The Evolution of Moral Cognition

Tony James Scott

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Philosophy

2009

For Bridget, with love and gratitude.

Abstract

Many modern approaches to the evolution of mind have claimed that the fundamental drivers of our cognitive capacities and cultures are genetically specified psychological adaptations, which evolved in response to evolutionary pressures deep within our lineage's history. Many of our cognitive capacities are innate. Recent approaches to moral cognition have similarly argued that moral cognition is innate. In this thesis, I argue that even though our capacity for moral cognising is an adaptation, it is a learned adaptation. Moral cognition is not innate. In arguing this thesis I will question many of the assumptions of traditional cognitive science and evolutionary approaches to the mind. By incorporating theory and evidence from cognitive science and the philosophy of mind, I apply the explanatory frameworks of embodied and extended cognition to the domain of morality: moral cognition is both embodied and extended cognition. This places particular importance on the role of our bodies and world in the fundamental structuring and scaffolding of the development and execution of moral cognition. Putting this in an evolutionary framework, I develop a dual inheritance model of the non-nativist evolution of moral cognition focusing on the roles of niche construction, biased learning and active learning in the transfer of moral phenotypes between generations. Morality is a learned adaptation that evolved through the dynamic and reciprocal interaction between genes and culture.

Acknowledgements

First and foremost, I must thank my primary supervisor Kim Sterelny. His support, comments, criticisms and wise words allowed me to wander into the philosophical terrains that interested me most and produce this thesis. For this I am very grateful. Our midnight cups of tea always included discussions of cricket and would frequently journey into philosophy. At these times, I was lucky enough to have my very own philosophy lessons and have learned more over the last few years than I could have possibly imagined. Thanks Kimbo.

I must also thank Richard Joyce for his secondary supervision. I am greatly appreciative of his patience, insightful and frank comments, as well as his dry sense of humour. I would also like to thank him for writing the book that inspired a great deal of my early thinking on the evolution of morality.

Victoria University has a great philosophy programme with fantastic faculty and graduate students. I can't begin to imagine the number of hours spent arguing the finer points of philosophy in the hallways of the Murphy building or in the coffee room. The faculty have always been friendly, helpful, encouraging, and approachable. The students have always been quick to challenge any argument, for better or worse, and keep everyone on their intellectual toes – although perhaps not so much after Thursday curry and a few bottles of red! These have been the most intellectually stimulating years of my life. I won't try to individually thank you all as I will inevitably leave someone out.

There is, however, one student I must thank especially. Frieder Lempp and I were writing up our PhDs at the same time and the 3am games of chess and pots of coffee kept us both sane I am sure. It is not clear that the thousands of games of chess actually improved our game, but, as you say Frieder, we did manage to solve most of the major philosophical problems of our time. Looking back, my wild scribbles on your whiteboard helped to clarify my thinking in many ways – the extended mind in action! Thanks Frieder.

Many of the ideas in this thesis have been presented, in some form or other, at numerous seminars, conferences and workshops. Thanks especially to those who offered valuable feedback from events at Victoria University, the Australasian Association of Philosophy annual conferences, and the evolution workshops at Australian National University.

Finally, and most importantly, I must thank Bridget Vaney for her unconditional support, patience, tolerance, love and understanding. Without her this thesis would never have happened. Our son Liam arrived right in the thick of the writing up phase. He was a wealth of joy and welcome distraction from the daily grind. Much love and thanks to you both.

Contents

List of tables, figures and boxes	viii
Introduction	1
1. The Evolutionary Problem of Morality	8
1.1 Morality is an adaptation	8
1.2 The problem of cooperation	10
1.3 Proximate and ultimate explanations	12
1.4 The evolution of cooperation	14
1.5 The evolution of cooperation and the proximate mechanisms of morality	29
1.6 Conclusion	30
1.7 Summary	31
2. The Wealth of the Moral Stimulus	32
2.1 Nativism vs. empiricism	33
2.2 The poverty of the moral stimulus	38
2.3 The poverty of the moral stimulus: development	42
2.4 The poverty of the moral stimulus: task analysis	46
2.5 Authority independence and the moral learner	66
2.6 Conclusion	74
2.7 Summary	76

3. Cognition: Embodied and Extended	77
3.1 Traditional accounts of cognition	78
3.2 Traditional cognition and the moral mind	84
3.3 Explanations beyond the brain	91
3.4 Cognition is embodied and extended	98
3.5 Embodiment, extension and evolution	101
3.6 Conclusion	102
3.7 Summary	103
4. The Embodiment of Moral Cognition	104
4.1 The role of the body in cognition	104
4.2 The embodiment of cognition	106
4.3 Emotional embodiment	114
4.4 Embodying more than emotions	124
4.5 Moral simulations of moral situations	127
4.6 Implications for moral cognition	132
4.7 Conclusion	135
4.8 Summary	136
5. Extending Moral Cognition	138
5.1 Beyond the organism	138
5.2 Moral cognition is also outward looking	141
5.3 Embodiment and moral cognitive extension	143
5.4 Extending moral cognition	147
5.5 Moral norms and cognitive extension	152
5.6 Questions of impermanence	156
5.7 Group interaction and the construction of moral scaffolds	157
5.8 A note on the moral niche	163
5.9 Conclusion	165
5.10 Summary	166

6. The Gene-Culture Coevolution of Moral Cognition	168
6.1 The making of moral phenotypes	169
6.2 Dual inheritance models in evolution	172
6.3 Niche construction	176
6.4 A dual inheritance model of the evolution of moral cognition	178
6.5 In group / out group similarities and differences	197
6.6 Conclusion	203
6.7 Summary	205
Conclusion	206
References	217

List of tables, figures and boxes

Tables

2.1 Conceptions of innateness	35
6.1 Children's Psychological resources that contribute to high fidelity transmission of moral phenotypes	187

Figures

3.1 Changing understanding of the interactions between the nervous system, the body, and the environment	99
6.1 Culture and the development of moral phenotypes	187
6.2 Moral phenotypes and their influence on the moral niche	191
6.3 The cultural elaboration of moral disgust and the moral niche	197

Boxes

1.1 Ultimate explanations informing proximate explanations	13
2.1 Against the poverty of the moral stimulus	40
2.1 Causal psychology, moral problems and the doctrine of double effect	64
4.1 Theories of Emotion and Embodiment	113
6.1 Case Study: The Gene-Culture Coevolution of Moral Disgust	193
6.2 Cake dividing games and the evolution of fair division	200

Introduction

The idea that morality has evolutionary origins is not new. Indeed, Charles Darwin proposed that our moral sense is a product of evolution by natural selection. In more recent times there has been a renewed interest in evolutionary accounts of individual moral capacities and society-wide moral systems. These accounts have been, by and large, nativist. Although they vary as to the exact nature of our moral sense and the degree to which it contributes to our capacities for moral thought and action, they typically maintain that the central causal determinants of these capacities are innate, domain-specific, information processing, psychological mechanisms. The positing of an innate moral cognition is thought to explain the origin, maintenance and structure of individual moral behaviours as well as society-wide moral systems. These mechanisms are proposed to have evolved in response to adaptive problems in our species' evolutionary past – our moral minds have been shaped by environments deep within our lineage's history. In this thesis, I propose an alternative account for the evolution of moral cognition.

This project will be largely integrative. Drawing on theory and evidence from evolutionary biology, developmental psychology, cognitive science and the philosophy of mind, I will present an empirically tractable hypothesis for the evolution, development and deployment of moral cognition. I will propose that, in contrast to the view that our moral minds were populated deep in our ancestral past, the fundamental structuring of our moral minds occurs within one's life history. This does not, however, mean that there is no evolutionary story to be told. Our moral environments, which scaffold the

development and execution of moral cognition, have an evolutionary history, as do the various non-moral psychological capacities which allow the establishment of moral cognition. It is these interacting non-moral biological and moral cultural components and their joint evolutionary history which give rise to the moral agents and worlds we see today.

The focus of this thesis will be on two types of explanations that are central to any account of the evolution of moral cognition. The first are ultimate explanations: population-level explanations that focus on the selective pressures that led to the various traits in question. The second are proximate explanations, which focus on the various mechanisms and processes (both ontogenic and synchronic) that give rise to the various traits within an organism's lifetime. From these, we can see that there are two important questions for any study of the evolution of morality: 'what are the evolutionary forces that gave rise to morality?', and 'what are the various mechanisms and processes that generate moral cognitions over one's life-history?'

With respect to the first question, the central problem in explaining the evolution of moral cognition is to account for the apparent fitness-sacrificing behaviours of morality. Often moral behaviours do not appear to be in our individual fitness interests. If this is the case, how did morality evolve? Proposed solutions to this problem based on kin selection and reciprocity have, for a long time, been popular in explaining the evolution of morality. Recently though, such accounts have been shown to be inadequate explanations for the large-scale group phenomena that we see in the real world. As such, alternative solutions have been proposed which focus on punishment and group-level cultural dynamics. These evolutionary explanations involve the modelling of proximate mechanisms beyond the physical bounds and behaviours of any individual agent, giving us some insight into the types of mechanisms that are central to the evolution of moral cognition.

In regard to the second question, many current accounts of the mechanisms of moral cognition have posited an innate moral faculty (as described earlier). These explanations typically make specific commitments to the complexity and information processing nature of internal cognitive mechanisms. I will critically assess one argument recently presented in support of an innate moral faculty. Rejecting moral nativism, however, does

not conflict with my assertion that morality is an evolutionary adaptation. Natural selection acts upon heritable variation in fitness and this variation can have genetic or non-genetic origin, or both. Variations with non-genetic origins can be inherited with reasonable fidelity through cultural learning and perhaps other mechanisms. Adaptations evolve by natural selection; they are selected for because of the fitness effects they have on their bearers. Accordingly, adaptations need not be of genetic origin. Therefore, under a gene-based conception of innateness, adaptations need not be innate.

Given the shortcomings of nativist accounts, I will argue that understanding moral cognition and its evolution requires a broader explanatory focus whereby our moral cognitive processes extend beyond our brainbound neural systems, outwards into our bodies and the world. The body and external world become proper parts of our moral cognitive system. Moral cognition is embodied and extended. Moral adaptations need not reside in the head of the moral cogniser. Such a view changes our explanatory focus from one which is brain-centred to one which focuses on the dynamic interplay between agent and their moral environment.

Instead of looking inside for solutions to adaptive problems we must also look outward; by focusing solely on internal mechanisms we fail to capture the essential contributions of beyond-the-brain mechanisms in the development, execution and evolution of moral cognition. Moral cognising alters our moral environments, but importantly it also alters the moral environments of developing children with profound evolutionary effects. If our evolved moral cognitive traits are complex skills that are scaffolded or even partly constituted by our social environments, then we need to account for the ways in which those environments exist over generations, allowing the cumulative evolution of the processes that they support. As I will explain, there are mechanisms through which our moral environments are reliably inherited between generations such that cumulative evolution can occur. A purely gene-based account of inheritance fails to explain this important role of the environment.

This thesis will proceed in four stages over the next six chapters. The first provides an overview of the different ways in which morality is thought to have evolved. The second argues against one specific argument in favour of the view that there is an innate moral faculty that underwrites moral cognition – we have reason to think that moral cognition

is not innate. The third presents an account of moral cognition that emphasises the essential role that our bodies and environments play as constituent parts of the development and execution of moral cognition. And the fourth presents a dual inheritance evolutionary account that incorporates the central role of our embodied interactions with our moral worlds, paying special attention to the role of extended cognitive systems that structure and scaffold our moral cognitive processes and those of our offspring.

The following is a brief summary of the six chapters of this thesis.

Chapter 1: The Evolutionary Problem of Morality

Moral cognition is an adaptation that evolved due to the cooperation and group coordination benefits it confers. In this chapter I will give an overview of different evolutionary paths to human morality, with the aim of establishing that there is a plausible evolutionary explanation. Reciprocity-based accounts have been especially popular amongst theorists exploring evolutionary accounts of morality. But they are probably insufficient to explain large-scale cooperative behaviours such as those found in human adherence to social-wide moral norms. Reciprocity-based solutions fail to “scale-up” to large groups. Recent approaches involving reputation, punishment and cultural group selection provide plausible solutions to this problem. These accounts explain morality in large groups by positing proximate mechanisms that operate beyond the physical bounds of any one individual. As we will see, this becomes important when giving accounts of both moral cognition and its evolution. This chapter summarises work from the evolutionary sciences.

