subtlety at work in their functioning. The result is a burden of proof problem where every error in judgment or weakness of will leads to a failure to appreciate the legitimacy of perfectly acceptable intuitions. Second, certain virtuous dispositions and character traits must be pursued unreflectively for their ethical value to be realized, and this fact ensures that seeking to justify their authority from a reflective perspective will necessarily lead to self-defeating results.

Considering these potentially corrosive results, it is likely no accident that cultural evolution has programmed us to internalize moral shortcuts in order to bypass the pitfalls of excessive critical reflection. Adopting proximate moral intuitions as indirect indicators of our ethical goals allows us to save time and information costs within complex cultural environments, but it also allows us to avoid the trappings of directly implementing the recommendations of a critically reflective perspective. By promoting our reliance on tangible heuristics instead of relentless critical scrutiny, cultural selection has artfully given us the means to avoid sliding into the depths of moral minimalism and unraveling our virtuous social dispositions. Thus, as promising as the methodology of uncovering the cultural genealogy of our moral memes might appear, it may be the case
that cultural evolution has obscured the underlying justifications for our deeply internalized ethical shortcuts for a very good reason.
Chapter 6

Striking A Balance: The Virtues of Reflective Instability

Introduction

It should now be clear that a memetic approach to ethics reveals a sharp dilemma for the task of prescriptively justifying our intuitions. The dilemma is revealed by the arguments of chapters four and five, which suggest opposing courses of action when it comes to the possibility of scrutinizing the authority that our moral intuitions ordinarily possess. Chapter four suggests examining the cultural genealogy of these intuitions in order to free ourselves from instances where our moral memes are distorted by anomalous selection forces (e.g. a cultural runaway process). It recommends that we appeal to the empirical information supplied by memetics as an aid to critical reflection in order to more effectively systematize the authority of our moral intuitions. However, chapter five points out that using critical reflection in this way has an unavoidably corrosive effect on many virtuous character traits and internalized social norms that legitimately serve our best interests. Hence, chapter five recommends that we ought to listen carefully to critics warning that the aim of explicitly systematizing ethics leads to unfavorable outcomes. In the end, a balanced analysis of the ethical implications of memetics simultaneously advocates and condemns the aim of appealing to theories of cultural transmission in order to better legislate the prescriptive authority that we assign to our moral judgments.

What are we to do? Memetics leaves us with both the realization that our moral instincts are imperfectly designed and the knowledge that attempting to correct their deficiencies invariably leads to a diminishment of many permissible moral principles. We are caught between the Scylla of leaving ourselves at the mercy of defective cultural selection forces and the Charybdis of degrading the prescriptive influence of plainly
justifiable moral memes. Both options present serious problems, but it is unclear how we can resolve this dilemma without succumbing to the disadvantages of either pitfall.

I believe there is no easy solution to this problem. In this chapter, however, I propose that the best possible compromise between studying the genealogy of our moral memes and inadvertently undermining their prescriptive authority is a delicate and unstable psychological balance that ethical theory is better off facing directly than it is tacitly dismissing. It is easy to write off the tension between critical reflection and our spontaneous adherence to moral precepts as a general feature of ethical decision-making - one that is enough of a platitude to be swept under the rug of meaningful discussion. The purpose of this chapter is to instead bring this feature out into the open. Specifically, I argue that the balance we must achieve to avoid excessively scrutinizing our moral intuitions (without placing too much faith in their ability to directly guide our behaviour) is a balance that is of a unique and particular type. We must not only maintain a balance between the competing recommendations of critical reflection and our spontaneous performance of certain moral actions. We must also routinely, but not permanently, insulate ourselves from the logic of critical reflection if we intend to maintain control over our ethical judgments and still successfully capture the value of actions that must be pursued for their own sake.

This peculiar type of balance is worth examining because it has unusual consequences for the way in which we conceive of our moral psychology. The unusual consequences in question are hotly controversial, since the process of insulating ourselves from critical reflection accentuates the division between justificatory reasons for action and our direct personal motives - a division we have seen diagnosed as a form of moral schizophrenia in chapter five. However, the conclusion drawn in the present chapter is that this supposedly pathological condition is inevitable if we want to reconcile the competing recommendations of chapters four and five in the least injurious way. The implementation of this rather odd compromise may seem awkward, but if we are willing
to take the ethical ramifications of memetics seriously we should not fool ourselves into thinking that any less complicated solution is an acceptable option.

The chapter is organized in three sections. Section one outlines three possible means of resolving the tension between chapters four and five. Only one of these options is presented as being viable, and in section two this option is explored in further detail. The aim of section two is therefore to show that the special balance required between critical reflection and our ability to act on insulated moral motives can be implemented as a precarious but plausible response to the choice between ethical runaway effects and the prescriptive dissolution of our intuitions. Finally, section three describes why the unstable compromise from section two is a predictable evolutionary response to complex moral landscapes. Moreover, this last section looks at evidence drawn from philosophers who seem to suspect that this instability in our moral psychology is necessary, even if they fall short of explicitly embracing the unusual form of self-deception required to insulate ourselves from the ability to critically reflect on our moral memes.

I. Three Options

There are perhaps many potential solutions to the dilemma created by memetics' conflicting recommendations for ethics (hereafter referred to as the 'genealogy dilemma', since the dilemma pertains to whether or not we ought to continue to investigate the genealogy of our moral memes). But these potential solutions can, I think, be classified within three general categories. As we shall see, only one of these categories proves to be viable. However, it is worth outlining all three options in order to gain some perspective on the scope and structure of the dilemma we seek to resolve.
A Ruling Elite

The first category of solutions is made up of responses that solve the genealogy dilemma by assigning the conflicting responsibilities of critical reflection and our spontaneous obedience to moral principles to different groups within a general population. The insulation required to prevent the corrosive effects of critical reflection is, in this case, achieved by allowing only a select few to critically evaluate the genealogy of our intuitions, leaving everyone else to straightforwardly follow the prescribed standards of their community. By splitting the suggestions of chapters four and five between separate groups of individuals, this option allows for the overall population to retain the capacity to investigate the legitimacy of its ethical principles without contributing to their prescriptive diminishment.

The most notable example of this solution can be found in The Republic, where Plato argues that the ideal state, the kallipolis, must be divided into three distinct classes of citizens. In this ideal state, soldiers and workers (the latter being composed of a variety of vocations, including farmers, craftsmen and businessmen) are governed by a ruling class of philosophers, who are uniquely equipped to govern because of their superior understanding of reason and the nature of the good. For the purpose of our inquiry, the interesting aspect of Plato's conception of the kallipolis is the considerable emphasis he places on the importance of strictly separating the role of the ruling philosophers from those of the other citizens. In fact, Plato equates the esteemed virtue of justice with the type of "moderation" or "harmony" required for each class to recognize its own proper role and to refrain from meddling in the affairs of others. (430d-434c)

The rigid, highly specific roles that the three classes fulfill are revealing, for they suggest

---

1 See (Plato 1992). The initial distinction between the three parts of the ideal city-state occurs in Book II, but the more interesting discussion of how these different parts of the state interact comes in Book IV. (Note that, as we shall see later in the chapter, these parts are also meant to represent the three basic elements of the soul, since the introduction of the kallipolis in Republic serves as a means of investigating the various constituents of the perfectly just soul at a larger scale)
that Plato is acutely aware of the need to cleanly differentiate the competing aspects of ethical decision-making.

Thus, in the *kallipolis*, the role attributed to philosophers is not kept distinct because of the scarcity of these specially educated individuals (or because Plato necessarily wants to protect their elite status). The role of the ruling philosophers is kept carefully distinct because Plato believes each class is uniquely designed to carry out its own specific function, and that mixing the predetermined duties of these classes will result in calamitous results. (434a-c) Plato is, of course, especially sympathetic to the role of the ruling philosophers. However, he is also aware of the fact that individuals so disposed to critical analysis will likely make a mess of what is required from the "appetitive" and "spirited" functions of the city-state (the functions that supply the state with its immediate desires and courage). Philosophers, it seems, make poor workers and soldiers. Similarly, soldiers make poor philosophers and poor workers, and workers make poor soldiers and poor philosophers. This basic incompatibility between the special skills required to perform each role within the state drives Plato to advocate an intriguing public balance where these classes operate in a highly integrated but *individually distinct* form of harmony. (439d-442d) While we may no longer agree with Plato's assumptions about the predetermined roles of social classes, his vision for the *kallipolis* and its separate internal functions provides an interesting means of retaining control over the community's ethical standards without allowing critical reflection to erode the widespread authority that these standards ought to possess.

Intriguing as it is, however, this proposal, and the general option of assigning the responsibility of critical reflection to a separate ruling elite, can be dismissed fairly quickly. The first and most obvious objection to this option is our overwhelming (and wholly understandable) political aversion to placing our faith in a distinct ruling class. Despite the *Republic's* charm, most of us believe that genuinely benevolent oligarchies exist only in thought experiments like the idealized *kallipolis*. The idea of actually
granting a group of citizens the power to determine the legitimacy of our ethical principles while insulating the rest of us from this responsibility seems a thoroughly unwise option. The prospect of placing ourselves at the mercy of a privileged few who will almost certainly not be as virtuous as Plato's philosopher-kings is therefore unappealing enough to render this solution to the genealogy dilemma virtually irredeemable.²

Moreover, even if we set aside our reservations about the integrity of a ruling elite, a second argument can be brought against publicly dividing the responsibilities of ethical deliberation. This second objection is a pragmatic concern: if we consider the capacity for critical analysis that exists within each citizen, cutting off a substantial portion of the community from critical reflection is simply not feasible. Even if we imagine a perfectly benevolent class of rulers who take responsibility for evaluating the legitimacy of ethical issues like the genealogy of our intuitions, it is unrealistic to think that the rest of society will be capable of successfully cutting themselves off from seeking to evaluate the legitimacy of our intuitions. Once we drop the assumption that a ruling class is naturally disposed to critical reflection whereas the rest of society is disposed to (and satisfied with) their own subordinate roles, the prospects for successfully insulating the greater population from engaging in critical analysis are slim to none.