Chapter 2: The Wealth of the Moral Stimulus

After developing a plausible case that morality is an adaptation in chapter 1, I will critically assess one recent argument in support of the claim that we have an innate, domain-specific faculty dedicated to the task of moralising. This innate moral faculty supposedly explains not only individual morality but also the generation of society-wide moral systems. There are various different accounts of what exactly a moral faculty is and the nature of the arguments in support of such a faculty. Here I will focus on one high

profile account that draws strong methodological and structural parallels with nativist arguments in linguistics. Poverty of the stimulus arguments have been—either explicitly or implicitly—central to a number of recent defences of moral nativism. The argument rests upon two claims: the first is that the information available to the child is impoverished and therefore inadequate to explain the acquisition of the relevant capacity. The second is that the learning task is too complex to be accounted for via domain-general learning alone. In sum, it is argued that there is a gap between informational input and achieved competence that can only be bridged by innate information. Drawing on evidence from developmental psychology as well as expanding upon some recent critiques of the moral nativists' argument, I will argue that neither of these two claims are sufficiently warranted to establish that we have an innate moral faculty.

Chapter 3: Cognition: Embodied and Extended

The traditional view of cognitive science and the philosophy of mind has typically been individualistic: the underlying metaphysical and methodological assumptions are that the structure of the mind is internal, and the methods by which it is characterised and individuated is by this internal structure alone. Individualistic accounts typically confine cognitive processes to what goes on in the head of the cogniser. In this way, individualism makes specific claims about the inner complexity of cognitive processes and the specific roles that are attributed to body and world with respect to the cognitive system. After detailing the ways in which traditional accounts of cognition are individualistic, I will look at two prominent accounts of individualism relevant to the project at hand: moral cognition as presented by the moral grammarians (discussed in chapter 2) and evolutionary psychology. An alternative view is that cognition is not confined to computational manipulations of amodal symbolic representational structures in the head of the cogniser. Cognition often incorporates both the body and the world. This has implications for evolution; if our moral adaptations incorporate more than brainbound psychological processes, then those additional processes must be inherited.

Chapter 4: The Embodiment of Moral Cognition

Building upon some of the ideas of embodied cognition introduced in chapter 3, I will explore how moral cognition involves reciprocal feedback between brain and body. Body

structures do some of our moral information processing and those structures become constituent parts of that system. This is important because much of the fundamental structuring of moral cognitive processes are not confined to the head of the moral cogniser. As such, an explanation that excludes the processes of brain-body feedback-loops will necessarily be incomplete. Drawing on recent work in embodied cognition detailing how many of our cognitive processes are embodied, I will argue that moral cognition is embodied and look at some of the implications of this a view.

Chapter 5: Extending the Moral Mind

In this chapter I develop the ideas of moral cognitive-extension, arguing that moral cognition extends into the world. Such a view places our external environments, their structure and our interactions with them as constituent components of the moral cognitive system. The moral environments in which we are embedded are structures over which we offload moral cognitive processes. Many of our moral cognitive problem-solving routines are thus instantiated by distributing moral cognition over brain, body and world. The body and environment make a non-trivial contribution to the establishment of the moral phenotype. Those cognitive extensions and the society-wide systems that support them are visible to natural selection and are inherited by future generations. We actively alter our moral environments and those routines have fitness bearing effects on generation N and $N + 1$. This chapter will explore work in the philosophy of mind and cognitive sciences demonstrating how our cognitive systems are not bound by our physical boundaries. I will apply these ideas to moral cognition, arguing that moral cognition is extended cognition.

Chapter 6: The Gene-Culture Coevolution of Morality

In the previous chapters I argued that morality is plausibly both an evolutionary adaptation and an embodied and cognitively-extended complex skill that we learn from our moral environments. In this final chapter I draw many of the earlier chapters together to propose an account of the evolution of moral cognition. Drawing on recent work in dual inheritance accounts in evolutionary theory, I present a gene-culture coevolutionary model of the evolution of moral cognition. The constituent structures involved in the development and expression of the moral phenotype are inherited in

both the genetic and cultural channels. I identify some of the key mechanisms that allow the reliable, high-fidelity transmission of the moral phenotype over generations required for the cumulative evolution of moral cognition. The model pays specific attention to the ways in which moral culture plays a unique evolutionary role by altering the fitness landscapes within which genetic evolution occurs. In turn, genetic evolution alters our cultural worlds. Each dimension influences the evolutionary dynamics of the other. Selection acting on culture becomes an evolutionary efficacious mechanism in its own right.

The Evolutionary Problem of Morality

This chapter gives an overview of evolutionary mechanisms that sustain cooperative behaviours, which will in turn provide us with some understanding of the individual and group mechanisms that instantiate and sustain morality. By showing that morality does not have a single evolutionary origin, I demonstrate that there are a variety of proximate mechanisms that support individual moral behaviours and social-wide moral systems.

1.1 Morality is an adaptation

Organisms evolve and many of their current traits exist because the effects of those traits furthered the reproductive interests of their ancestors. Moral cognition is no different. We are the moral agents that we are today because our ancestors' capacity for moral cognising contributed positively to their reproductive fitness.¹ The capacity for cognition was selected for by natural selection and is an adaptation. But the claim that morality is an adaptation seems counterintuitive when we consider the *apparent* fitness sacrificing nature of much of moral behaviour. We often refrain from cheating when it appears to

1. Whether or not moral cognising currently positively contributes to fitness is an interesting empirical question, but not one that will be considered here. What is important for our present explanatory project is the claim that moral cognition, broadly conceived, evolved as an adaptation.

be in our best interests to cheat. People go out of their way to prevent others from cheating, even when such actions are of no benefit to them. Many people regularly give their time and money to charities, again, often at no obvious benefit to themselves. One of the central tenets of Darwinian thinking is that natural selection favours those traits that further an organism's reproductive interests. If moral behaviours were obviously fitness enhancing then explaining their evolution in Darwinian terms would be relatively straightforward. Yet this is not the case. Many moral behaviours appear to run counter to an individual's reproductive interests, and hence ought to be selected against. From an evolutionary point of view, the fact that individual organisms are prepared to sacrifice their fitness *on such a wide scale, persisting over generational time-frames*, cries out for explanation. Explaining these behaviours is *the evolutionary problem of morality*.

The problem fits into a broader one of explaining the evolution of cooperation. Individual moral behaviours and wider moral systems are largely solutions to cooperation problems between individuals and groups. Although some moral norms are not obviously solutions to cooperation issues, a large and significant number of moral norms are. Cooperative interactions have the potential to generate benefits that positively affect the reproductive interests of agents. Yet, in many cooperative situations there arises the problem of maintaining the cooperative interaction in the face of temptation to free-ride. Because many of the mechanisms that maintain cooperation are moral, understanding cooperation will provide us with some understanding of moral cognition and its evolutionary history. Further, many dyadic and group-wise cooperative interactions in human populations are underwritten by norms which regulate the cooperative behaviours of agents. Many of these norms are moral norms: those that pertain to betrayal, stealing, and cheating. They operate in the service of cooperation by prescribing the particular behaviours one *ought* to follow in cooperative interactions.

Morality also fits into the explanatory framework of the evolution of cooperation because the adherence to moral norms is itself a cooperation problem. Group-wide moral norms regulate social behaviours, but individual agents will often have an interest in violating moral norms because moral adherence often incurs a cost, such as the loss of potential individual benefits of non-moral behaviour, acting in altruistic ways, or enforcing another's compliance with moral norms. In this way, individual moral adherence can be considered a cooperative act, whereby the moral agent provides

benefits to others often at an individual cost to oneself.

Morality maintains the adherence to moral norms through the controlling or influencing of both our own and others' behaviours in relation to those norms – morality is a means of regulating norm-based behaviour. Moral thinking does this by contributing a particular motivational element to our judgements and actions. Normal people are not indifferent to the moral judgements they make and are motivated to comply with their moral obligations. As we will see, this regulating of behaviour may be self-interested or altruistic, individually driven or enforced by third parties or groups. Many self-directed moral judgements, for example, can be seen as solutions to cooperation problems by increasing individual commitment to cooperative norms and removing the need for costly mechanisms of external sanction (Joyce 2006). Feeling guilt about cheating is an effective way of preventing future cheating without needing to invoke external threat as a deterrent. Even self-directed moral judgements and norms that are not obviously cooperative in nature, such as those pertaining to sexual purity, can indirectly aid cooperation by inducing group conformity. As we will see later in this chapter, and again in chapter 6, group conformity is one of the mechanisms that facilitate cultural group selection, which in turn is proposed as a driver for the evolution of cooperation (Boyd and Richerson 2004).

From this perspective we can see how the benefits of cooperative action were the drivers for the evolution of our individual moral cognitions and group-wide moral systems. Being moral allowed our ancestors to have more babies primarily because it facilitated the expansion of large cooperative groups and the fitness benefits that came with those groups. Understanding the evolutionary forces that drove cooperative activity will therefore give us some purchase on the evolution of moral systems and the various moral cognitions that maintained them.

1.2 The problem of cooperation

Humans are unique in the primate group in the extent of cooperation that occurs between agents. Cooperation is a powerful mechanism that allows the generation of benefits typically unavailable to groups of non-cooperating individuals: multiple pairs of eyes are more successful at identifying threats and potential food sources than a lone pair

of eyes; hunting in a team is more successful in bringing down large game, from which all can enjoy the spoils, compared to the difficulty of the task for a lone hunter. But it is also the case that agents can reap the rewards of others' cooperative behaviours without contributing to the costs of cooperation themselves. A big-game hunter may deviously fail to contribute as much as the others in a dangerous hunt, yet still feast on its spoils. Such free-riding behaviours can undermine the generation of benefits from cooperative activity, and various mechanisms are needed to restrain such behaviours.

It comes as no surprise that valuable resources such as food are often associated with moral norms prescribing collective action associated with their collection, distribution, preparation and consumption. As archaeologist Martin Jones explains:

Meals that are strictly bounded by moral code are ... widespread among human societies around the world. Indeed, human meals of all kinds are framed within moral codes about sex, age, rank, and ethnicity, and the dinners do not typically sense that these rules are negotiable. They are set at some other time, by some other authority, part human and part divine. The rules of conduct are passed down from each generation to the next. This seems to mark us apart from our closest relatives. (2007:37-38)

In cooperative interactions each agent has the opportunity to cooperate or defect (free-ride). Cooperation entails individually costly behaviour that benefits the other partner in the interaction. Defection entails not cooperating while still receiving any benefits offered from other cooperating agents. Both agents cooperating will receive net benefit from the cooperative enterprise. In cases where only one agent cooperates, and the other defects, the cooperating agent pays all the cost while recouping no benefit, or incurring additional cost, while the defector receives benefits without paying the costs. Although everyone will be better off if everyone cooperates, defectors are individually even better off if they defect while others cooperate. Defection in these cases is, *prima facie*, the evolutionarily favoured strategy. The problem of cooperation therefore involves explaining how the group beneficial cooperative behaviours we see in human populations can be maintained in the face of individual temptation to free-ride.

There are two types of explanations that are important for gaining an understanding of the mechanisms which support cooperation: proximate and ultimate.² It is worth briefly discussing these two types of explanations, the relationship between them, and the ways in which they can mutually inform each other.

1.3 Proximate and ultimate explanations

Proximate explanations involve mechanisms that generate a particular behaviour over an individual's life history. These include individual psychologies, biologies, ecologies, genetics, learning, reproductive strategies, environmental resources, group structure and interaction. Affection for close family members, for example, helps maintain cooperative behaviours towards those family members. We protect and cooperate with them, in part, because our affections motivate us to. Ultimate explanations, on the other hand, depend on more general patterns of population-level dynamics over generations, and involve mechanisms such as kin selection, reciprocal altruism and group selection. For example, consider two explanations of human preference for sweet foods. The proximate explanation is that we have taste buds that are sensitive to sugar molecules, inducing pleasurable sensations upon contact. The ultimate explanation is that we have preferences for sugary foods because eating sugary foods provided a ready source of energy for our ancestors, increasing their relative fitness compared to those who did not have such preferences, thereby spreading the preference for sugary foods in our ancestral populations.