Humans are intensely curious, reason-seeking beings. We sometimes act spontaneously according to prior social norms; however, we also routinely seek to discover the cultural origins of these norms and question the grounds for their justification. It is therefore unlikely that the vast majority of a population can be guided by norms powerful enough

²Similar condemnations of this option occur in the form of criticizing the possibility of so-called 'Government-House utilitarianism', which assigns an elite group the task of deciding on the rules that most effectively maximize happiness so that they can provide these rules to (and presumably enforce them on) the rest of society. See (Sidgwick 1893, pp. 487-490), who finds the idea unappealing but does not entirely rule it out, (Williams 1973, pp. 135-140), who obviously rejects the idea, (Parfit 1984, pp. 40-43), who hedges by entertaining the idea while claiming that the few who are not picked out for self-deception about the aims of morality "need not live in Government House, or have any other special status" (p. 42), and (Williams 1985, pp. 108-109) who gives the idea its current title and expediently rejects it a second time.
to suppress our tendency to question authority and the ethical principles of our community. Without positing a dubious class-based teleology (or implementing a politically intolerable means of enforcement), the harmony required to successfully separate the competing suggestions of chapters four and five among separate public groups is not only politically unfavorable, it is highly impractical as well.

_Bury our Heads in the Sand_

The second potential method of resolving the genealogy dilemma is to take the view that the hazards of excessive critical reflection are dramatically worse than those associated with defective memetic selection forces. This option dissolves the stalemate between the recommendations of chapters four and five by maintaining that the choice between optimizing our existing moral intuitions and risking their diminishment is really no choice at all. On this view, the possibility that our moral instincts and legitimate ethical principles will be jeopardized by investigating the details of their cultural history provides us with only one clear course of action: *stop investigating the cultural history of these instincts and principles*! In doing so we allow ourselves to fall victim to the distortions that arise because of imperfect cultural selection mechanisms. But if we refrain from opening the can of worms that critical reflection indiscriminately unleashes for the authority of our intuitions, this method of responding to the genealogy dilemma insists that we will nonetheless be left far better off.

There are two possible formulations of this option, one strong and one weak. The strong formulation claims that the consequences associated with using memetic genealogy as an aid to critical reflection are so pernicious that we ought to fully abandon

---

3 One does see situations like this occur in close-knit religious groups, where difficult moral decisions are referred to an elite group of elders for adjudication. In fact, the same structure can perhaps be seen even in larger organizations, such as the Catholic church. However, in cases like Catholicism where vast numbers of community members are involved, the fact that there is an extremely broad spectrum of compliance with the prescriptions of the Vatican demonstrates just how difficult it is to maintain a level of spontaneous conformity in large populations where members are meant to be insulated from making critical judgments about their religious beliefs.
the project of using this information to systematize our moral intuitions. Like some unstable scientific process that creates radioactive waste, the aim of scrutinizing the evolutionary merits of our moral intuitions is, according to this view, a methodology that ought to be renounced in order to protect us from its imperspicuous but highly injurious by-products. It would of course be unhelpful to give up completely on the general aim of systematizing our intuitions, for we cannot abandon the basic tenets of logical consistency without producing even more disastrous results. However, the stronger version of giving up on examining the merits of our moral memes argues that above a certain critical threshold we should repudiate the attempt to optimize the authority of our intuitions and desist from actively seeking out instances of prejudice and distortion in our ethical judgments.

The claim, then, is that memetics itself gives us sufficient reason to keep ourselves from appealing to the information it supplies as a source of ethical guidance. By documenting the corrosive effects of implementing its suggested revisions for our moral intuitions, the methodology of appealing to memetics in order to further our ethical interests ends up suggesting, oddly enough, that we do not attempt to follow its direct recommendations. According to this strong formulation of our second option, the aim of using memetics to advance our ethical objectives is self-effacing, because it ultimately recommends that we ignore the information memetics provides if we want to best achieve our original objectives. The result, then, is a curious defense of common-sense morality. If we believe that the dangers associated with allowing ourselves to reflect on our intuitions are prohibitively harmful, then the most prudent course of action is to cut ourselves off from critical reflection and to take our chances with our moral intuitions as they stand. (Better to face the devil we know than to risk letting loose a corrosive mechanism with no discernible limit foreseen in terms of the damage it might do.) The strong formulation of this response to the genealogy dilemma thus interprets the goal of
optimizing the authority of our ethical memes as a self-effacing enterprise that succeeds
only in paving a circuitous route to endorsing our existing intuitions.

To its credit, this option (if successful) certainly manages to solve the genealogy
dilemma. Whether one accepts the permissibility of self-effacing strategies\(^4\) or believes
that having to resort to these measures constitutes a reductio of the idea, the claim that
critical reflection generates harmful enough effects to require self-effacement suggests
the same clear course of action: ignore the potential benefits of chapter four and embrace
common-sense morality without disputing the finer details of its legitimacy.

Decisive as this option is, I think it is severely misguided. First, although it is
difficult to quantitatively prove the point one way or the other, I think the claim that the
warnings of chapter five grossly outweigh those of chapter four underestimates the harm
that can be done by distortions like ethical runaway processes. Allowing ourselves to be
trapped into internalizing memes that propagate because of exaggerated selection forces
leads to very serious errors in judgment about what we consider prescriptively justified.
(If this point needs reinforcement, ask a member of the gay community how they feel
about the meme that 'what is natural is necessarily virtuous' and how this meme affects
their struggle for equal moral consideration.) Thus, it strikes me as entirely too
complacent to write off these distortions as occupational hazards for the larger design of
ethics. The anomalies that defective evolutionary forces create within our ethical
thinking result in chronic prescriptive distortions for very important issues, and they
should not be brushed aside as isolated examples or 'tough cases' that ethics need not
seriously address.

Furthermore, even if one remains agnostic about the possibility that these
distortions are inconsequential compared to the sweeping damage unleashed by critically
examining our intuitions, the option of permanently cutting ourselves off from a source of

\(^4\)For a defense of the claim that self-effacing theories are peculiar but nonetheless coherent, see Parfit
knowledge about our moral thinking is presumably a very unwise decision. If we carry out this objective, we risk the possibility that circumstances will arise in the future where an understanding of memetics could prevent results more harmful than we can currently imagine. If so, we will have foolishly purged ourselves of the ability to defend against these more serious circumstances before completely understanding the extent of their gravity. Without knowing just how serious these future circumstances might be, it seems profoundly short-sighted to give up on investigating their significance. Thus, as a simple matter of prudence I think it is clearly a bad idea to permanently sever the cord to critical reflection and risk irretrievably losing a resource that could prove to be beneficial in ways we cannot yet fully conceive.

The weaker formulation of this option, however, may be able to avoid the short-sightedness of its overzealous sibling. This formulation claims that it is unwise to give up entirely on the need for some mechanism to act as a critical check on our commonsense intuitions. It proposes instead that we ought to specifically give up on using systematic implementations of critical reflection to examine our intuitions, because theory-driven attempts to revise common-sense morality (like systematically evaluating our intuitions on a case by case basis according to their cultural history) will inevitably draw the ruinous consequences outlined in chapter five, whereas less formal methods of revising our ethical beliefs will not. Hence, this weaker formulation solves the genealogy dilemma by claiming that theoretically structured methods of systematizing our intuitions will be self-effacing, but that alternative methods of reforming our intuitions are available that can fulfill the role required of critical reflection without incurring its corrosive side-effects.

Moreover, this option looks to be an unrealistic decision as well. It is doubtful that we could permanently cut ourselves off from pursuing the study of memetics and its impact on ethics for much the same reason that it is impractical to think that we could insulate everyone but a ruling elite from seeking this information. We are intensely curious creatures. If important information about ourselves is available, it will be extremely difficult to keep ourselves from investigating its significance no matter how many 'flat-Earth' style obstacles we create to prevent this from happening. Realistically, then, we are probably better off confronting the implications of critical reflection and memetics than trying to bury the project entirely.
What exactly will these alternative methods look like? The answer is best articulated, perhaps, by the recent move away from formal theory building in contemporary ethics. According to this movement, our ethical aims are not promoted most effectively by extensive theoretical analysis or the construction of fixed decision-making procedures for determining the rightness of our actions. Instead, the claim is that ethics ought not be formalized in this way, and that it ought to instead be pursued through more personal, less axiomatic channels.6 These channels usually involve broad social virtues and the pursuit of admirable character dispositions that enable us to recognize the appropriateness of our actions from the details of our particular circumstances. For our purposes, the important feature of this general approach is the fact that the mechanisms proposed to keep a critical eye on the legitimacy of our moral judgments are assumed to be emergent social functions that guide our collective attitudes in holistic but highly indistinct ways. These mechanisms are presumed to be sufficiently effective to guide our intuitions in appropriate directions, but they are nonetheless described as being irreducible to any predetermined heuristic blueprint. Thus, they are thought to be the natural checks and balances that spontaneously emerge from human social interaction, and these properties, it is claimed, cannot be systematized by any concrete theoretical formula. Bernard Williams famously describes the process as follows:

How truthfulness to an existing self or society is to be combined with reflection, self-understanding, and criticism is a question that philosophy, itself, cannot answer. It is the kind of question that has to be answered through reflective living. The answer has to be discovered, or established, as the result of a process, personal and social, which essentially cannot formulate the answer in advance, except in an unspecific way.