Both types of explanation are empirically constrained. Their explanatory success or failure is ultimately an empirical matter. Also, ultimate explanations *inform* the proximate explanations and vice versa (see Box 1.1). Mathematical models of evolutionary dynamics can be used as a way of testing hypotheses about the roles of various proximate mechanisms. For example, the modelling of psychological theories of culture and learning in evolutionary models allows one to establish the effects of learning over evolutionary time frames, providing useful empirically testable hypotheses into the roles

2. Tinbergen (1963) proposes an alternative four way distinction amongst explanatory projects: adaptive, evolutionary, developmental and proximal. I will view adaptive and evolutionary explanations as broadly corresponding to ultimate explanations and developmental and proximal explanations as broadly corresponding to proximate explanations.

of various proximate mechanisms in the transmission of cultural variants. Without mathematical models of cultural transmission it would be difficult to establish under what conditions maladaptive cultural traits might be transferred within a population (Laland 1993). Because the two types of explanation are mutually informing, understanding the mechanisms involved in the evolution of cooperation will help inform us of the relevant psychological mechanisms, or otherwise, that underwrite cooperative (and in terms of this project, subsequently moral) activity.

Box 1.1 Ultimate explanations informing proximate explanations and vice versa in reproductive conflicts in *Dinoponera* queenless ant colonies

Ultimate explanations inform proximate explanations and vice versa. A clear example of this comes from the use of inclusive fitness modelling in eusocial insects. Ratnieks and colleagues (2001) detail how Hamilton's (1964) inclusive fitness theory has provided some novel predictions with respect to facultative sex allocation and worker policing in social insects.³ In the queenless ant colonies of *Dinoponera quadriceps* and *D. australis*, all female ants can potentially become reproductive in their lifetime. When they do they are called *gamergates*. Because reproduction is limited to one or a few gamergates at any one time, dominance hierarchies of a small number of potential reproducing gamergates emerge, below which are the worker ants. This hierarchical ranking system means that the higher ranked females will spend most of their time engaged in aggressive struggles to increase their rank and therefore the probability of being gamergate and passing on their genes (i.e., increase reproductive fitness).

Inclusive fitness models make empirically confirmed predictions about how the length of the hierarchies will be affected by colony size and the shape of the hierarchy itself. The models show under which parameters it is more fitness enhancing to become a worker and help further one's genetic interests via the genetically related colony, or to directly further one's genetic interests by attempting to become the next gamergate. Here we have a case in which models of the ultimate causes (i.e., inclusive fitness models) make predictions which are empirically confirmed by the subsequent studies of proximate causes, such as specific dominance behaviours of gamergates (Monnin and Liebig 2008).

3. As the authors state, this is important in that the scientific theory can make novel predictions as opposed to, what might be seen as, *post hoc* rationalisations.

Inclusive fitness models make predictions about individual behaviours within the colony. Under certain conditions selection can favour individual ants overthrowing a gamergate to more directly further their own reproductive interests. Genetic modelling by Monnin and Ratnieks (unpublished, reported in Ratnieks, Monnin et al. 2001) details which conditions this type of strategy will arise in the population, namely when the gamergate is old and the overthrow is highly likely to benefit the overthrowing ant. But their models also predicted that “other workers and the gamergate should oppose early overthrow” (2001:204). These predictions are in line with previously poorly understood observations that workers will seemingly punish other ants who challenge the active gamergate by physically restraining them for up to three days. Here, ultimate explanations give understanding of what certain proximate behaviours are for, namely informing us that the ants are immobilising other ants which have prematurely tried to usurp the gamergate. These models also predict that ants need some method of age-assessing the reproducing gamergate. Dominance behaviours that the gamergate performs with the other higher-ranked females, such as blocking and abdomen rubbing, are “honest signals of vitality” which can be used by other females to gauge the age of the gamergate, and therefore whether or not she should be overthrown. Ant behaviour is better understood because of the mutually informing nature of ultimate and proximate explanations.

1.4 The Evolution of Cooperation

In human populations, no single explanation is sufficient to understand all observed cooperative phenomena and it is very likely that cooperation evolves and is maintained by multiple mechanisms. The explanation as to why we have norms of moral behaviour towards our kin, for instance, may be different from explanations as to why we have norms of moral behaviour towards our friends, and again, towards complete strangers. The norms that prescribe those different behaviours are likely to be understood through a variety of different mechanisms (Harms and Skyrms 2008). I will now outline some of the most prominent accounts in the literature on the problem of cooperation, providing insight into the types of psychological and other mechanisms involved in the evolution and execution of moral cognition.

1.4.1 Kin selection (inclusive fitness)

If we are to classify an organism's fitness in terms of its ability to pass its genes on to future generations then we can see how the helping of one's offspring is a fitness enhancing act. The improvement of the health and welfare of one's offspring, even at a cost to oneself, will increase the likelihood of one's genes being passed on to future generations. In such cases, what would appear to be an individually altruistic act of helping one's offspring is, from a genes-eye perspective, actually a genetically-selfish act. By viewing such behaviours in terms of one's propensity to pass on one's genes, we can broaden our explanatory scope to include altruistic behaviours towards relatives that are not direct decedents. Under the right conditions, it is in one's genetic self-interest to aid genetically related relatives even at an individual cost to oneself. In these cases, the degree of relatedness multiplied by the reproductive benefits to the recipient is greater than the reproductive cost incurred by the altruist, mathematically expressed as $rB > C$ (Hamilton 1964).

Viewing cooperative behaviours through the lens of inclusive fitness, helping offspring and helping a relative are to be understood within the same explanatory framework: both are in one's *genetic self-interest* because both further the transmission of one's genes from one generation to the next (Hamilton 1964). However, it is not clear that inclusive fitness models can explain many of our moral behaviours relating to kin. Looking after one's genetically related parents or grandparents when they are beyond breeding age isn't explained by inclusive fitness models – the costs are high but the inclusive fitness benefits appear to be somewhat less. In such cases, our fondness for our non-reproducing relatives might just be an artefact of already developed familial-directed affective mechanisms that were selected for via kin selection. Furthermore, kin selection requires some means for preferential assortment of genetically related individuals, such as the ability to identify genetically related family members. This is not always plausible. More generally though, explanations for cooperative behaviours based on inclusive fitness are limited to groups of closely related organisms and therefore fail to explain cooperative interactions between unrelated individuals. As such, inclusive fitness is inadequate to explain the large-scale cooperation amongst non-related individuals witnessed in human populations.

1.4.2 Reciprocity

Cooperation with non-kin can be favoured by natural selection if the costs of cooperative actions are returned in the future. Robert Trivers proposed that the prospect of future reward could drive the evolution of cooperation between non-related individuals (1971; Axelrod 1984). Reciprocity (often referred to as direct reciprocity) of this sort is summarised in the well known saying, “I’ll scratch your back if you scratch mine”. Reciprocators are agents who will cooperate in the first instance and continue cooperating so long as their partner does so. If their partner does not cooperate or ceases to cooperate they will in turn refuse to cooperate any further. Reciprocators are first-shot unconditional cooperators and subsequent conditional cooperators. In the case of one-off interactions or the last interaction in a series of repeated interactions, it is in one’s interests to defect and receive the benefits of the other’s cooperative actions without incurring the costs – in this way reciprocators are selected against and defection becomes the favoured strategy. Yet this is not necessarily the favoured strategy in cases where there is a significant probability that the agents will meet again in repeated interactions. Here, one might receive higher payoffs by cooperating through repeated interactions rather than defecting and forgoing those cooperative benefits. Selection for cooperation based on reciprocity depends on an adequate number of repeated future interactions. In models of prisoner dilemma scenarios, the probability of an encounter between the same two agents needs to exceed the cost-to-benefit ratio from the altruistic act for cooperation to evolve by reciprocity of this kind (Nowak 2006).

Yet, models of cooperation based on reciprocity are limited in explanatory scope. Firstly, they still face the problem of defection because free-riding agents can move from player to player in a group, thereby accruing the benefits of defection without contributing the costs of cooperation. Individual-level selection therefore acts against the altruistic cooperative behaviours of reciprocators. One proposed solution to this problem is that agents have the ability to recognise each other, keep a track of all their past interactions, and preferentially cooperate with those whom they have previously successfully cooperated (I will address this type of reciprocity in 1.4.3). If such conditions are met, defection will not remain a viable strategy because free-riders will eventually run out of agents to interact with. Long-term cooperation will be selectively favoured over short-term defection strategies.

Secondly, reciprocity is limited to interactions in small groups (Bowles and Gintis 2003; Boyd and Richerson 2004). Reciprocity models do not scale up from two person dyadic cooperative interactions to those of n -person interactions involving cooperation between large numbers of agents. If reciprocators are cooperating in a group, where the cooperative activity is between all group members, then if one defection (or even a mistaken defection) occurs, all reciprocators in the group will cease to cooperate and everyone, including other non-defecting reciprocators, will be denied the benefits of future cooperation. Reciprocity is therefore not a sufficiently discriminatory mechanism to sustain successful cooperation in large groups. Evolutionary models by Boyd and Richerson (1988) and others have shown that reciprocity based cooperation can only be sustained in small groups, because in large groups of reciprocators, a small number of defectors will break down cooperation. These models of cooperation using n -person prisoner's dilemma show that cooperation is unlikely to be favoured by natural selection in groups of around 10 or more unrelated individuals (see also Henrich and Boyd 2001:79). Successful reciprocity also requires that interactions are repeated, which is not always possible in large, dynamic groups.

Thirdly, although we witness reciprocity in many of our dyadic relationships such as those between husband and wife, family, friends and some economic transactions, beyond these, evolutionary explanations based on reciprocity do not accord with much of the empirical evidence. Reciprocity fails to explain why agents in real world and experimental settings will cooperate with strangers in both non-repeated interactions as well as the final round of a repeated interaction. As Bowles and Gintis point out, in times of war and group threat for example, "precisely when a group is most in need of prosocial behaviour, cooperation based on reciprocal altruism will collapse" because the prospect of one's cooperation being repaid in the future is diminished and hence so is the incentive to cooperate (2003:434).

Finally, reciprocity, like kin selection, requires some means of preferential sorting, such as by way of agent identification. For reciprocation to work agents typically need to be able to preferentially cooperate with those who have previously cooperated and avoid who

have defected, as well keep track of past interactions.⁴ This is an unlikely scenario in the case of large groups where many interactions are often fleeting and the vast number of interactions makes bookkeeping of this sort impractical. In the human case, we do not have the requisite psychological (proximate) mechanisms for the evolution of cooperation under reciprocity alone.

It is most likely that kin selection and reciprocal altruism explain many of our cooperative behaviours. Although they may be limited in explanatory scope, they still tell us something about the types of proximate mechanisms that we might expect to see for the limited class of human cooperative activities they do explain. Kin selection requires an ability to identify family members and motivation to cooperate with them. Reciprocity requires some form of discriminatory bookkeeping mechanisms so one can remember who to cooperate with – or not – in future interactions. Neither kin selection nor reciprocity can adequately explain the evolution of cooperation in large groups of unrelated individuals because they rely on strong probabilistic assumptions involving positive correlation between types, population size, cost-benefit ratios, and numbers of repeated interactions. Their applicability is limited with respect to the large, dynamic, often random and non-repeating interactions we see in human populations. More recently, two proposed solutions to the problem of cooperation in large-scale societies have proven to be both more theoretically and empirically viable than either kin selection and/or reciprocity. The two solutions involve reputation based and punishment based strategies.

1.4.3 Reputation

We saw above that reciprocal altruism involves cooperative acts between two people which result in both gaining net benefit from the act. Any subsequent acts of cooperation are conditional on the outcome of previous interactions. What reciprocity does not take into account is an agent's cooperative track record with third parties. An agent's previous behaviour towards others is irrelevant to the decision-making process of the altruistic reciprocator. In contrast, reputation-based strategies present a more promising account of large-scale, *n*-player cooperation. Reputation based strategies for the evolution of

4. It should be noted that there are some situations in which agents can satisfy the cost/benefit structures of a reciprocal exchange even if they never encounter each other.

cooperation, called *indirect reciprocity*, have been proposed by Richard Alexander (1987).