(Williams 1985, p. 200; my emphasis)

Thus, the movement against theory in ethics suggests that mechanisms exist within our social systems that serve to critically reflect on our intuitions and prevent gross

---

6For examples of this broad approach to ethics, see (MacIntyre 1984), (Taylor 1982, 1989), (Foot 1985), (Baier 1985, 1986) and (Williams 1985). Not coincidentally, many of these writers are the same individuals that were identified as opponents to instrumentalist conceptions of value in chapter five.
distortions from occurring within our judgments. The important difference being that these mechanisms can be understood as theoretically unstructured phenomena that naturally surface from the interaction of our broader ethical concepts.

This trend in ethics introduces a convenient response to the genealogy dilemma, for it suggests that less formal methods of promoting justified ethical intuitions and principles will succeed where systematic means of scrutinizing the authority of our intuitions ultimately fail. Through means such as 'reflective living', viewing virtue in terms of 'narrative unity' or perhaps placing our faith in a 'fusion of horizons', those who support less theoretical approaches to ethics defend the possibility that cultural distortions can be avoided via a collective capacity to critically reevaluate our beliefs without falling prey to the corrosive pitfalls of systematic critical reflection.

Unfortunately, this weaker formulation of the option of renouncing critical reflection is far too convenient. After studying the undirected character of cultural evolution, there is simply no reason to believe that unsystematized social mechanisms will be any less immune to distortion than our unexamined common-sense intuitions. The emergent properties of what it is to successfully engage in "reflective living" will still be subject to the undependable forces of cultural evolution. Thus, unless these properties involve explicit methods of inspection and critique within their (intangible) conception of critical reflection, they will still be liable to produce distorted ethical beliefs and judgments. The fact that they are described as the spontaneous products of social interaction does not change the fact that we have no guarantee that such forces will promote our best interests. The price of this kind of guarantee is exactly the slew of unfortunate consequences that follow from systematically analyzing the merits of our intuitions. Sadly, we cannot have our cake and eat it too. Either we give up on meaningful critical reflection and place our faith in the results generated by undirected social processes, or we take our chances with the pitfalls of carefully investigating the merits of what we reflectively justify. To reject this disjunction and praise some abstract
third option that necessarily leads to our collective advantage is to indulge in an unhelpful form of unsubstantiated optimism.

This weaker formulation of our second option therefore fails because it falls prey to one of the following three possible interpretations: a) the social mechanisms in place that direct our collective ethical beliefs toward appropriate conclusions are *truly unsystematic*, in which case there is no reason to think they will be effective because they will be just as vulnerable to unperceived cultural selection forces and therefore subject to distortion (meaning that this variant reduces to the strong formulation of giving up on critical reflection discredited above); b) the social mechanisms in place are *systematic enough to function effectively*, despite their opaque appearance, in which case they will inevitably produce the same unwanted by-products described in chapter five; or c) the social mechanisms in place are unsystematic but nonetheless effective because they rely upon a *teleological conception of human development* which ensures that spontaneous social interaction leads to beneficial results - a pleasant thought, but one that relies on an implausible empirical premise.

It is this third interpretation that is perhaps most likely to attract support. Yet if its underlying assumption is exposed and articulated, I think it is clear that this option is not a realistic possibility for those of us who no longer believe in antiquated conceptions of human development and emergent social functions. Moreover, the decisive knowledge that cultural evolution operates according to imperfect transmission mechanisms makes the aim of updating this solution to allow it operate within a naturalistic framework a hopelessly lost cause. Once our faith in some higher evolutionary power is extinguished by the existence of anomalous selection mechanisms, there is no plausible way to avoid the fact that the weaker formulation of giving up on critical reflection falls into one of the three unfeasible interpretations outlined above. Appealing to seductively obscure accounts of how our interests are promoted by holistic, pre-theoretical reflective mechanisms only unfairly plays on our unspoken hopes about social progress and
provides us with a false sense of comfort which distracts us from facing an unavoidably difficult problem.

A Delicate Internal Balance

With the previous two options eliminated, only one remaining alternative provides a potential response to the genealogy dilemma. This option involves achieving the necessary balance between critical reflection and our immediate ethical motives by maintaining a delicate equilibrium within the psychology of each individual agent. Rather than splitting the conflicting recommendations issued by memetics between separate groups of individuals, as we saw in the first option described above, this option situates the competing suggestions of chapters four and five within the independent minds of all moral agents. This amendment removes the ominous political problems associated with assigning the task of critical reflection to a ruling elite, yet it nonetheless manages to ensure that the critical reevaluation of our existing intuitions and the protection of our capacity for spontaneous moral action are both carefully preserved.

Moreover, this amendment to our first option mirrors the fact that Plato's presentation of the kallipolis in Republic serves as a metaphor for the internal balance that exists in the ideally just soul.\(^7\) The ruling class of philosophers that govern the city-state serves as a rhetorical tool designed to represent the sense of personal wisdom required to guide us towards virtuous action. Similarly, the guardians and working class in the kallipolis are intended as representations of individual courage and our immediate motivating passions respectively. Keeping these three separate aspects of the soul in a balanced state of harmony is the key feature that distinguishes a virtuous soul from one

\(^7\)In fact, it may be argued that the public implementation of the kallipolis is meant only as a thought experiment designed to elucidate the details of the soul, and that it therefore ought not be taken seriously as a political proposal at all. While this may be so, the fact that Plato clearly intends this metaphor to be convincing at the level of the kallipolis (in order to prove his point about the soul) demonstrates that he believes that the political measures of the kallipolis must be at least prima facie plausible.
pathologically influenced by excessive proportions of any one of its mutually indispensable components.

Considered in the abstract, this option presents an exceedingly sensible resolution to the genealogy dilemma. If each person can maintain the necessary balance between scrutinizing the merits of her intuitions while internalizing those that must be pursued unreflectively in order to function properly, then we can secure a means of protecting ourselves against anomalous cultural selection forces without placing the remainder of our legitimate ethical principles at risk of dissolution. In effect, we can allow our cultural programming to function effectively without giving up the capacity to keep this programming on a leash, so to speak, by retaining the ability to reevaluate the prescriptive status of its guidance. If this option is psychologically possible, it combines the best of both worlds and gives us an efficient solution to the genealogy dilemma that is limited only by our ability to maintain a virtuous internal balance within our moral psyches.

But is this option psychologically possible? At the very least, maintaining an equilibrium between these two divergent ethical objectives is a significant challenge. To successfully resolve the tension between chapters four and five, this psychological balance must be able to tell us when to shift into a critical perspective towards our intuitions and when we ought to ignore this urge and unreflectively follow their immediate advice. The option therefore relies on both the possibility that we can operate on at least two separate levels of moral thinking as well as the assumption that we can achieve an ongoing psychological arrangement that will notify us when we ought to take up each one of these two separate levels. This will by no means be an easy assignment. We may be able to appeal to certain social institutions in order to buffer the difficulty imposed by this challenge. (For example, we are cued to critically assess the legitimacy of our ethical principles in certain predetermined social contexts, such as political debates or community meetings. In these social contexts we are given clear indications that we
are expected to take up a critical perspective while taking part in these events.) However, these measures cannot alleviate the problem entirely. At some point we will have to find a way to strike a precarious balance within the moral psychology of each individual, and we will have to learn to coordinate the continuous shifting of this balance as circumstances demand that we adopt different levels of moral thinking.

Despite the obstacles facing this third option, I think it is the most viable means of reaching an acceptable compromise between the conflicting directions suggested by an investigation into the ethical implications of memetics. Once we recognize the tension between investigating the genealogy of our intuitions and letting these memes directly guide our behaviour, it is only by locating the attempt to resolve this tension within each person's psychological capacities that we can avoid resorting to objectionable political measures without giving up on the project of minimizing the repercussions associated with both sides of the genealogy dilemma. However, the obstacles facing this option are very real. By claiming that it is the most promising means of tackling the issue of reconciling a critical perspective with our common-sense moral memes, I am certainly not suggesting that the potential defects of this project are inconsequential. In fact, as we shall see in the following section, I think the difficulties associated with implementing this project are of an important and peculiar type.

II. Triggers, Anchors and Moral Alarm Clocks

It is important to realize that the conflicting results of chapters four and five are not unrelated. The problem of ethical runaway effects and the subsequent need to inspect the genealogy of our moral intuitions arise because the memes we internalize as shortcuts to more complicated moral objectives can, in some cases, fall prey to exaggerated feedback loops - loops that distort the beneficial quality of these shortcuts if they are internalized deeply enough to remain hidden from critical reevaluation. Conversely, the corrosive effects of inspecting the genealogy of our intuitions arise when we tamper with cultural
shortcuts that are useful \textit{precisely because} they enable us to internalize moral dispositions for unreflective action - dispositions whose insulated quality enables us to avoid the self-defeating outcomes that plague our straightforward attempts to pursue virtuous moral aims.

These two phenomena are therefore not incompatible by coincidence. They are directly antithetical to each other, for they both revolve around (opposite poles of) the unique feature of \textit{insulating} our moral dispositions from critical scrutiny. It is this feature that provides us with both an escape from self-defeating outcomes and a glaring vulnerability to the creation of cultural runaway effects. Both sides of the genealogy dilemma thus share a common ancestry, meaning that distortions in our moral intuitions occur in exactly the same structures that we purposely insulate in order to prevent the diminishment of our advantageously unreflective moral aims.

This common ancestry is perhaps not surprising, since it makes sense for the phenomena described in chapters four and five to be products of a single evolutionary tendency: the propensity to solve complex computational problems with expediently imperfect solutions. When it comes to questions of design, natural selection is the undisputed champion of using \textit{satisficing} methods to overcome complicated algorithmic challenges. (The fact that it is a champion without foresight being all the more galling...) As we saw in chapter four, the term 'satisficing' is used to denote a decision-making strategy that searches for options until it reaches a satisfactory - but not necessarily optimal - result, at which point it then settles for this proximate solution. Because evolution works blindly and extremely slowly, it creates similar short-term solutions to immediate environmental pressures and then piles these solutions one on top of the other over vast stretches of time. The apparent design that emerges from natural selection is for this reason a prime example of satisficing at work, and it ought to come as no surprise that many elegant but less than ideal solutions are generated by cultural evolution through this uniquely imperfect process.
A Peculiar Balancing Act

Unfortunately, the fact that this satisficing process often involves the need to insulate proximate moral dispositions presents a very specific and very odd problem for our moral psychology. The problem arises because the cultural mechanisms responsible for the internalization of our moral dispositions are often not just examples of satisficing for the sake of saving time and information costs. They are, as we have seen in chapter five, examples of satisficing for the sake of protecting ourselves from excessive temptation and self-defeating situations. Thus, while it may come as no surprise that these traits have been generated by an evolutionary satisficing process, it is worth special attention that, as proximate goals, the traits recommended by satisficing in these cases are beneficial not just because of constraints related to limited information. The traits are adaptive, rather, because deviation from these proximate goals in the form of straightforward optimization produces worse results than if we continue to follow our indirect satisficing guidelines.

This difference is important because it explains why insulating our moral dispositions from critical reflection is so essential to their proper functioning. If a satisficing strategy generates superior results because direct optimization is for some reason deficient as a course of action, then the proximate goals of this satisficing strategy cannot be abandoned whenever critical reflection judges that these goals are suboptimal in light of new information (or what our immediate circumstances cause us to perceive as new information). They must instead be securely insulated from the logic of direct optimization and its self-defeating outcomes in order to serve as a reliable means of securing our long-term best interests.8

---

8 Note that this is not the case when satisficing is used merely as a means of making the best possible decisions under circumstances of limited time and information. In cases of this type, if new information arises that can lead to more efficient results there is no reason not to drop one's proximate satisficing goals and switch to a direct optimization strategy. Self-defeating situations are therefore unique in that new
This leads to an unfortunate implementation problem for our moral psychology. Insulating our proximate moral goals means that we have to make ourselves immune to the compelling reasons that lead us to directly pursue our moral ends. To allow our proximate goals to function, we have to intentionally cut ourselves off from being able to step back and reevaluate their ongoing merits. (Otherwise, we are right back into the problems outlined in chapter five.) This fact about what is required in order to insulate our moral dispositions creates an arduous problem for the prospect of reconciling both sides of the genealogy dilemma within a single, coherent pattern of thinking. Our third option from section one requires us to strike a delicate psychological balance between two competing recommendations, but the need to insulate unreflective moral aims tells us that when we try to achieve this balance we will be unable to recognize when it is appropriate to switch from one level of moral thinking to the other. This makes what is already a very delicate balance that much more difficult to execute. To prevent ourselves from getting trapped in self-defeating situations, we have to cut ourselves off from precisely the kind of deliberation that would enable us to recognize when we ought to switch from being directly motivated by our moral dispositions to being compelled to question the these dispositions from a critically reflective perspective.

This insulation problem suggests that the second and third options described in section one are closely related, because the psychological balance required to resolve the genealogy dilemma depends on critical reflection being partially self-effacing in order to capture the value of moral actions that must be performed unreflectively. Unlike our second option, however, the attempt to construct an ongoing psychological resolution to the genealogy dilemma requires that this self-effacing character be only partial. This makes our third option more plausible as a long term solution, but it leaves us with the tricky question of how to manage a 'partial' cutting of the critical reflection parachute.

Information may appear as if it legitimizes abandoning one's proximate goals when this will paradoxically bring about less profitable results.
How do we decide when to return to the fold of direct optimization once we internalize traits that are purposely insulated from allowing this restoration of critical thinking to occur?

What is required is some kind of subconscious mechanism that will alert us to the special instances where critical reflection is required - a mechanism that will otherwise remain dormant within the operation of our ordinary moral psychology. Of course, in reality this function need not be represented by any one, clearly delineated module. Speaking in terms of a single mechanism serves only as a convenient abbreviation for a more general ability to recognize the exceptional conditions under which one ought to 'switch over' to critical moral reasoning - an ability that may be fulfilled by more than just one psychological device. The point is that to successfully resolve the genealogy dilemma within our individual minds, we must each possess the capacity to be impelled to reevaluate our moral commitments in certain key situations without being aware from moment to moment of the reasoning that determines when this switch to critical reflection ought to occur. The reason a lack of awareness of this kind is required is that if we are aware of the justificatory criteria that control the timing of the switch to critical reflection, then our ordinary moral motivations will be tainted by considerations attached to our critical level of moral thinking - a result that risks the self-defeating outcomes that arise when virtuous actions are not sufficiently spontaneous.

*The Need for Reflective Instability*

So far, the difficulties surrounding psychological mechanisms of this kind have been discussed almost exclusively in the literature related to indirect forms of consequentialism. Because the direct implementation of consequentialism is dramatically inconsistent with the value of such ethical aims as spontaneously acting virtuously, maintaining close personal relationships and pursuing the "ground projects" that are
essential to a healthy moral life⁹, indirect forms of consequentialism have postulated various means of reconciling the ability to act on non-consequentialist goals with an underlying commitment to the guidance provided by consequentialist evaluation.

The two most prominent examples of this kind of reconciliation project are Peter Railton's "sophisticated consequentialism" and Philip Pettit's "virtual consequentialism".¹⁰ The former recommends that the motivational structures of moral agents be flexible enough to include various non-consequentialist projects, but that agents ought to still be guided by a broad 'counterfactual condition'. This condition ensures that even if an agent "ordinarily does not do what he does simply for the sake of doing what's right, he would seek to lead a different sort of life if he did not think his were morally defensible" (pp. 152-3). The result is a form of consequentialism that requires agents to periodically reevaluate their moral aims and dispositions from a critical perspective without requiring that agents be aware of the consequentialist criteria governing these reevaluations from moment to moment. Similarly, Pettit's articulation of virtual consequentialism proposes that the calculations required by critical reflection are only virtually present in agents' moral deliberations. This virtual presence allows agents to routinely operate without any direct reference to consequentialist calculations, yet it nonetheless ensures that they retain psychological "triggers" whose activation cause them to reflect in a consequentialist manner on their behaviour. (p. 15)

What Railton and Pettit's versions of consequentialism share is thus an understanding of both the fact that critical moral thinking cannot intrude into the ordinary performance of moral decision-making and the belief that critical thinking must nevertheless find a way to retain some form of distanced control over our actions in order to make sure that they remain sufficiently justifiable. The balance they seek is in this

---

⁹ See (Williams 1973).
¹⁰ See (Railton 1984) and (Pettit 1994) respectively. Railton also outlines a similar project in (Railton 1988) which he describes as "valoric utilitarianism" (p. 409).
way exactly parallel to the balance required to solve the genealogy dilemma. Although the corrosive form of critical reflection they favour is consequentialist calculation rather than a genealogical analysis, they are attempting to insulate our spontaneous moral intuitions from their own version of critical reflection in the same way that is required of a memetic approach to our moral intuitions.

Unfortunately, however, neither philosopher believes that reconciling critical reflection with our intuitions is an easy assignment. The strategy involves constructing ethical systems that are reflectively unstable in the sense that the multiple psychological levels they make use of are not stable under reflection from any one point of view - a characteristic that is sufficiently extraordinary to have drawn criticism from those, like Bernard Williams, who object to the awkward nature of the mental demands involved in such projects. Nevertheless, Railton and Pettit believe that a reconciliation of this kind is possible and that it is necessary if we want to construct a justifiable system of ethics that takes into account the motivational structures and long term ethical interests of moral creatures such as ourselves.

---

11 This criticism is made most explicit in (Williams 1981a) and (Williams 1988).
12 Note that there is a version of consequentialism that is conspicuously absent from this discussion of theories with two levels of moral thinking: the utilitarianism of R.M. Hare. Hare defends a form of utilitarianism that is deceptively similar to the positions outlined by Railton and Pettit because it also relies on a distinction between critical and intuitive levels of moral thinking embedded within the psychology of individual agents. However, while Hare views the intuitive level of thinking as both psychologically and practically necessary for creatures such as ourselves, he appears to view this fact as a necessary evil that arises because of our limitations as human agents - a necessary evil that ought to be avoided when, at any particular point in time, we can somehow manage to elevate ourselves to thinking like an "archangel" rather than a "prole". Hare therefore seems to be unaware of the dangers associated with critical thinking and its potential for driving us toward self-defeating outcomes. Instead, Hare's theory assumes that we can operate seamlessly on a sliding continuum between our two levels of thinking, and that the only relevant factor determining how close we ought to approach "archangel-like" thinking is the "powers of thought and character each one of us, for the time being, thinks he possesses". (Hare 1988, p. 289) This seems overly optimistic considering the paradoxes described in chapter five, and Williams, in his critique of Hare's theory, is I think right to point out that the two levels are not always stable under reflection. (I simply disagree with Williams that this fact necessarily constitutes an objection to two-level theories.) The difference between Railton/Pettit and Hare is important, because Hare's assumption that coordinating critical and intuitive moral thinking is a purely epistemic problem underestimates the corrosive effects associated with critical thinking and subsequently trivializes the importance of promoting insulated virtuous dispositions at the intuitive level. See (Hare 1984), especially chapters two and three.
Though the debate over two-level forms of indirect consequentialism captures the general tension between critical reflection and our spontaneous execution of virtuous actions, it is important to note that this tension is by no means restricted exclusively to consequentialist ethics. The fact that the basic aims of consequentialism are so especially clear and forthright makes it an easy target for criticism when it comes to deciding how two levels of moral thinking can be reconciled within a single psychology. However, balancing the task of reevaluating our moral intuitions with the corrosive effects this process invites is a problem that all ethical theories must come to grips with, even if it is a problem that is commonly ignored in many contemporary ethical frameworks.