Indirect reciprocity occurs when one does not receive the benefits of cooperation from the agent to whom one has behaved altruistically. Instead, one is the recipient of altruistic acts from some other agent. In a group of indirect reciprocators, agent A will help agent B who will help agent C and so on. Indirect reciprocity can be summarised as “You scratch my back, and I’ll scratch someone else’s” and “I scratch your back and someone else will scratch mine” (Nowak and Sigmund 2005:437). There have been various mechanisms proposed by which cooperation via indirect reciprocity can be achieved. ‘Image scoring’ is one such mechanism, whereby cooperating with others increases your image score, while refusing to cooperate with others decreases your image score. Discriminating group members can then preferentially cooperate with those who have a good image score over those who do not. Models show that under the right conditions cooperation can evolve via image scoring mechanisms (Nowak and Sigmund 2005).⁵ There is also empirical evidence that people do actually discriminate by image scoring (see Milinski (2002) and discussion below).

One problem with simple reputation based strategies is that although agents gain poor reputations for non-cooperation, those reputations don’t take into account the reason for non-cooperation. All acts of non-cooperation are judged the same. This means that one can gain a reputation as a non-cooperator for not cooperating with those agents with poor reputations! A proposed solution is to introduce the notion of a ‘good standing’, which individuals lose if they fail to help others with a good standing, but retain if they refuse to cooperate with others who have a bad standing (Leimar and Hammerstein 2001). In contrast to image scoring, this is a more discriminatory strategy because it is responsive to not only the cooperative reputation of agents, but also the reputations of those with whom they do or do not cooperate with. One’s standing is not damaged through non-cooperation with non-cooperators.

Like direct reciprocity above, these cooperative strategies are conditional strategies. But unlike direct reciprocity, indirect reciprocal strategies are based upon the other agents’

5. One of the proposed problems with the image scoring account is that agents damage their image score when they refuse to cooperate with other agents with poor image scores. In other words image scoring in this way doesn’t sufficiently discriminate between non-cooperation with good vs. poor image scorers. For discussion see Leimar and Hammerstein (2001) and Nowak (2005).

interactions with other agents. Indirect reciprocators preferentially cooperate with those who have cooperated with others in the past. Of central importance for the sustaining of cooperation in these models is reputation: agents cooperate conditionally with respect to another's reputation as a cooperator. Interactions witnessed by third party, non-participatory observers provide information to be used as part of their own decision-making process in future interactions. In this way, agents do not have to interact with other agents to establish their cooperative credential; they need only observe, or by some other means obtain the relevant information about, their behaviours.

This means that reputation has consequences for one's ability to either cooperate or defect on future interactions. A system of reputation-based reciprocity fosters individual concern for one's own reputation because it becomes in one's best interests to maintain a reputation and build up the trust of other agents. Failure to do so will result in the exclusion from future cooperative activities, not only with those with whom one has defected, but also all those who have heard of your defection. If one gains a poor reputation, one not only loses the potential to gain rewards from cooperative interactions, but also the rewards to be had from defecting in future interactions. Discriminating against agents with poor reputation acts as a form of punishment whereby they are denied the potential rewards of others' cooperative activity. Under such circumstances, selection can favour the evolution of cooperation where the decision to cooperate is based on the reputation of the interacting agents. Evolutionary models by Panchanathan and Boyd (2004) linking indirect reciprocity with norms of collective action have shown that the threat of exclusion from a system of indirect reciprocity can sustain cooperation in large groups. Helping others cultivates a good reputation. This in turn encourages others to cooperate with those who have good reputations and provides motivation to contribute to cooperative activity, even for purely self-interested agents.

Intuitively, reputation counts. We care about what other people think of us, whether we like it or not. And we seem to especially care about our moral reputations as well as those of others with whom we interact. Not only is indirect reciprocity theoretically and intuitively appealing, empirical evidence supports the claim that it plays a significant role in the maintenance of cooperation. In public goods games players are more generous to those whom they have witnessed contributing to the public good (Milinski, Semmann et al. 2002). People are also more generous to those whom they know have previously

donated to a charitable organisation (Milinski 2002; Nowak and Sigmund 2005). Additionally, we can see reputations playing a central role in online auction sites such as eBay. These use image scoring as a means of regulating cooperative behaviours by placing self-interested value on having a good reputation. In this way, reputation acts as a means of deterring defection and signalling trustworthiness to third-party potential cooperators.

In the above discussion, the assumption has been that individual agents are witness to the interactions of others. Clearly this cannot always be the case. Within groups of interacting agents, every interaction is not witnessed by all other agents, especially in large groups. Agents will not always be party to the information required to successfully employ the conditional strategies and maintain cooperation. For indirect reciprocity to work in large groups, there must therefore be some mechanisms which allow the relevant information to be distributed amongst the group, in what McElreath et al. (2003) call “distributed bookkeeping”. In this way, cooperatively relevant information is stored and transmitted via social relationships. Other agents become a central part of our moral and cooperative decision-making processes and the establishment of moral trust. This idea of social image, reputation and distributed bookkeeping accords well with our propensity for moral gossip: one is much less likely to cooperate with those one has had failed cooperative interactions with before, and people tend to tell others about this. As we will see later in chapters 5 and 6, distributed mechanisms such as moral gossip play a large and centrally important part in moral cognition.

Indirect reciprocity provides a sound theoretical and empirical basis for much of our cooperative activity, but it is not the explanatory silver bullet for the evolution of cooperation. Although language and culture provide relatively cheap means of increasing the probability of knowing someone else’s reputation through gossip (Dunbar 1996), the degree to which indirect reciprocity succeeds will be constrained by the nature of the population and the relationships between its agents: “In a fluid population, in which most interactions are anonymous and people have no possibility of monitoring the social scores of others, indirect reciprocity has no chance” (Nowak and Sigmund 2005:1296). Like the models of direct reciprocity, many indirect reciprocity models are also based on dyadic interactions which cannot account for n-player cooperative interactions (an exception to this is the previously mentioned work by Panchanathan and Boyd (2004),

which links reputation to norms of collective action).

However, indirect reciprocity is a significant evolutionary mechanism sustaining cooperative acts and is indicative of some of the proximate mechanisms facilitating cooperation. It requires mechanisms to make appraisals about the cooperative actions of others (like we do in moral judgement), track and memorise the behaviours and reputations of others, communicate that information to group members, and to perceive and judge one's self-image. Furthermore, culture, group structure, and the nature of the interactions and social relationships within groups all have a central place in the mechanisms that facilitate group cooperation through indirect reciprocity. As we will see in later chapters, group structure and agent interaction play a central role in moral cognition.

The mechanism of indirect reciprocity demonstrate how one's reputation, and hence one's cooperative prospects, can be altered depending on one's cooperative history. In this way, reputation can encourage punishment for failing to cooperate. The link with morality is clear. Moral reputation and associated punishments enforce moral norms which give rise to large-scale cooperation.

1.4.4 Punishment

One of the proposed solutions to the problem of cooperation is the introduction of targeted third-party punishment. An example of this is *strong reciprocity*, or what is often referred to as *moralistic punishment* (Trivers 1971; Henrich and Boyd 2001; Fehr, Fischbacher et al. 2002; Boyd, Gintis et al. 2003; Bowles and Gintis 2003; Fehr and Fischbacher 2004; Henrich, McElreath et al. 2006). Strong reciprocators are conditionally altruistic cooperators as well as conditionally altruistic punishers; they will initially cooperate and continue cooperating with other cooperators but will also undertake individually costly punishment of those who fail to cooperate with them, as well as on behalf of others. Importantly, they will punish even when there is no future payback for such punishment. The strong reciprocator's actions are a benefit to other group members (by punishing defection), even when at a cost to themselves, and therefore strong reciprocators are providing a public good (Boyd, Gintis et al. 2003; Bowles and Gintis 2003). If the punishment meted out by the strong reciprocator is sufficiently severe and

appropriately targeted, then the individual benefits of defecting diminish and the relative benefits of cooperating increase. Cooperation becomes the more profitable strategy. “Once the cooperative equilibrium becomes common, it is plausible that natural selection acting on genetic variation will favor genes that cause people to cooperate and punish because such genes decrease an individual’s chance of suffering costly punishment” (Henrich and Boyd 2001:87). Boyd and Richerson (1992) have shown mathematically that only a small number of strong reciprocators are required to induce group cooperative activity. Strong reciprocators can invade a population of self-regarding individuals and reach equilibrium (Bowles and Gintis 2003).

Intuitively, punishment-based models such as strong reciprocity appear to model real world phenomena and the intimate connection between punishment and morality. Not only do we typically punish those who commit moral infractions, we desire and are motivated to punish them. We get angry and want to punish those who betray us. Importantly, we also get angry and want to punish those who betray others, and not only when we have had relationships with those betrayed. Disgust at child abuse and slavery for example are not limited to instances in our own social groups, ethnicities, or countries. And neither is a desire to punish the perpetrators of such actions. Experimental evidence also shows that strong reciprocator behaviours are common. A large number of subjects will punish low contributors in public goods games at a cost to themselves even when there is no chance of recouping the costs, and even when they are only witness to the violation and not directly affected (Fehr, Fischbacher et al. 2002). Costly punishment is also a cross-cultural phenomenon. Experiments by Henrich et al. (2006) show that people from a diverse range of different cultures will similarly punish at direct cost to themselves when they are victims of non-cooperative norm violations. (For a brief overview and further evidence see Gintis 2000.)

However, punishment does not come without cost to the punisher, although this cost varies substantially depending on the type of punishment. Physical punishment, for example, can be a potentially dangerous affair. Ostracism by group members on the other hand is a comparatively cheap form of punishment. So too is the spreading of another agent’s poor reputation amongst potential cooperators. But neither is without cost. There are costs involved in the observation and monitoring of behaviours, and subsequent information dissemination. Additionally, these acts can elicit potentially

vengeful reactions from the recipient of the punishment. Cost can be significantly reduced if they are distributed over a number of punishers, whereby individual costs of punishment decrease with an increasing number of punishers. However, there is still a cost involved in punishment, no matter how small.

The fact that punishment is costly raises a cooperation problem which is often referred to as the *second-order defection problem*. Punishment may be an effective means by which to sustain various norms, but why should an agent undertake individually costly punishment? Agents who punish are providing a public good and are open to exploitation, thereby posing the second-order defection problem (Henrich and Boyd 2001; Boyd and Richerson 2004). Agents can cooperate, avoiding potential punishment themselves, but refrain from inflicting costly punishment on others (so called non-punishing cooperators). They are enjoying the public benefits of the strong reciprocators administering punishment on group members but are not prepared to bear the costs of punishment themselves. In this case the second-order free-rider benefits from higher payoffs than the punishers do. Therefore punishment is not an evolutionary stable strategy (Henrich and Boyd 2001).

There are a number of responses to this problem. In cases where there is a large (enough) number of prospective punishers, the number of first-order non-cooperators will be low, because as we saw above, the costs of defection become greater than the costs of cooperation. If this is the case, instances of first-order defection will be rare and the need to undertake punishment will itself be correspondingly rare. The costs of punishment will therefore be low as will the fitness benefits from being a non-punishing cooperator (i.e., second-order free-riders) since the probability of defection decreases as the cost of punishment also decreases. If this is the case, instances of second-order defection will be rare and, although there will be fitness advantages accrued to second-order defectors, these will be small relative to the strong reciprocators (Boyd, Gintis et al. 2003).

Another solution to the problem of second-order defection is the introduction of higher-order punishment. In this instance, punishment is meted out to those who fail to punish first-order defectors, imposing costs on being a non-punishing co-operator such that it is no longer in one's interests to be a second-order free-rider (Boyd and Richerson 1992).

Although evidence for higher-order punishment is rare, evidence from audience effects on the behaviours of third-party punishers is suggestive. Kurzban et al. (2007) found that in the presence of an audience, third party punishers are more likely to punish, and more likely to punish more severely. Reputational effects would appear to play a role in our propensity to punish norm violators, whether or not we are aware of these effects. If we consider that reputational mechanisms are closely aligned with punishment (as per above) then we can tentatively see how the threat of punishment (i.e., loss of reputation) may play a role in the maintenance of higher-order punishment. Still, the problem with meta-punishment strategies such as this is that they give rise to infinitely recurring orders of free-riding and corresponding punishment strategies. Transferring the cost of punishment over infinitely recurring levels seems an unrealistic mechanism by which to stabilise punishment (Henrich and Boyd 2001).

One further solution is that the selective forces favouring second-order free riding are overcome by other adaptive behaviours. If the relative costs of being a strong reciprocator are low, then it will only require relatively weak mechanisms to stabilise punishment. Henrich and Boyd (2001) argue that because of this, the propensity of humans to adopt the behaviours of others through learning can potentially prevent second-order free-riding. Humans use a variety of psychological mechanisms and social learning rules to adopt the adaptive behaviours of their groups (such as “copy the successful” and “copy the majority”). These mechanisms will also see the adoption of norms of behaviour that underwrite punishment. If the conformist transmission of punishing is sufficiently strong (and it need only be very weak) then it will counteract the small fitness advantage that exists in being a second-order free-rider and hence stabilise costly punishment. In this way, the within-group evolutionary forces of conformist transmission that drive other adaptive behaviours can stabilise punishment and sustain cooperative behaviours.⁶

Experimental evidence shows that institutions adopting sanctioning mechanisms that maintain cooperative behaviours have a competitive advantage over institutions that do not sanction. In an experiment by Gurerk et al. (2006), subjects were allowed to freely

6. Henrich and Boyd (2001) point out that the stabilisation of punishment itself is in fact a maladaptive side-effect of conformist transmission. i.e., it may be in one's individual self-interest to be a second-order defector, but the capacity for conformist transmission which has been selected for other adaptive reasons, means that we inadvertently copy the social norms that entail we punish defectors.

choose whether they joined a sanctioning or non-sanctioning group of potential cooperators. Each group then partook in a public goods game to establish conditions of conflict between self and collective interests within each group. Agents were allowed to freely move between the two groups after each round of the game. The authors found that over time subjects moved from the non-sanctioning institution to the sanctioning institution until nearly all agents had moved to the sanctioning group. The presence of free-riders in the non-sanctioning group resulted in the collapse of cooperation, thereby reducing the potential benefits derived from cooperation as well as the number of cooperators which the free-riders could exploit. Additionally, in the sanctioning group, punishment increased the group contribution of agents by making free-riding more costly. As such, it became more profitable to be a non-free-riding cooperator in the sanctioning group than a free-rider in the non-sanctioning group.

In sum, strong reciprocity is one mechanism that can maintain cooperation with non-related individuals in large groups with both dyadic and non-dyadic interactions and is intuitively, theoretically, and empirically sound. The link with morality is clear. Punishment is mediated by norm recognition and recognition of norm violation, both of which are central to morality. In moral terms, punishing minds are moralising minds.

1.4.5 Group selection and norms of collective action

Mechanisms based on reputation and punishment can explain the stabilisation and maintenance of cooperative behaviours in the face of defection. However, neither mechanism is sufficient to explain why it is that we see *specifically* cooperative behaviours on such a large scale. This is because, as Boyd and Richerson (1992) have shown, punishment-based mechanisms can stabilise just about any norm of behaviour: “if evolutionary change is driven only by individual costs and benefits, then moralistic punishment can stabilize cooperation, but it can also stabilize anything else.... Since cooperative behaviors are a tiny subset of all possible behaviors, punishment does not explain why large-scale cooperation is so widely observed” (Boyd 2007:201).⁷ Richard Joyce also explains: if “a good reputation means sharing food indiscriminately with the group, then an indiscriminate food-sharing trait will develop; if a good reputation means

7. This is an especially important point when we consider the vast array of moral norms and behaviours with which we find across a multitude of different domains.

wearing a pumpkin on your head then a pumpkin wearing trait will develop” (2006:33). And similarly, Joseph Henrich tells us that “the same reputational and sanctioning mechanisms that can stabilize cooperation can also sustain maladaptive practices such as consuming the brains of dead relatives, flattening the foreheads of infants, or binding the feet of young girls” (2006:60). Cooperative norms of behaviour are only a fractionally small portion of all possible norms that exist, so we must ask why we see a disproportionate profusion of group beneficial prosocial norms as opposed to say group harmful norms. The reason is, quite simply, that groups with prosocial cooperative norms will out-compete those groups with neutral or harmful norms (Bowles and Gintis 2003; Boyd and Richerson 2004; Joyce 2006). This is group selection.

Natural selection operates on variation to produce adaptive change. Within populations individual agents vary in their phenotypic traits which will often differentially contribute to the survival and reproduction of bearers of those traits. Selection in this case occurs between individuals within the population but can also, under the right conditions, occur *between* groups. In cases where there is heritable variation between groups, sufficient and sustained homogeneity within groups and competition between groups, then selection will act on between-group fitness differences. In most cases, these conditions will not be adequately met. The main problem with group selection when cashed out in terms of genetic selection, for example, is that it is hard to maintain variation between groups when there is even small migration between those groups (Boyd and Richerson 2004). But if we understand group selection as being driven by culture and various within-group mechanisms that resist the corruption of within-group similarities, we can see how group selection can occur. Cultural learning, for example, helps explain both between-group differences and within-group similarities required for such selection.

Between-group conflict is common.⁸ While some forms of inter-group conflict are direct conflict, such as in the case of warfare, there are also non-direct conflicts whereby one group will out-survive another in times of famine and pestilence. Groups can also compete against each other through cultural practices such as hunting and food gathering. One group out-competes another when the latter group becomes extinct or

8. Selection between the groups does not necessitate the extinction of the individuals within one or other of the competing groups – only that those individuals from the losing group either adopt the practices of the more successful or disperse, thereby maintaining the winning group’s integrity and intra-group homogeneity (Boyd and Richerson 2004).

adopts the behaviours and practices of the former. In many instances, groups that are cooperative will have a better chance of survival than those that are not. As we also saw above, there are a number of mechanisms that can buffer within-group variation but these mechanisms can also amplify and sustain between-group differences. Human traits such as imitation and conformist learning, moralistic punishment, and normativity do just that. As Kim Sterelny points out, these mechanisms magnify “the influence of group selection by increasing the variation between groups and decreasing it within groups, for in different groups, different innovations appear and are picked up ... Once the conditions for group selection are in place, between-group selection will act powerfully in favour of cooperative groups” (2003:127). Within-group cultural evolution will mean different groups reach different behavioural equilibria with different fitness payoffs. Selection at the level of groups operates on these between-group fitness differences. Where cooperative behaviours positively contribute to group fitness differences, inter-group competition will favour mechanisms that give rise to those cooperative behaviours. Hence, we see the profusion of norms that promote group beneficial cooperative behaviours (Henrich and Boyd 2001; Boyd and Richerson 2004; Henrich 2006). There is obvious conflict between individual-level selection operating within the group, and group-level selection operating between groups. But as we have seen above, mechanisms such as punishment can work to mitigate the disparity between the two and allow weak selective processes such as group selection to select for group beneficial mechanisms. Theoretical mathematical models give credence to this picture by showing how cultural group selection can favour particular group-beneficial behaviours over non-cooperative behaviours (Boyd, Gintis et al. 2003).

Earlier we saw that systems based on reputation and punishment provided explanations as to how group cooperative norms are maintained. What they didn't explain was why there is a profusion of group cooperative norms as opposed to non-cooperative norms. Group selection provides us with an explanation of why we have specifically group beneficial cooperative norms such as many of our moral norms.

1.5 The evolution of cooperation and the proximate mechanisms of morality

None of the above explanations – kin selection, reciprocity, reputation, punishment and cultural group selection – are complete explanations for the evolution of cooperative behaviours. It is most likely that they all play a part, possibly in addition to other, as of yet, less thoroughly investigated evolutionary mechanisms of cooperation, such as sexual selection (Miller 2007) and costly signalling (Gintis, Smith et al. 2001). Importantly though, they provide some insight into the proximate mechanisms that are involved in cooperative action. Although there is cross-over, each of the mechanisms discussed select for some different proximate mechanisms. When we think of human populations, kin selection requires, among other things, a way of identifying family members. Reciprocity requires some means of identifying cheaters⁹, bookkeeping and the ability for future discrimination. Reputation-based strategies again require capacities for cheater detection, tracking behaviours and intentions over time, agent discrimination skills, awareness of one's own and others' reputations, and group-wide communicative systems. Punishment-based strategies require various methods of punishment, cheater detection, and special motivational mechanisms. Group selection requires transmission of cultural traits and mechanisms that maintain within-group homogeneity and between-group differences such as punishment and conformist mechanisms. Many require an ability to identify and follow norms of behaviour and all require some form of motivational mechanisms such as emotions. These are just a sample of the proximate mechanisms involved in cooperative activity.

Morality acts in aid of cooperation. And if morality is, as I have claimed above, constituted by a range of mechanisms that sustain cooperative activity, then many of those mechanisms are going to be central to moral cognition. We have seen individual and group-level traits that arise via multiple levels of selection. Kin selection, for example, involves the selection for traits that benefit groups of related individuals, while cultural group selection involves traits that benefit groups of unrelated individuals. We can see how many of the proximate mechanisms that sustain cooperation are recognisably central to our moral lives: punishment, norms, reputation, judgement and

9. See, for example, the work by Tooby and Cosmides on cheater detection (1989; Cosmides and Tooby 1992).

gossip. Morality is not a single trait, like, for example, colour vision. It is composite. This plurality of mechanisms helps explain the existence of our conflicting moral intuitions. Ought one to steal food to feed their family if they are hungry but cannot afford food? Ought one turn in a close friend to police if he has committed a crime? The mechanisms that drive preferential support for kin might conflict with one's adherence to group beneficial norms. One's reputation amongst friends (depending on who your friends are) might also conflict with one's belief that it is a moral duty to be a law abiding, group beneficial citizen.

What is also clear is that many of the adaptive mechanisms that sustain cooperation involve mechanisms external to any one individual agent. In some cases we see individual benefits, and in other cases group benefits, involving both individual and group-level adaptations. In the case of the group, extending the 'object of selection' beyond the organism places special importance on beyond-the-organism environments in our explanations. But, as Dawkins (1982) has made clear, adaptations need not be traits that reside in an individual organism. The beaver's dam is an adaptation that resides outside the bodies of many beavers. Morality (and moral cognition), I will later claim, is no different in this respect. It emerges from the integration of various disparate mechanisms that span across brain, body and world. These are causally integrated into our moral problem-solving routines in such ways that they are best considered an extended part of our moral cognitive processes.

1.6 Conclusion

Morality is an adaptation. We are moral agents and have the social-wide moral systems that we do because being moral conferred adaptive advantage on our ancestors: morality primarily allowed the formation and sustaining of cooperative social relationships. The evolutionary mechanisms that gave rise to this are numerous, as are the various proximate mechanisms that underwrite individual and group cooperative behaviours. Understanding evolutionary mechanisms works in tandem with evidence from moral psychology, developmental psychology and other cognitive sciences. Jointly, these provide the tools to construct a more comprehensive understanding of those phenomena.

1.7 Summary

In this chapter I have argued that morality is an adaptation and that moral cognition involves a suite of proximate mechanisms that gave rise to and sustained cooperation in our evolutionary past. By looking at the evolutionary mechanisms that sustained cooperation we can gain some insight into those proximate moral mechanisms that evolved in aid of cooperation. The main points from this chapter are as follows:

- Morality is an adaptation primarily selected for as a means of generating the fitness benefits from cooperative interaction.
- Morality is an individual and group-level adaptation involving multi-level selection
- Kin selection, reciprocity, reputation, punishment and group selection are amongst the evolutionary processes that gave rise to morality.
- These evolutionary processes inform the nature of the proximate mechanisms that underwrite morality.
- Evolutionary considerations suggest that moral action is sustained by several proximate mechanisms, not just one.

In the following chapter I will reject one particular but very influential argument in support of the claim that morality is innate. This in turn raises the issue of how morality, if it is not innate, can be an adaptation. I will address this issue in subsequent chapters.

2

The Wealth of the Moral Stimulus

What is in our minds, nativists argue, is richer than what could be learned from experience. Hence, nativists conclude, what is in our minds must be innate.

(Cowie 1999:31)

It *has to be* the case that the human mind has certain innate capacities that enable us—but not chimpanzees, dolphins, or parrots—to see certain moral distinctions and appreciate their significance for our lives and the lives of others ... there *must* be some innate structure to guide which bits of experience are taken on as part of one's moral knowledge.

(Hauser 2006b:297-298, italics added)

The advent of evolutionary psychology has witnessed the proliferation of nativist arguments about a great many of our psychological capacities (see for example Barkow, Cosmides et al. 1992; Wright 1994; Pinker 1997). Moral psychology has not been immune. In fact, moral psychology appears ripe for nativist hypotheses: it is a human universal, it appears relatively early in human ontogeny and, as was shown in the previous chapter, it is likely to be an adaptation. But not all evolutionary hypotheses need be nativist hypotheses (for recent examples see Sterelny 2003; Boyd and Richerson 2004). It is my contention that, even though morality is an adaptation, our capacity for moral cognising is not innate. This will have important implications for the type of evolutionary explanations that we give.