The reason the problem will not go away is simple: ethical aims that must be unreflectively pursued will be undermined by critical reflection of any kind, not just reflection that seems particularly alienating because it is concerned with calculating best possible outcomes. The value of performing virtuous actions is tenuously dependent on these actions being pursued for their own sake. The admirable quality they possess is thus exceedingly fragile when our motivations are potentially influenced by critical reflection. As chapter five demonstrated, one cannot be reflectively aware of the justificatory reasons for engaging in virtuous behaviour without running the risk of tainting one's motivations and stripping the behavior of its moral value. To do so is to invite self-defeating outcomes no matter what criteria one is using as a foundation for one's critical perspective.13

---

13 Take for example the virtuous disposition to act for the sake of a friend or loved one. The point is often made that this disposition cannot be governed by consequentialist considerations without forfeiting its value, since we ought to be disposed to act for the sake of our friends and loved ones and not for the sake of promoting the impartial value of such relationships. Yet it is clear that this disposition is equally jeopardized if we act in the interest of promoting a Kantian conception of ethics, for example, rather than a consequentialist one. Your friends will be just as disillusioned to hear that you are acting in their interest because such behaviour is the moral duty of a transcendentally free agent than if you tell them you are maximizing agent-neutral outcomes. Moreover, even if we strive to live according to the recommendations of virtue theory (a seemingly more personal, less alienating option), we cannot aspire to behave as ideally virtuous agents unreflectively. We must eventually critically assess our progress and judge how our actions are meeting our criteria for what constitutes virtuous behaviour. Once we take up this critical perspective, we undermine the virtue of friendship because we act out of reasons related to perfecting our own virtuous character rather than promoting the interests of our friends.
As a result, any framework seeking to capture the value of unreflective moral aims without giving up on the ability to critically evaluate these aims will have, at some point, to face the problem of insulating our different levels of moral thinking from one another. Investigating the genealogy of our moral intuitions is one such framework (one that provides both a novel means of systematizing our intuitions and an explanation for why we can expect cultural selection to favour satisficing solutions), but any framework that takes up a reflective perspective towards ethics will face their own variant of the genealogy dilemma and the subsequent need to insulate our unreflective virtuous aims.

Once we accept this fact about the way certain moral aims must be pursued, we can see that the move adopted by indirect consequentialism towards a reflectively unstable mix of separate levels of moral thinking is an appropriate response to an unavoidable problem. Strange as it may seem, we need at least two separate levels operating within our moral psychology. Hence, we need the psychological mechanisms required to help us shift between these levels - mechanisms that must remain concealed until they are called upon to snap us out of our reliance on spontaneous intuitive action and prompt us to switch over to the perspective of critical reflection.

Pettit refers to these mechanisms as 'triggers', a term that appropriately conveys the idea that they remind us to examine the broader ethical implications of our actions by abruptly diverting our attention to a different evaluative perspective when exceptional conditions present themselves. Michael Stocker, on the other hand, pejoratively refers to these mechanisms as "mental alarm clocks" to stress the absurdity of being periodically awakened from our ordinary patterns of thinking without external provocation. The idea of a mental alarm clock is not particularly appealing, but the disparaging nature of the term is highly dependent on how much faith we place on the 'ordinary' perspective from which we are being awakened (if not our general loathing for our alarm clocks). We would not want to only perform periodic checks on our intuitions, since we presumably want critical reflection to kick in when we find ourselves forced to make decisions in
certain unusual circumstances. But *pace* Stocker, there is nothing obviously wrong about being periodically impelled to check our intuitions from a reflective point of view. How often should we check our intuitions? Under what conditions should we be triggered to make unscheduled checks? Answer to these questions are extremely difficult to predict. I therefore prefer the term 'moral anchors' to describe the psychological mechanisms that govern our capacity to take up a reflective point of view, since this term allows us to remain agnostic about the difficult empirical questions related to how often and under what conditions the mechanisms ought to take effect. The term also nicely implies that at the surface level (where we can 'float' without having to constantly be aware of our position or direction) these mechanisms allow a certain amount of deviation from our ideal location while simultaneously preventing us from drifting too far away from our underlying target.

*Two Separate Functions*

While it may seem extraordinary that such mechanisms exist in our minds, introspection reveals that we sometimes think about our moral choices from at least two separate perspectives, and that we often feel compelled to switch to a critical level of thinking because of anchors that reveal themselves when we encounter unusual situations. In particular, when we think of cases where we are compelled to question the authority of ethical principles that we normally take for granted as plainly justified, we find that such cases are very reasonably interpreted as instances where the ordinary recommendations of our intuitive level of thinking appear inadequate and we are subsequently driven to evaluate the principle from a critically reflective point of view.

Consider, for example, even the very basic principle that each person ought to be treated equally. That all individuals be given equal consideration and opportunity is a fundamental moral tenet few would want to question. Yet we often second-guess this principle in special circumstances, one of the more notable of these being the
controversial case of affirmative action. When faced with this issue, many believe that the normally acceptable principle that all persons be treated equally is suddenly not as obviously intuitive as it appears in its stark, non-contextual form. Thus, many choose to support policies that (at least temporarily) contravene the principle by giving privileged status to groups that have been previously disadvantaged. My aim here is certainly not to defend either side of this thorny debate. The point is that no matter what one's final opinions on the subject may be, the case is one where our otherwise unreflective ethical intuitions about fairness are called into question and we are subsequently forced to address the appropriate implementation of the principle from a critical point of view. We obviously do not ordinarily think about the equal treatment of persons from this point of view, because we normally internalize this intuition to the point where our adherence to it is all but automatic. But when extenuating circumstances arise, like the restitutive appeal of affirmative action, it is possible to find that we are prompted (or 'triggered' if you will) to examine the authority of even this deeply held intuition at a reflective level of moral thinking.

As revealing as this introspective data is, to many it is not immediately obvious that this case, and other examples like it, are most plausibly interpreted as evidence that we operate at more than one level of moral thinking. Instead, they are often interpreted simply as instances where conflict and indecision occur within a single ethical perspective. On this interpretation, the difference in the way we think about our ethical principles is accounted for entirely by the fact that our normal way of thinking about ethical considerations is distorted by an inability to decide on an appropriate course of action in an especially difficult situation. On this view, it is not that we suddenly find that we have to switch to another level of thinking; we simply 'switch' to being confused about what our lone ethical perspective demands of us.

Yet this explanation wrongly ignores the fundamental difference that exists between the way we think about ethical standards from a critical versus an intuitive point
of view. This difference is not so stark that the different levels do not sometimes overlap or exist simultaneously within our psychology. But a difference there nonetheless remains. (Again, if this difference did not exist we would not be able to both achieve and periodically reevaluate those aims that must be pursued unreflectively to retain their moral value.) This difference is that intuitive moral thinking is dependent on our acceptance of the legitimacy of ethical principles and our commitment to their authority. Critical thinking, on the other hand, is explicitly focused on the disengaged evaluation of whether our principles promote the successful implementation of our foundational ethical objectives (whatever these objectives may be). Although these two levels of moral thinking are often closely intertwined and difficult to distinguish clearly in our everyday psychology, they clearly fulfill two very different functions.

Hence, the mistake is to regard our critical level of thinking as the imperfect execution of a single moral psychology. Carrying out the evaluative function of critical reflection is a difficult task, but this difficulty should not blind us to the fact that there is a separate function at work here, and not merely the impaired realization of a continuous mode of ethical thought. The presumption that one pattern of thought is attainable that can, if sufficiently clarified, fulfill both of the functions of intuitive and critical thinking is the most common obstacle to accepting the idea that we rely on two separate levels of moral thinking. Yet, rarely are substantive arguments offered to demonstrate that we ought to expect our moral psychology to be unified. Instead, critics of the idea that moral thinking involves more than one level usually rely on the unapologetic presumption that a unified ethical psychology is achievable, and that this supposedly unified mode of thinking is obviously more desirable than any bifurcated alternative. However, this presumption is suspiciously naive, for it assumes that there is some exclusive means of implementing our foundational ethical objectives that magically alleviates the difficulties that arise because critical and intuitive thinking are so ill-suited to successfully executing each other's psychological function. If we give up on this excessively romanticized (but
surprisingly common) notion of a perfect level of moral thinking that lies just beyond our reach, the conclusion that we are equipped with two different levels within our moral psychology becomes far more plausible. The fact that these two separate levels must be insulated from one another makes their management a thoroughly arduous enterprise. But if we set aside the false comfort of believing in a flawless single moral perspective, we ought to be able to recognize that we need a reflectively unstable balance between two levels of thinking to accommodate the separate tasks required of our evolved moral psychology.