In this chapter I will argue against one of the central arguments for modern moral nativism, namely the poverty of the moral stimulus (POMS) argument. We will see that much of the evidence used in support of an innate moral faculty can be explained

without appealing to any moral domain-specific structures or bodies of information. This gives us good reason to doubt the moral nativist's positing of an innate moral faculty.¹ I will proceed by outlining the basic disagreements between nativists and empiricists before detailing the specific argument moral nativists have employed in defence of the view that we have an innate moral faculty. This argument rests on the claim that there is insufficient information available for the learning child to acquire the moral competence that they display and the requisite information must therefore be innate. There are two facets to this argument which I will address. The first rests on the claim that development is largely invariant to changes in informational exposure. The second is that the move from data to output is an especially difficult one. Neither claim is sufficiently supported. Finally, I will take a close look at one nativist argument that focuses on the specific criterion judgements that we make about moral and conventional norms. In particular, how children recognise that these norms differ in their dependence relations with authority. Some moral nativists claim we could not learn this relationship. I will argue that nativists have mischaracterized this relationship and that the learning task is not as difficult as proposed. The distinction between moral and conventional norms can be learned.

2.1 Nativism vs. empiricism

Nativism is the view that particular traits are innate. Nativism about the mind is, accordingly, the claim that some particular psychological/mental capacities are innate. Empiricism about the mind, on the other hand, claims that particular psychological capacities are learned rather than innate traits. No one in the current nativist vs. empiricist debates about the mind claims that all and every mental trait is innate, or that all and every mental trait is learned. The distinction is one of degree, rather than being absolute. Clearly some traits are in fact innate. Children show a remarkable visual sensitivity to the faces of others within a few hours of birth. Because it is visual it cannot be learned prenatally, nor is it likely to be learned in the short time after birth. Children must therefore be born with some innate visual information (Plotkin 2007). Yet they also show remarkable ability to acquire novel and often complex knowledge, such as learning

1. It should be noted that I am not arguing against nativist accounts of the mind per se. I am only arguing against a nativist account of the moral mind, which is still perfectly consistent with all other mental capabilities being underwritten by innate, domain-specific faculties – however likely or unlikely that may be.

the rules of games like chess and cricket at an early age. No one thinks that we have innate knowledge of chess or cricket. It is clear that the entire mind is not a *tabula rasa*, nor is it wholly innate. Modern debates between the nativist and the empiricist are not about whether the mind is innate, but what particular aspects of the mind are innate and what are acquired experientially.

One of the problems confronting this debate is that there is no clear definition of what it means for something to be considered innate. The literature on both our folk and scientific concepts of innateness is extensive and growing (see for example Bateson 1991; Samuels 2002; Griffiths 2002; Griffiths, Machery et al. 2009). Paul Griffiths (2002) has claimed that innateness is a “confused notion”, while Matteo Mameli and Patrick Bateson (2006) have identified no less than 26 different candidates for the scientific conception of innateness. Table 2.1 gives a summary of many of those surveyed by Mameli and Bateson. Even though there is no consensus on a single conception of innateness, it does not mean that debates about nativism need to be postponed until we have one. The moral nativists we will encounter below often appeal to more than one of these conceptions of innateness. Marc Hauser (2006), for example, claims that our innate moral sense is genetically determined, follows regular patterns of development, is not learned, and is a species-typical Darwinian adaptation. By looking at some of the specific claims that the moral nativist makes, we can critique their nativism. Furthermore, we have many clear, intuitive examples of what we consider innate. Hair, arms, hearts, and legs are innate, as is our ability to visually locate faces. Language, considered by many to be the paradigm case of mental nativism, is also innate.² In contrast, the making of clothes, riding a bicycle, playing cricket and playing the flute are not innate. Although these examples do not define the necessary and sufficient conditions for innateness, they do provide us with some idea of the terrain. They vary in both the types of innate structures they posit and specific contributions that the environment makes to the establishment of the trait. Is moral learning, and the structures and processes that underwrite that learning, more like the case of language or is it more like riding a bicycle? To answer this question we need to take a close look at the specific commitments the nativist makes to the moral acquisition process.

2. Although see Cowie (1999) and Tomasello (2003) for sceptical views on language nativism.

Table 2.1 Conceptions of innateness

Basis of innateness	Conception of innateness	Some problems
Non-acquisition and regularities in development	Innate traits are not acquired; they are present at birth and/or appear at particular developmental phases.	All traits are acquired in some sense, including prenatal traits. Regular development is still compatible with a trait being learned.
Genetic influence	Innate traits are genetically influenced or genetically encoded. Development doesn't involve extraction of information from the environment.	All phenotypes are genetically influenced, including learned phenotypes. It is hard to account for the distinctive contributions genes make. It is difficult to distinguish between environments which are developmental supports vs. those that make specific information contributions.
Heritability	If heritable phenotypic variation within a population is dependent on genetic differences then a trait is innate.	If the innate status of a trait depends on the statistical composition of the population, then changes in the population can determine the innate status of a trait, irrespective of whether individual organisms have themselves changed. Heritability of traits is also compatible with learning.
Lack of learning and adaptive plasticity	Innate traits are not learned.	All traits have some environmental influence. It is difficult to differentiate between learning processes and those processes which merely trigger the expression of innate traits.
Non-malleability	An innate trait is robust in terms of its development and is difficult to modify it once it has developed.	All traits, including learned traits, are to some degree insensitive to environmental variation. There is no principled way to decide the degree of malleability an innate trait is permitted.
Species typicality and adaptation during evolution	Innate traits are species-universal traits. Innate traits are Darwinian adaptations.	Issues with what is species typical. Queen bees, for example, are not typical members of their species. Learned traits can be universal. Adaptations need not be genetic, some adaptations are learned.

This table summarises some of the concepts of innateness presented in Mameli and Bateson (2006). For more detailed analysis and examples of these and many others, refer to original article.

The nativist vs. empiricist debate can be broadly framed around two central claims. The first is about the division of labour between genetic and cultural factors in the development and expression of a particular trait. Both sides agree that genetic and experiential factors make important contributions to phenotype expression. Yet fundamental disagreements exist with regard to the extent these factors contribute to the learning situation: the issue is not about whether there is a division of labour, but where the division of labour resides.

The second claim is to do with the mechanisms involved in that division of labour. Both the nativist and empiricist think that we have innate learning mechanisms, but have different views as to the nature of these mechanisms. Nativists believe that the acquisition of a particular trait requires some specific machinery dedicated to that acquisition. There are different suggestions as to the makeup of those innate structures; perhaps they are concepts, representations, informational biases, neurological structures, or functional modules. Either way, what is central to the nativist's claim is that certain learning tasks are skill specific, requiring innate learning mechanisms that are task- or domain- specific: "each is peculiarly adapted to the performance of one of the various learning tasks with which we will, as we mature, be confronted" (Cowie 1999:90).³ Empiricists on the other hand stress the generality of our innate learning mechanisms: the mental machinery involved in learning a specific capacity is not confined to learning in that specific domain.

One of the mistakes nativists often make in their criticism of empiricism is the assumption that empiricist learning is purely an inductive problem (Sripada 2008a; Sterelny forthcoming). I take 'induction' in this instance to mean "reasoning from observation alone". For example, in her criticism of moral empiricism, Susan Dwyer writes: "How does she then, *just by observation*, learn that some rule-governed regularities are merely conventional ... while others are moral" (2006:240-241, italics added). As we shall see, moral acquisition is not a purely inductive task. It involves active 'learning by doing' through embodied agent-environment interactions, biased cognitive mechanisms

3. Domain specificity can be read as either a claim about the specialisation of proximate mechanisms, or a selective claim about specific adaptive problems for which the mechanism was selected to solve. Although the moral nativists addressed here typically talk of the former, I take the moral nativist's position to endorse both claims.

and highly structured moral environments with explicit directives from caregivers.

We now have a picture of the moral nativist's position: to some degree our moral capacities are the product of an innate moral structure that makes a specific and necessary contribution to the moral acquisition process. It is not enough for the moral nativist to claim that biologically inherited faculties are employed in the task of moral acquisition, those faculties **must** be task- or domain- specific: they must be devoted to operating in the domain of morality. The moral nativist is "committed to the existence of some cognitive mechanisms that are specific to the domain of morality. These we term the moral faculty" (Hauser, Young et al. 2008a:172).

Moral nativists differ in their views on the nature of the moral faculty. A number of authors have recently broadly identified three kinds of nativist hypotheses about the moral mind. (For various examples of these and some critiques see Nichols 2005; Joyce 2006; Hauser 2006a; Mikhail 2007; Hauser, Young et al. 2008b; Prinz 2008a; Sripada 2008a). One proposal is that the moral faculty contains specific information or propositional knowledge in the form of moral imperatives such as "Do not steal" or "Do not kill". Such a view is not widely held. It would mean that there are very tight constraints on the impact that culture and development could have on specific moral behaviours and there is too much variation in our moral norms and moral development to make it a likely hypothesis. Another option is that the genetically endowed moral faculty contains universal moral principles to which culture sets the various parameters. These cultural parameters are said to determine individual moralities and the various moral systems that human populations have (see Hauser 2006a, 2006b; Dwyer 2006; Mikhail 2007). A third proposal is that the innate moral faculty gives us, minimally, the capacity to make moral judgements – to see the world in terms of right and wrong. One recent proponent of this view is Richard Joyce, who maintains that it is the concept of a moral norm that is innate: the moral faculty "enables us to categorize the world in morally normative terms" (2006:131).

These hypotheses can all be seen as differing along the two central claims discussed earlier. The division of labour varies between the three options above from very little input from culture in the first example, increasing through to a substantial cultural

contribution in the last. We can also see variations in the type and richness of innate structure that they posit. The first case posits explicit propositional knowledge; the second posits underlying schematically structured principles; while in final example, we see the positing of innate moral concepts.

With these possibilities in mind, we can now turn to an analysis of the moral nativist's argument in favour of the moral faculty, where the focus is on moral learning. With the mental and environmental resources that we have, how can we come to be the moral creatures that we are? Moral nativist Marc Hauser states the issue as thus: "what issues like this boil down to is a question about the relationship between a learning mechanism and the input it grabs on to.... The empirical question here is whether this capacity could be acquired through observation or teaching" (2006b:302). I will now outline the structure of one specific argument in favour of an innate moral faculty, namely the poverty of the moral stimulus argument. I will then proceed to critique this argument and conclude that we have reasons to reject the moral nativist's position (see Box 2.0).

2.2 The poverty of the moral stimulus

Arguments in support of nativism typically consist of two threads involving a negative and a positive thesis. The negative thesis involves the denial of domain-general learning as a plausible explanation for the acquisition of the trait in question. If domain-general learning cannot account for acquisition, then we are led towards a nativist conclusion (Cowie 1999). The positive thesis involves positing the requisite innate, species-universal acquisition mechanism with which to account for the particular trait. The negative and positive theses can be seen at play in the paradigm nativist case, language. Chomsky's famous poverty of the stimulus argument is based on the claim that there is inadequate data in the environment for a child to learn the complexities of language required to achieve mature competence. Therefore, the requisite data must be innate. This data comes in the form of a universal grammar which specifies the structural information required for language acquisition in the impoverished environment (Chomsky 1986; Pinker 1994).

The poverty of the stimulus family of arguments draws attention to a supposed

explanatory problem: that our experiential input is inadequate to account for particular mental resources that we have. There is nothing in our experience that could account for some of our cognitive abilities so we must appeal to some form of innate, domain-specific information or structure. Poverty of the stimulus arguments combine claims about development and task analysis with respect to a particular task- or domain- specific activity (Sterelny 2003):

- The developmental claim relates to the ontogenic expression of a particular mature competence in question – although dependent on information to trigger expression, it is relatively independent of variance in informational exposure: change in developmental environment does not necessitate a difference in development of competence.
- The task analysis claim “takes mature competence in the focal domain to rest on a theory of that domain, and argues that the task of generating a theory from perceptual evidence is an excessively challenging one” (Sterelny 2003:195). The move from perceptual data to theory formation can only be accounted for via innate faculties.

Recently, philosophers and psychologists have drawn strong analogies between language and morality. These ‘moral grammarians’ have exported aspects of the Chomskyan linguistics research programme to the domain of morality. Here again, we see the negative and positive threads of the nativist argument. The moral grammarians claim that the informational exposure in the environment is inadequate to explain moral competence through domain-general learning alone. The gap between informational exposure on the one hand and achieved moral competence on the other cannot be bridged by domain-general learning mechanisms and the environmental resources available to the moral learner. Therefore, some form of innate moral information or structure must be posited to bridge the gap between exposure and competence. This is the crux of the poverty of the moral stimulus (POMS) argument. Analogously, they also argue that there is an underlying Universal Moral Grammar, based upon the Principle and Parameters model in linguistics, which can explain the universality, diversity, compositionality, and computational nature of moral acquisition and competence (Dwyer 1999, 2006; Harman 2000, 2008; Mikhail 2002, 2007; Hauser 2006a, 2006b; Hauser,

Young et al. 2008b).

Although the focus of this chapter will be on acquisition, the structure of moral competence is relevant. How much evidence is required in the learning environment will depend on what one has to learn, because the possibility of moral competence through learning will depend in part on the nature of that competence. If, for example, one thinks that moral competence involves the possession of abstract principles, as the moral grammarians do, then the learning environment needs to contain the relevant information to learn those principles. If the learning environment is impoverished with respect to this information, the innate components of the moral mind must contain the relevant information. One of the reasons for focusing on acquisition in this chapter is that the POMS argument is not confined to specific structural claims of the moral grammarians. It is endorsed, in some form, by all moral nativists (see for example, Joyce (2006)). (For detailed criticism of the structural parallels between language and morality see Dupoux and Jacob 2007.; See also Prinz 2008a; Sripada 2008a; Sterelny forthcoming.)

Box 2.0 Against the poverty of the moral stimulus

Since the advent of the poverty of the moral stimulus argument there have been a number of critiques directed at it on both empirical and theoretical grounds. Shaun Nichols (2005) argues that domain diverse emotions play a special role in generating non-conventional judgements (including moral judgements). Jesse Prinz (2008a, 2008c) looks at the role that emotional consequences of norm violations and the differences in the enforcement of moral and conventional norms play in the moral acquisition process. These he argues, provide sufficient learning resources for the non-nativist moral child. Chandra Sripada (2008a, 2008b) maintains that compared to language acquisition, the learning target in moral learning is less complex. He also argues that moral learning, unlike language learning, need not be inductive; it often involves explicit instruction, whereas language learning does not.

I will, among other things, provide further analysis of these ideas as well as the implications of further empirical evidence on the poverty of the moral stimulus argument.

The acquisition argument, in the form of poverty of the stimulus arguments, has been, either implicitly or explicitly, central to recent defences of moral nativism.⁴ A significant piece of evidence cited in support of the POMS relies on the early emergence of sensitivity to a distinction between moral and conventional rules. Evidence for the moral/conventional distinction came from research into social domain models of social cognition, in which morality was identified as a distinct developmental and conceptual domain (cf. domain-specificity above). All children, except those with psychopathic disorders, show proficiency at drawing the distinction from an early age along the same dimensions as adults do. Moral norms are typically identified as more authority independent, more generalisable, more serious than conventional norms, and justifications in the moral domain typically appeal to issues of harm and welfare. Children show this ability at an early age, some younger than three years old. (Turiel 1983; Turiel 1998; Smetana 2006; Tisak, Tisak et al. 2006).⁵ This is taken as evidence by the proponents of the POMS argument that firstly, morality is a distinct domain of social knowledge, and secondly, that children's proficiency at this task indicates that moral competence comes on-line at an early age, a level of competence beyond anything that the child has been taught (Dwyer 2006; Hauser 2006a, 2006b; Joyce 2006; Mikhail 2008b). I will argue below that the evidence offered by the nativist in support of the POMS argument, including proficiency at the moral/conventional distinction, is not adequate to establish the argument for nativism.

In terms of the developmental claim, the POMS argument asserts that moral competence comes on early in life, too early to be explained by reference to the information exposure that children receive. Further, moral development is invariant to differences in moral information exposure in the child's environment. All normal children attain moral competence in varying moral environments. With respect to task analysis, the POMS argument makes two claims. Firstly, a child's moral learning environment is noisy; it is awash with norms that are not explicitly nor adequately identified as moral or

4. From here on, I will understand the term 'moral nativist' to refer to those who employ the poverty of the moral stimulus argument. It is an interesting and open question as to whether all nativist arguments necessarily employ poverty of the stimulus type arguments, but not a question that will be addressed here.

5. It should be noted that some authors have recently questioned the moral/conventional distinction. Kelly and Stich (2007; Kelly, Stich et al. 2007) point out that the class of transgressions over which it has been tested has been very narrow. They provide evidence that the distinction is not robust over a larger class of transgressions. This seriously questions the basic datum of the moral nativists who appeal to the moral/conventional distinction in support of nativism. See also Prinz (2008a).

conventional, so the evidence is inadequate to explain the acquisition of the capacity to draw the moral/conventional distinction. Therefore, the information available to the child from the environment alone is impoverished. Secondly, the complexity of the learning task is especially difficult because “to solve it one must identify a class of considered moral judgements and a set of principles from which they can be derived” (Mikhail 2008b:353-354). The task of rule extraction from the surrounding environmental data is too complex to be accounted for in terms of empiricist learning alone because “empiricist accounts radically underestimate the complexity of the task that faces the young child with respect to rule recovery” (Dwyer 2006:239). Moral nativists also often claim that the principles that guide our moral behaviour are unconscious and cognitively inaccessible, and therefore unlikely to have been learned. (As I will explain in section 2.4.3 below, I see this last claim as being, in part, about the complexity of the learning task.) In summary, we have an informational gap that is due to both the inadequate resources available to a non-nativist moral learner and the complexity of the moral learning task. According to the nativist, this gap between informational input and behavioural output can only be bridged by some form of innate, domain-specific, information or mechanism – i.e., the said moral faculty.

I will now argue against the POMS argument by looking at both the developmental and task analysis claims, showing that the moral stimulus is not impoverished; there is no gap between the informational input of experience and the output of behaviour. Before moving to this analysis, there is an important point to note with respect to a non-nativist account of moral acquisition: there are no constraints on the nature of the inputs and cognitive mechanisms of moral acquisition except that the internal cognitive resources must not be specific to the moral domain. In other words, any non-moral cognitive capacities (be they innate or otherwise) and any external (including moral) resources can freely operate as inputs and mechanisms in the acquisition process. To this extent, the moral non-nativist need not claim that the mind is a *tabula rasa*, only that the moral mind is.

2.3 The poverty of the moral stimulus: development

The developmental trajectory with which a particular trait develops can often be indicative of the trait’s innate status. Again, take the paradigmatic nativist case, language.

Linguistic competence arrives rapidly relative to environmental stimulus, and linguistic development is independent of variations in linguistic informational exposure. It is claimed that the attainment of moral competence is analogous. All morally competent adults are believed to make the moral/conventional distinction. Very young children's ability in making this distinction is touted as evidence that moral competence comes on very early, outstripping the environmental inputs available. Susan Dwyer writes:

The capacity to distinguish between different socio-moral normative domains and the heightened sensitivity to permission rule violation appear to be central aspects of adult human moral competence. These capacities do not represent a sort of proto-morality limited to childhood. Rather it would appear that, over a remarkably short period of time, human children acquire moral capacities that are shared with adult members of their communities. (2006:239)

And the developmental trajectory is said to be invariant to variations in moral experience. Marc Hauser writes:

If experience really drives the ability, then it should be possible to accelerate the timing of its expression by early training, giving children additional exposure to permission rules. If, on the other hand, the capacity to detect violation is part of our innate moral faculty, then children living in different cultures, with wildly different experiences in school and home, should emerge with this capacity at around three years of age. One signature of an innate faculty is a narrow time window for expressing a skill that is relatively immune to differences in experiences. (2006b:303)

The developmental trajectory of moral acquisition does not give the nativist grounds for their case. When we look at the broad developmental trajectory of the moral child, we can see that it corresponds to increasing levels of complexity, and in comparison to language, over comparatively longer time-scales. What we see is that children as young as two and a half years old can draw the distinction between moral and conventional transgressions. However, this applies only for a very small class of transgressions, namely those involving physical harm and welfare but not psychological harm, and is less consistently identified with transgressions involving fairness (Smetana 1981, 2006). In middle childhood, development progresses to understanding fairness associated with

equality between people. As the child moves into preadolescence, concerns with equality broaden into notions of equity taking into consideration people's different needs and status. Then, during adolescence, moral concepts such as fairness become "more broadly comprehensive, universally applicable, and generalizable across situations" as well as more context sensitive (Smetana 2006:124). There is change in the child's moral capabilities over the course of ontogeny, and this change is slow. Children do not start off life as mature moral agents; they progress through gradual stages of increasing complexity over time.

This is important for several reasons. Firstly, it may be the case that "over a remarkably short period of time, human children acquire moral capacities that are shared with adult members of their communities" (Dwyer 2006:239) but these capacities are very limited in scope in early ontogeny. It is true adults and children both make the moral/conventional distinction, but the range of norms over which adults make this distinction is vast. Very young children on the other hand, only make such distinctions over a limited range of norms, namely emotionally salient instances involving harm and welfare. This is important because at this early age non-moral emotional abilities aid the categorising of moral, as distinct from conventional, norm violations (Nichols 2004, 2005; Dunn 2006; Smetana 2006; Prinz 2007, 2008b). As I will detail in section 2.4, emotional responses do this by providing different information on the two types of norms. Very early competence at the moral/conventional distinction is more indicative of emotional abilities rather than any specific moral competence. This raises the question of when one can legitimately say that the child acquires the moral capacities of the adult. From the picture of general moral development above, full moral competence, including the ability to generalise complex moral concepts and principles, would appear to come late. This is in contrast to language, where children show near-adult competence at early ages.

The developmental trajectory is also important because the longer the learning child has to achieve moral competence the more they experience relevant informational exposure. This will include first-hand moral experience; being the subject or perpetrator of a moral violation and any subsequent consequences such as parental admonishments. In addition, they also receive second-hand moral experience, such as the many moral stories and fables children encounter throughout childhood (see section 2.4.2 below).

Finally, the longer the acquisitions period, the more time the learning child has for the development of non-moral mental capacities (Prinz 2008a) and the arrival of certain concepts with which to aid domain-general learning. This goes some way towards explaining why there are limitations on accelerating the timing of various moral competencies through extra moral training. It is no surprise that children apply concepts of welfare to physical harm much earlier than they do to those which entail psychological harm (Smetana 2006:124). To identify the latter, children need to be able to reason about the effects of acts upon the often opaque psychological states of others. Such abilities will require, among other things, theory of mind to have developed. Children cannot, for example, expand their notions of equality to equity until they understand that people are not all the same. But, once they have this understanding, it allows that development and expansion of other concepts and abilities. Here we can apply Dennett's (1995) concept of a crane: products of one process enable and amplify the development of other processes. In this case, we acquire cognitive machinery which helps build further cognitive machinery.

Children do not follow a fixed developmental timeline; some children develop faster than others and gain moral competencies earlier. As we will see below in section 2.4.2, children vary over ontogeny in moral competence, including that ability to make the moral/conventional distinction, due to variations in informational exposure. For example, studies show that parental disciplinary practices affect children's moral development. Children whose parents undertake other-orientated reasoning during disciplining commit fewer moral transgressions than those whose parents undertake power-assertive techniques (Smetana 1997; Thompson, Meyer et al. 2006). Psychologist Judith Smetana concludes, "children's failure to apply moral criteria to moral events and differentiate morality from social convention can be attributed at least partly to differences in parents' child rearing practices" (2006:139). Individual developmental trajectories are sensitive to, and constrained by, both the onset of relevant cognitive capacities and variations in environmental stimuli (Prinz 2008a). Evidence from a vast array of cultures does show that children consistently make the moral/conventional distinction (Smetana 2002). But even this evidence fails to show invariance. Some cultures demonstrate greater competency at this task than others, which is thought to be due to variations in informational exposure through parenting practices (see section 2.4.2

below). Also, even though the ability to make the moral/conventional distinction is a cross-cultural phenomenon, this does not mean we need favour a nativist explanation. If the ability can be explained through recourse to non-moral learning resources and it is also the case that all normal children are in possession of those resources, then given similar learning environments this is what one would expect to see, even under non-nativist accounts. Not all cross-cultural phenomena are innate.