III. Memetic Schizophrenia

The presumption that we possess only one unified level of moral thinking carries with it the accusation that theories postulating more than one level of thinking invariably lead to a form of *moral schizophrenia*. Michael Stocker, the originator of this term, claims that the split between one's motives and one's justificatory values is indicative of a "malady of the spirit", and that, conversely, if we are to lead a good life we must be both moved by our justificatory values and value that which moves us. (p. 454) Thus, while it may be *possible* to construct ethical theories that require two separate levels of thinking, Stocker's objection is that this option invites a pathological condition into our moral psychology. The use of the term "schizophrenia" is surely an exaggeration of this condition. (Moreover, because the term is used to criticize the dividing of our justifications from our motives, it seems to refer to a diagnosis of having a 'split personality', a condition that is expressly *not* the same as schizophrenia, despite the common misperception.) However, the term does convey the idea that two-level ethical theories seem to encourage a disjointed experience within our psychology. The question, then, is whether this quality is indicative of an aberrant form of moral thinking. It may be possible to separate levels within our psychology and appeal to moral anchors to decide
when to switch between these levels, but will doing so make us mentally ill in some important respect?

It will not. The idea that divisions within our psychological faculties are necessarily indicative of a 'malady of the spirit' or some pathological deviation from an otherwise healthy state of consciousness is, again, an assumption based on a romantic ideal rather than established fact. Furthermore, adopting a memetic perspective toward our moral psychology is an especially helpful way of dispelling the romantic ideal in question. If we take the memetic picture of the mind seriously, there is no reason to presume that an undivided unity exists among the flurry of cultural replicators that inhabit our brains. If we are willing to accept that the structures of the brain and the cultural contents that fill it are both products of an intricate coevolutionary process, then there is no reason to expect the outcome of this process to be one that results in a completely undivided level of thinking.

*Explaining our Unified Experience*

Introspective experience tells us that there is at least *some* basic sense of harmony that exists within our minds, and that this sense of harmony is present within our capacity for moral decision making.\(^1^4\) However, this harmony does not necessarily indicate that we operate on just one level of moral thinking. Nor does it indicate that promoting the use of more than one level of thinking invites the dissolution of the harmony we experience. For if we interpret our experience as the result of a long and complex evolutionary process, it is instead likely that the harmony present within our moral psychology is the product of a continuous balance between the competing demands of our different psychological propensities. Evolution generates stable outcomes if the selection

---

\(^{1^4}\) Without at least an elementary sense of harmony or continuity in our moral thinking it would be difficult even to establish that there is for each of us a person to which we can attribute a moral psychology at all. See (Parfit 1984).
forces acting on competing replicators reach a state of equilibrium, but this is not always achieved by producing a homogenous result. Instead, evolutionary stable outcomes are often the result of balances struck between equal selection forces working in opposition to one another.\textsuperscript{15} Thus, rather than building a completely indivisible final outcome, it is possible that evolution has produced a stable sense of harmony within our minds in the form of a self-correcting balance between competing psychological forces that remain in constant flux.

What this possibility suggests is that we ought to think carefully about how to best interpret the sense of harmony we experience in our moral thinking. It is tempting to assume that the apparent coherence that exists in our mental states reflects a natural uniformity, the disturbance of which would presumably lead to unhealthy consequences. But if we reflect on the evolutionary origins of our psychological capacities, I think this coherence is best interpreted not as the manifestation of some indivisible unity, but as a carefully balanced equilibrium between competing selection forces acting on our separate patterns of thinking. Moreover, this is exactly what we ought to expect from a memetic interpretation of our mental faculties. The most thoroughly worked out account of a memetic view of the mind is arguably Daniel Dennett's \textit{Consciousness Explained} (Dennett 1991), and, not coincidentally, Dennett's primary aim in the book is to attack the assumption that a unified underlying structure (what he calls a "Cartesian Theatre") exists somewhere in the mind. In place of this assumption, Dennett defends a more fluid conception of the mind where consciousness is the product of the "multiple drafts" of experience that are created by the brain. Instead of a single stream of consciousness that

\textsuperscript{15} See (Maynard Smith 1982) for examples of such cases in the form of stable genetic polymorphisms. The most memorable of these is the evolutionary origins of the sex ratio in organisms such as humans. While it may seem advantageous to have a much higher ratio of females to males (because of the dramatically different reproduction costs associated with each sex), the ratio remains relatively even because if the balance of females to males becomes too high then the relative fitness of being male suddenly increases and will eventually restore a more even ratio. What is important about this example for our present purposes is the fact that the stable equilibrium achieved is not homogenous or invariable from moment to moment. Instead, an evolutionary stable outcome is generated from the balance between two competing selective forces that work in opposition to one another.
is generated within some central location in the brain, Dennett claims that the brain creates multiple steams of independent experiences which subsequently rise to the surface of consciousness according to memetic selection forces that are not necessarily governed by any single neurological module.

Dennett’s view leads to many interesting and controversial conclusions for traditional problems in the philosophy of mind (e.g. qualia and personal identity), but his primary argument for the multifaceted structure of experience is extremely persuasive. If the mind is made up of memes competing for space in the neurological architecture of our brains, then presumably we have every reason to believe that our psychological faculties are the result of a ongoing equilibrium between cultural selection forces acting on the competing drafts of our experience. How exactly we end up with even a basic sense of harmony out of this continuous process of selection between drafts is, at present, a mystery not even Dennett can presume to solve. Yet the mystery is no less peculiar if we arbitrarily postulate a unified psychology to explain the miraculous harmony we experience than if we attempt to account for it. A memetic explanation of our experience as the product of an equilibrium between multiple strains of thought at least has the advantage of providing a credible hypothesis for this miracle, rather than simply offering a Panglossian restatement of the harmony that is (thankfully) normally present in our mental states and then assuming that this harmony is the result of some natural uniformity.

Dennett’s memetic explanation of consciousness is important for our inquiry into the nature of our moral psychology, because it reinforces the idea that the sensation of harmony we experience in our moral thinking is the product of a delicate balance between separate levels of thinking, as opposed to simply a reflection of some pre-existing unified structure. The product in question is reflectively unstable, but Dennett makes a convincing case that such structures represent a plausible naturalistic response to the computational requirements placed on organisms attempting to survive in complex
environments. Far from inducing schizophrenia, the divisions between separate levels of moral thinking ought to be interpreted as part of the ordinary functioning of a healthy moral psychology that has evolved from natural origins.

An Emerging Awareness

Though a 'multiple drafts' model of the mind is at least persuasive as a naturalistic interpretation of our psychology, it is often difficult to shake the assumption that appealing to separate levels of thinking will disrupt a valuable part of what is required to live a coherent and fulfilling moral life. I think this assumption holds a powerful intuitive grip on us because it is closely linked to the tradition of philosophy, beginning with Descartes, which views the self as an indivisible whole that exists outside the confines of our material existence. This classical dualist view of the mind is by now all but extinct, but its tacit influence continues to exert an intuitive pull on contemporary philosophy despite the fact that few would want to commit themselves to its full implications.\(^\text{16}\)

Thus, thinkers like Stocker would almost certainly not want to postulate a non-naturalistic basis for the unity they attribute to our moral psychology. Yet the claim that divisions in our moral thinking necessarily lead to schizophrenia is predicated on a view of the self that is suspiciously similar to the full-blooded dualist view of the self as an indivisible immaterial object. In fact, the brevity of Stocker's argument is reminiscent of the argument for reliable knowledge of the external world in the Third Meditation: just as Descartes claims that God's benevolence ensures that we \textit{cannot possibly} have been designed in such a way as to be constantly deceived, Stocker claims that (for some undisclosed reason) we \textit{cannot possibly} have been designed to function with a divided

\(^{16}\) Most secular academics would deny that we believe our identity is the reflection of an immaterial soul that survives the death of the body, yet most of us tacitly assume that our identity is somehow a permanent, naturally unified object without acknowledging the connection between these two theses. To return to the methodology of chapter four, investigating the genealogy of the two ideas would likely show that the latter meme is a cultural leftover (although not specifically a runaway effect) of the former, and that this 'natural unity' meme continues to propagate even though its parent meme has been for the most part eradicated within credible academic circles.
psychology and separate levels of moral thinking. Once the structure of Stocker's argument is revealed, however, we can see that he, and critics with similar agendas, have little ground to stand on without the outdated appeal of the Cartesian tradition and its reliance on a kind and all-powerful designer. Unless some other story can be told about why we have been designed with a unified psychology, the argument against our experience being the result of multiple levels of thinking reduces to a work of ungrounded faith. If we are willing to take our naturalistic origins seriously it seems clear that we ought to sober up and accept the fact that there is no prima facie reason to presume that an undivided psychology is the natural state of our thinking, and we ought to similarly resist the idea that all other psychological structures are necessarily pathological.

This is not to say that there is nothing interesting, however, about the possibility of systems that make use of multiple levels of moral thinking to form a seemingly coherent whole. Such systems are not pathological, but they are certainly very unique and well worth our attention. Even if systems of this kind represent an entirely 'natural' psychological state for us, they are still peculiar enough that it is difficult to grasp how we manage to successfully balance more than one level of thinking within a reflectively unstable psyche. Therefore, even if a delicate balancing act of this sort is the likely result of cultural evolution's construction of our psychology, it is nonetheless a very strange prospect and one that deserves further investigation.

In fact, certain theorists seem to acknowledge this fact even if they are not completely aware of its implications. Consequentialists have, of course, wrestled with the problem out of necessity. However, other theorists, particularly those working from

---

17 See (Descartes 1986), Third Meditation.
18 Parfit is a good example here, although as a consequentialist thinker he shows an awareness of the scope and difficulty of reconciling our critical and intuitive levels of thinking that is above and beyond the call of duty. See his closing remarks in part one of Reasons and Persons (p. 111-14). Here, Parfit discusses the prospects for a "Unified Theory" (the development of which he says would take a book of itself), the purpose of which is to combine the conflicting recommendations of his analysis of Consequentialism and
a naturalistic perspective, also address the fact that striking a balance between the critical and intuitive functions of moral thinking is required for the implementation of any viable ethical system.

The most obvious of these is understandably Dennett, who finishes *Darwin’s Dangerous Idea* with an examination of the inability of ethical theories (including both consequentialism and Kantianism) to account for the pressures of real-time decision making.\(^{19}\) Dennett is acutely aware of cultural evolution's reliance on satisficing shortcuts, and he clearly recognizes the self-defeating character of applying straightforward instrumental rationality as a means of solving all ethical questions. He therefore recommends that we adopt a "Moral First Aid Manual" to make our ethical decisions computationally tractable. By proposing this solution Dennett clearly grasps the dual role required of our moral psychology, for he admits that this first aid manual must involve a delicate balance between excessive rational reflection and our unexamined obedience to the provisional norms he describes as pragmatic "conversation stoppers" (p.506).

Although Dennett is very much aware of the two levels of thinking that play a part in his Moral First Aid Manual, he falls short of explicitly acknowledging that these two levels must be insulated from one another and that they must therefore be combined within a reflectively unstable whole. He notes at one point that our adherence to certain strategically unquestioned norms must be internalized to such an extent that our obedience is spontaneous "even in the face of contrary rational challenges designed to penetrate [our] convictions". (p. 507) Yet he still curiously fails to follow through on the

\(^{19}\) See (Dennett 1995), especially pp. 494-510.
next logical step by neglecting to spell out exactly how this peculiar characteristic affects our moral psychology.²⁰

One need not, however, be well versed in the terminology of memetics to recognize the dual role required of our moral thinking. In the book Wise Choices, Apt Feelings, Allan Gibbard outlines a theory of human normative judgment that is impressively mindful of the questions we have seen arise for the task of critically reflecting on our existing moral intuitions. Gibbard's theory, which he labels "norm-expressivism", seeks to provide a naturalistic explanation for the broadly normative character of human experience. In his discussion of morality, Gibbard presents an "ecology of broadly moral sentiment" (p. 274) in which our capacities to act according to shared norms and guiding moral sentiments (e.g. shame and guilt) are described as evolutionary adaptations that have been selected because they dramatically improve social coordination. In fact, Gibbard claims that the key to human moral nature lies in coordination (p. 26), and he argues that the propensity to accept and commit ourselves to norms serves as the basic foundation for our evolved moral capacities because it so effectively allows us to coordinate our behaviour and moral sentiments. Gibbard's norm-expressivism is a rich and far-reaching theory, and I will not attempt here to give it a thorough examination. However, I want to point out some of the interesting ways in which Gibbard's methodology shares the same structural challenges that our genealogical analysis has uncovered. First, Gibbard acknowledges that we must ultimately work from within the boundaries of our moral intuitions (broadly understood), and he shares the view that we can nonetheless take up a critical perspective towards these intuitions in order to evaluate their prescriptive authority. In Gibbard's

²⁰ There are two likely reasons for Dennett's failure to seize on the psychological details of what his Moral First Aid Manual implies. The first is the fact that Dennett is skeptical about the legitimacy of any guiding ethical principles - even those derived from a critical level of thinking. This underlying pessimism about the justifiability of our moral aims leads to a more passive approach to the task of reconciling theoretical goals with our unreflective intuitions. Second, Dennett is for some reason willing to consider a political solution to the problem in the form of something like establishing a ruling elite - the first option discussed earlier in this chapter (and for good reason quickly dismissed). See pp. 506-9.
terminology, morality consists in norms for what are ultimately moral sentiments, yet we need not appeal to these sentiments directly as a guide to the kinds of norms that ought to govern them. (p. 275) We can instead deliberate about the pragmatic merits of our moral sentiments and adopt norms that seek to reshape these sentiments in more systematic ways. Thus, Gibbard accepts a picture of ethical justification that is similar to Rawls' reflective equilibrium, since he believes we can proceed with systematic moral inquiry by starting with tentative views and then working towards refining our common-sense moral sentiments. (pp. 25, 277-8)

Second, and more importantly, Gibbard is evidently aware of the risks associated with systematically investigating our moral sentiments. He does not explicitly use the terminology of self-defeating outcomes. However, he notes that if ideal moral inquiry were to strive to build a rigorously consistent system of moral sentiments, it may succeed only in driving us away from a morality we can thrive on - one that will allow us to live a life most worth living. (p. 324) Gibbard therefore proposes that in building ethical systems we must find devices that can both gain our reflective conviction and mesh sufficiently with our existing moral sentiments in order to work in action:

They must not put us so far out of touch with ordinary ways of thinking and discussing that we lose the advantages of a human normative makeup, the coordination moral discussion can bring. They must combine hardiness under scrutiny with daily moral power. (p. 325)

This realization contributes, I think, to Gibbard's ability to recognize the two different functions required of moral thinking. Gibbard claims that in moral inquiry we need two separate "prongs" for assessing our feelings. (pp. 283-84) We need a pragmatic prong to evaluate the gains and costs of our moral sentiments, for these sentiments are often deficient (despite the fact that they may be generally trustworthy). Plus, we need an intuitionistic prong to directly ascertain how to live according to the norms governing our feelings - the norms that we spontaneously commit ourselves to and accept as being plausible. Gibbard is confident that these two separate prongs each play an important
role in the overall picture of moral inquiry, and he seems keenly aware of the fact that these two different functions must be delicately balanced within an interdependent whole. However, he nevertheless stops short of recognizing that this balance must be reflectively unstable for the two prongs of moral inquiry to operate effectively, and he therefore neglects to investigate the effect this peculiar quality implies for our moral psychology.

Thus, the writings of philosophers like Dennett and Gibbard suggest that a growing awareness of the dual function of moral thinking has begun to emerge within ethical theory, even if this awareness has not yet reached the point where philosophers are willing to recognize the unique structure required for our moral psychology to successfully implement two insulated levels of thinking.\footnote{This awareness even shows itself in fiction related to naturalistic interpretations of our ethical systems. In Stephenson\'s \textit{Snow Crash}, the story\'s Mafia henchmen are asked why they are pursuing their seemingly pointless mission of capturing a fifteen year old girl with whom they have had previous dealings. Their deadpan admission that they have no idea (they are simply following orders) is curiously self-aware: "the Mafia way is that we pursue larger goals under the guise of personal relationships. [...] This is how we avoid the trap of self-perpetuating ideology. Ideology is a virus. So getting this chick back is more than just getting a chick back. It\'s the concrete manifestation of an abstract policy goal." (1993, p. 350)} That this awareness is beginning to emerge is not surprising. As we have seen, taking up a naturalistic perspective towards our psychology and its evolutionary origins tells us that cultural evolution favours satisficing shortcuts to solve the complex computational tasks present in our social environments. We ought to therefore expect these shortcuts to play a noticeable role in the psychology of creatures such as ourselves (i.e. creatures who must routinely navigate complex cultural landscapes). If these shortcuts play an important role in our thinking, then it should come as no surprise that an underlying two-level structure should begin to reveal itself - even if it is not explicitly recognized as being implied in what is required to adopt insulated satisficing strategies.

What is surprising is the fact that the reflectively unstable character of this underlying two-level structure is so dramatically under-appreciated among philosophers who nonetheless realize that we possess the capacity to address ethical issues from two separate ethical perspectives. It is only a small step to realize that these two ethical
perspectives are incompatible with one another and that they must therefore be combined within a reflectively unstable final product. Yet this last step and the implications it has for our moral psychology are almost always conspicuously absent. It is as if the growing awareness of the dual functions required of our ethical thinking has led us to the edge of a profound realization concerning the structure of our moral psychology, yet it has not quite given us the nerve to jump.

This, I submit, is the most valuable contribution a memetic approach to ethics can provide that is of broad theoretical significance. By uncovering the evolutionary causes for our two levels of moral thinking, memetics can help us to understand why we need these two psychological levels to implement our ethical objectives. In terms of practical ethics, this understanding can help us to more effectively manage the implementation of our intuitive level of thinking (as was shown in chapter four using the notion of genealogy). Yet this understanding can also explain enough about the naturalistic basis of both levels of thinking to force us to come to grips with the delicate balance that must be struck between our commitment to each individual pattern of thought. In doing so, a memetic approach to ethics can have a valuable impact on the way we theorize about our moral psychology by helping to ease us into accepting the fact that reflective instability is an ordinary part of our moral experience.

*Psychological Harmony as Virtue in Practice*

For many this will not be an easy pill to swallow. Accepting that we must keep the individual levels of our moral psychology insulated from one another will alarm many philosophers, for it implies that something not unlike self-deception constitutes part of a healthy moral life. This is an odd conclusion, but if we take the implications of memetics seriously it is a conclusion we are better off facing and investigating in further detail, for if we quietly dismiss the problem we will be adding confusion to what is already a very strange and delicate issue.
Moreover, the information provided by memetics really only forces us to \textit{reinterpret} the internal harmony required to live a healthy and fulfilling moral life. Rather than viewing this harmony as a preexisting form of unity, we ought to instead view the perfection we attribute to an ideally structured psychology as something to strive for - a virtuous balance that we must continuously aspire to achieve. The difference, however, is that this balance must now be interpreted as something fragile and unstable. If we accept the memetic interpretation of our moral psychology, we can no longer believe that \textit{normal} ethical thinking exists in a state of perfect balance and that experiences which appear to disrupt this unity represent some pathological abnormality. We must instead view the balance that exemplifies a virtuously harmonious life as a hard-fought equilibrium between competing strains of moral thinking - one that captures the value of ethical aims that must be pursued unreflectively but still periodically evaluates their long term justifiability.

The perfectly virtuous agent will thus be one who internalizes both the appropriate moral dispositions and the appropriate moral anchors required to act spontaneously without abandoning the capacity to critically reflect on her actions. Such is the ideal: the agent who acts from pure motives until exactly the point at which these motives ought to be re-examined in order to best serve our underlying ethical objectives. This is a lofty goal, but it would truly be a harmony worth seeking. Thus, philosophers such as Michael Stocker and Bernard Williams are entirely right to claim that the virtue of harmony is an important indicator of a good life. A memetic approach to ethics insists only that this virtue is something that must be achieved rather than preemptively assumed.

\textbf{Conclusion}

In this final chapter, we have grappled with the realization that there is no easy solution to the dilemma created by the conflicting outcomes of chapters four and five. If we fail to
investigate the genealogy of our moral intuitions we leave ourselves at the mercy of cultural runaway effects, but if we proceed with this investigation, we invite the prescriptive dissolution of many justifiable moral memes. I have argued that the least injurious means of resolving this dilemma is to strike a delicate internal balance in the moral psychology of each individual agent - a balance between the level of critical reflection required to ensure that our moral intuitions remain justified and the level of unreflective performance required to ensure that our immediate efforts to promote these intuitions retain their morally praiseworthy character.

Although I believe this psychological balance is the most profitable way of reconciling the costs and benefits of critical reflection, I have attempted in this chapter to describe how the implementation of this balance will be a nonetheless precarious task. The need to insulate our intuitive level of moral thinking from being undermined by critical moral thinking makes the balance in question a very peculiar one, for it requires both that we possess two separate levels of thinking and that these separate levels must be combined within a reflectively unstable whole. The implications of this unique balance for our moral psychology are somewhat alarming, because they suggest that we must intentionally conceal certain aspects of ourselves from the critically reflective function of our moral thinking. The result is an ongoing form of pragmatic self-deception that is easily interpreted as a pathological mental condition we ought to unconditionally reject.

I have argued, however, that it is a mistake to interpret the need for a reflectively unstable moral psychology as a strategy that invokes a pathological condition into our moral experience. We cannot assume in advance that the harmony we perceive in our moral thinking is necessarily the result of a unified underlying psychological structure. Consequently, we also cannot assume that any deviation from this hypothetical unified structure necessarily represents a pathological way of thinking. In fact, if we adopt a memetic view of the mind it seems far more likely that the harmony that emerges in our moral experience is the product of complex selection processes operating on the separate
patterns of thought competing for space in our brains (and within our broader cultural environment).

The difference is important, because it indicates that we can no longer cling to outdated assumptions about living out some preexisting moral harmony that has been granted to us by a mindful designer. Instead, I believe that a thoroughly naturalistic approach to our moral capacities suggests that the virtue of harmony we ought to strive to achieve is an ongoing struggle to retain the appropriate balance between our critical and intuitive levels of thinking. It is what is required to successfully implement a reflectively unstable psychology that will most effectively promote our ethical objectives. This memetic view of what is required to achieve virtuous harmony may initially appear rather odd and exceptionally demanding. However, by allowing us to recognize the scope and structure of the task we face memetics at least allows us to address the issue without illusions, and this is undoubtedly a step in the right direction.
Conclusion: Moral Monkey Traps and Naturalized Perfection

There is an interesting myth about how hunting camps in Africa trap monkeys. Monkeys, it seems, can be a terrible nuisance to these camps by stealing food, rifling through garbage and sometimes making off with family pets. According to legend, however, these monkeys can be trapped in a most curious way. By attaching a hollowed out coconut to a tree and placing a piece of food (or a small shiny object) inside the coconut, hunters are said to be able to capture monkeys with ease. How does it work? The key is to ensure that the piece of food placed inside the coconut is the same size as the hole that allows the monkey to reach inside. Apparently, once a monkey has latched on to the food inside the coconut, it is unable to free itself quickly enough to escape from an approaching hunter. With the combined size of the piece of food and its own clenched fist, the monkey cannot remove its hand from the coconut without letting go of its prize. Hence, the 'trap' is that monkeys are built with a rather unfortunate design flaw: it is simply not in their nature to let go of food while attempting to escape.¹

As far as I can tell, this story about monkey traps is just that - a fictional story used as a parable to convey a unique message. Normally, this message is a lesson about human behaviour rather than the conduct of monkeys - it is told in order to warn us of the fact that greed and obsession can sometimes blind us to the recommendations of prudence. However, the monkey trap story also illustrates something important about our assumptions regarding the eccentric qualities of animal behaviour, because it is predicated on an understanding that animals can be constructed with very specific imperfections that tend to work against their best interests. Like the panicked flight of lemmings into the sea, the moth drawn to the flame or the laboratory mouse choosing to

¹As with most meaningful lessons in life, the point can be concisely summarized by an episode of The Simpsons. In this case, the myth of the monkey trap is reproduced when Homer gets his hand stuck trying to steal a can of pop from the office vending machine. The problem is resolved only moments before the fire department is about to cut off his arm, as they finally ask him, "Homer, you're not still holding on to the can, are you?".
receive sexual stimulation instead of food until it reaches the point of starvation, we are meant to view the inability of monkeys to release themselves from the coconuts of their captors as an evolutionary glitch - an unfortunate defect in their biological construction that leaves them vulnerable to a very unusual snare. The myth of the monkey trap may be fictional, but it relies on the fact that we recognize the possibility for peculiar imperfections to arise within the otherwise adaptive structure of natural selection.

I offer the story of the monkey trap as a fitting conclusion to my thesis because it captures the most basic message I have endeavored to present. This message is that evolutionary systems cannot be trusted to promote our best interests, for these systems are driven by structural mechanisms that can lead to chronic imperfections in the output they produce. These imperfections are more than just the ordinary limitations we accept as being part of our natural existence (e.g. that we grow old, that we require sleep, that we cannot leap ten metres in the air, etc.). They are products of selection anomalies that require special explanations for their existence in an evolutionary environment. In short, they are unexpected design flaws in the results produced by biological evolution in specific environments. I have focused on a particular type of design flaw in this thesis (those created by a runaway process), but the general theme being argued for here is that the mechanisms at work in natural selection can generate surprisingly maladaptive outcomes and other curious imperfections when they are distorted by unusual circumstances.

If we accept that such imperfections occur in biological evolution, I submit that we ought to expect similar results to occur in our cultural systems. Despite our ability as

---

2 Specifying an environment is important in order to characterize evolutionary results as design flaws. For example, whatever strange characteristic is responsible for moths being drawn to light may have been perfectly adaptive before their environment included objects like 'bug-zappers'.
intentional agents to direct the flow of cultural evolution in certain ways, the parallels between the structural mechanisms of cultural and biological evolution suggest that cultural transmission and selection will be affected by the same distortions that we observe in biological environments. As a result, we can anticipate strange and unforeseen outcomes being generated within our cultural landscapes, even though we appear to have more control over this medium than our biological system of inheritance.

I have argued in this thesis that because such cultural imperfections are likely, the most effective application of modeling cultural evolution for ethics will be to try to identify instances where selection anomalies have caused cultural evolution to provide us with moral intuitions that are deeply internalized but do not promote our best interests. In short, I have argued that the best option for ethics is to protect us from moral monkey traps. By appealing to the empirical information supplied by models of cultural evolution, we can obtain an additional means of evaluating the legitimacy of our common-sense moral intuitions, and we can therefore more effectively systematize the prescriptive authority that these intuitions possess.

However, the difficulties associated with implementing this application of cultural evolution models to ethics are significant. Specifically, the only feasible means of purging ourselves of cultural 'design flaws' (without inviting a corrosive potential into the normative authority our legitimate moral intuitions possess) is to adopt a reflectively unstable moral psychology - one that can operate at two levels of moral thinking, one critical and one intuitive, without either of these two levels being aware of the evaluative criteria guiding the other. The result is a very delicate and unstable balance between the competing claims of critical and intuitive thinking within our moral psychology - a result that is sufficiently bizarre to be often considered pathological.

I have suggested, however, that a moral psychology of this kind is not pathological. Instead, I believe we ought to consider the balance required to manage two separate levels of moral thinking part of what it means for moral agents to strive towards
the sense of harmony we attribute to the perfectly virtuous individual. This is fairly radical reconstruction of the notion of virtuous harmony as it is presently understood, but it is one that fits well with a more general view about the consequences we can expect to follow from taking up a rigorously naturalistic perspective towards ourselves as moral creatures. If we cease to view ourselves as latently perfect individuals attempting to maintain or reclaim our rightfully unified moral psychology, we can finally begin to appreciate our true status as products of an elaborate coevolutionary process, whose best course of action is to try to achieve an arduous sense of naturalized perfection within our moral psychology to the best of our modest abilities. If the study of cultural evolution can help us to conceive of ourselves and our ethical projects in this more discerning manner, then I think it will supply ethics with a very valuable contribution indeed.


**Bibliography**


244


------. Forthcoming. *Morals From Motives*.