Given the above, we have reason to doubt the nativist claim that the trajectory of moral development supports the nativist argument. Still, these are not complete arguments against the POMS. Even if the developmental trajectory taken by the moral learner is slow, susceptible to environmental variations, and consistent with domain-general learning accounts, it does not follow that the acquisition process is. It may be that the resources available to the learner are completely inadequate to facilitate moral learning without positing an innate moral faculty. To address this issue, we need to look at the second claim of the POMS argument: task analysis.

2.4 The poverty of the moral stimulus: task analysis

Task analysis gives an account of the move from perceptual input to mature competence. The moral nativist claims that this could not occur via non-moral learning mechanisms alone because the move from data to theory is radically underdetermined. In what follows, we will see that the nativist has mischaracterised the learning strategies employed in moral learning and underestimated the resources available to the moral learner. Furthermore, the moral nativist has over-estimated the complexity of the move from perceptual input to mature competence. Various non-moral capacities of the learning child as well as resources in the external moral environment allow the child to distinguish moral from non-moral transgressions as required for moral learning. Although each of these resources may not be individually sufficient to attain moral competence, together they provide the requisite information for moral learning. In section 2.5 I will look at how this information allows the child to make the specific criterion judgements of the moral/conventional distinction.

It will be helpful to divide task analysis into the following three components in order to understand where the moral nativist goes wrong:

- Internal resources available to the learner
- External resources available to the learner
- Complexity of learning target and tacit moral knowledge

I will address each of these in turn.

2.4.1 Internal resources available to the moral learner

What, if anything, is contributed to the learning situation by the child?

Moral nativists claim that the non-moral learning resources available to the empiricist are inadequate to account for the complex task of moral acquisition. The POMS argument is severely weakened if an account can be given of how the child's moral learning resources can be enriched. There are numerous capacities involved in making the moral/conventional distinction that arrive long before children actually show the ability to make that distinction. Here I will look at these non-moral resources and show how they contribute to the acquisition process. (In section 2.5 we will see how these resources contribute to the specific properties of the moral/conventional distinction.)

We saw earlier that young children initially only make the moral/conventional distinction over a small class of norm violations, namely those pertaining to emotionally salient moral issues involving physical harm and welfare. This is important, because it suggests that the young child's emotional sensitivities are one of the primary resources they have for making this distinction (Nichols 2004, 2005; Dunn 2006; Smetana 2006; Prinz 2007, 2008a). We are born primed for assessing our emotional worlds – not only our own emotional reactions but also the emotional displays of others (Saarni and Harris 1989).

Although pro-social emotions alone are not sufficient for genuine moral judgement (Joyce 2006), they make the task of moral learning significantly easier. Emotional capacities contribute to moral learning by making particular aspects of the environment more salient, narrowing the learner's search space. Children use the affective consequences of acts to make inferences about whether they are moral or conventional norm violations (Arsenio 1988). At an early age, they have the resources to see that primitive rule violations are moral when the consequences of those violations elicit emotions such as empathy, and that they are conventional when they don't.

Experimental studies support this claim. Children consistently rate moral transgressions as more affectively salient. They also attribute negative emotions to the recipients and observers of the transgression due to harm, loss or injury to the victim. In the case of conventional transgressions, young children attribute "neutral or somewhat negative" emotions to the participants, more specifically attributed to those in positions of authority who did not want their rules to be violated (Dunn 2006; Turiel 1998; Smetana 2006). Additionally, the justifications that children give for their moral classifications pertain to the emotionally salient features of the acts such as the harm and welfare of victims, whereas their justifications for conventional classifications typically appeal to authority. Smetana concludes that "[c]hildren's direct experiences (as victims and observers of transgressions) provide one source of knowledge about the intrinsic consequences of acts for others' welfare and rights" (2006:136; see also Dunn 2003; Dunn 2006). We have good reason to think that *early* proficiency on the moral/conventional task is affect based and therefore is more indicative of the child's affective capacities than moral competence.

Soon after birth infants show an ability to adopt the emotional states of others around them (Saarni and Harris 1989). 'Emotional contagion' provides a direct, pre-theoretical resource which forms the developmental foundation for interpreting the emotional states of others. By the end of the first year the learning child has developed the capacity for 'social referencing', using the emotional expressions of others to evaluate novel objects or situations. For example, children are more likely to cross a glass-covered cliff or play with novel toys if their mothers display positive emotional states, whereas the opposite is true if they display negative states. In this way children use affective information in their environments to make decisions and guide their actions (reviewed in Thompson, Meyer

et al. 2006). Both of these skills are essential for the future ability to assess the emotional states of others when identifying moral violations and using that information to make decisions with regards to those violations.

Between the ages of 12 to 18 months toddlers start to respond to others' negative emotional states in the form of 'distress reactions'. They will show concern and display prosocial behaviour towards another's distress by attempting to comfort them through both physical action and talk (Dunn 2003). Here we see that from a very early age children are motivated to prevent or reduce the harm caused to others. This is important for marking the qualitative difference between emotion-inducing moral violations and non-moral violations, as well as providing a motivational dimension to moral learning. Children are motivated to prevent harm, not because they are necessarily told to by an authority figure, but because they are independently motivated to do so. Distress reactions also indicate the perceived seriousness in the eliciting event. These facts are significant when we consider that moral violations are deemed independent of authority and more serious than conventional norms (see discussion below in section 2.5). Infants also show the ability to use facial and vocal displays as a means of identifying different emotional states in others, such as fear and sadness (Harris and Saarni 1989). Abilities like these are vital tools for identifying the various emotional states of victim, transgressor and admonisher in moral learning. Learning the emotional consequences of actions enables us to identify distinctions between the various causal components of both moral and conventional violations.

Evidence from studies of atypical populations also shows that capabilities are essential for the ability to make the moral/conventional distinction. Psychopaths, who show diminished emotional responses towards the suffering of others, also show diminished abilities in making the distinction from an early age (Blair 1995; Nichols 2004). Yet emotions are not the only mechanisms involved in the moral acquisition process. Autistic children who show deficits in theory of mind and intentional attribution can successfully make the distinction but have trouble with respect to violations that require mind reading capacities to understand the event, such as instances involving lying (James and Blair 1996; Nichols 2002a). This is consistent with autistic children showing less of an ability to lie than normal children (Sodian and Frith 1992). Moral acquisition involves many cognitive abilities: the ability to identify causal chains and consequences; reason

counterfactually; follow norms; make abstractions and generalisations. This is especially true in the case of complex moral competencies. Many of these are in place in early infancy. 12-month-old children, for example, start to show an awareness of others as “deliberate and subjective partners” as seen through such displays as expectation, imitation and joint attention (Thompson, Meyer et al. 2006). The ability to perceive intentional behaviours of others is an important tool in generating the relevant causal structure of more complex events involving the attribution of purposeful vs. accidental behaviours and deceptive behaviours.

By integrating additional informational resources into their evaluations, children make more nuanced and complex moral judgements. Arsenio and Love (1995) argue that young children’s sensitivity to the emotional states of others provide them with the information needed to anticipate the morally relevant consequences of their behaviours. These ‘emotion-event expectancies’ provide information from which children, over time, can construct abstract and generalisable moral principles. However, Arsenio and Love also point out that it is not solely the emotional capacities that allow the construction of these principles; it is through the integration of various experiential and developing cognitive resources that the child comes to make more subtle evaluations of their moral worlds. Young children initially view all perpetrators of moral violations as happy because the child thinks the perpetrator has gained from the transgressing act. But, as the child’s social and affective experience increases and is combined with the development of various other cognitive skills, the child comes to appreciate that the perpetrator may not necessarily be a “happy victimiser”. Children come to understand that perpetrators may also be sensitive to the negative effects of their actions. The important point is that moral cognitive abilities develop in conjunction with the development of other capacities. As these come online, the resources available to the moral learner grow, allowing the acquisition of greater moral competence.

The upshot so far is that the data-to-theory gap has been narrowed by the child without employing innate, domain-specific moral mechanisms. The internal resources contributed by the learning child help to identify basic norms specific to the moral domain, resources which form the basis of moral acquisition. Whether any of these specific capacities themselves are innate or not is an open question but one that is not under examination here. What is important is that these capacities are not specifically moral. Each is readily

employed in domains other than moral cognition. This is an important point that Jesse Prinz (2008a) emphasises in his non-nativist account of moral acquisition. He explains that the awareness and assessment of the emotional states of others, for example, arise in domains far beyond that of morality; emotions such as empathy are invoked in instances such as car accidents and natural disasters, as well as other contexts such as interpreting the mental states of others (Shaun Nichols (2005) also makes this point). Looking at other resources available to the child we see that situational assessment via social referencing is, again, not specific to morality. Nor are consequential, counterfactual, general reasoning, theory of mind and norm following abilities. These internal resources are not specific to morality; they cannot be considered a unique part of any domain-specific moral faculty.

So far we can conclude that the child's mind is a biased learning machine – non-moral mental resources contribute substantially to reducing the learning gap. Importantly, children also learn by doing. As we will see later, it is their active, embodied participation in a highly structured moral world that facilitates the transmission of external moral information to the learning agent (see chapters 4 to 6). Children interact with and explore their worlds, and importantly their moral worlds interact with them. We will now look at some of the ways in which the structured moral world contributes substantially to moral acquisition.

2.4.2 External resources available to the moral learner

What, if anything, is contributed to the learning situation by the environment?

We have seen that the child does not learn through dispassionate observation alone. The many emotional biases and cognitive capacities that are used in navigating their social worlds contribute substantially to the evidence required to differentiate moral from conventional transgressions. But the child's social relations and the subsequent structuring of their moral environments are an essential source of *directed* information for the development of the moral learner – enriching the evidence available to the learning child.

In her account of the POMS, Susan Dwyer (2006) makes the claim that there is very little

evidence concerning rule violations that could cue the child as to whether a moral or conventional rule had been broken. Yet, intuitively, it does not seem to be the case. Given that normal adults identify the two domains along lines of authority dependence, seriousness and generalisability, and differentiate between the types of justifications they give, one would expect them to treat them as such (Prinz 2008b). Empirical studies of parenting techniques confirm this intuition - parental disciplinary responses to children's moral and conventional transgressions are qualitatively different (Smetana 1989, 1997, 1999; Dunn 2006; Thompson, Meyer et al. 2006).

Judith Smetana (1999) identifies two components of parental transgression responses that are relevant to the acquisition process. The first is the *cognitive component*. The ways in which parents communicate with their children are domain specific and impact upon moral development. Studies of maternal parenting practices show that verbal responses to moral transgressions typically focused the transgressor towards the consequences of the act in terms of injury, loss, emotional reaction, rights and welfare. With the exception of rights, these are consequences that children are emotionally sensitive to at a very early age. In contrast, maternal responses to conventional transgressions typically focused on social order and regulation. (Smetana 1989, 2006). Here we have clear evidence that parenting techniques provide explicit informational differences between the two domains.

The second component of transgression responses is an *affective component*. The emotional displays of caregivers help children to understand and encode moral information by focusing the child on the act and consequences of the transgression (Arsenio and Lover 1995). As with the cognitive component, studies in maternal parenting practices show that their affective responses are qualitatively different in the two domains. In the moral domain parents typically exhibit more negative emotions and increased anger than they do in non-moral domains, providing a qualitatively different experience for the learning child between the two transgressions (Smetana 1999; Thompson, Meyer et al. 2006). In this way, the affective and cognitive elements are complementary sources of different information to aid moral acquisition.

Importantly, research also shows that parental disciplining methods that focus on domain-specific explanations and consequential reasoning improve moral and social

development (Zahn-Waxler and Chapman 1982; Smetana 2006). Children of caregivers who explicitly undertake other-oriented reasoning during admonishment, are reported to commit fewer moral transgressions. Children of parents who frequently discussed feelings during evaluations of moral behaviours were more likely to exhibit guilt after committing a transgression compared to those of parents who did not. Furthermore, the degree to which parents undertake reasoned argument, provide justifications and discuss emotions during conflict resolution with toddlers is a good predictor of future moral development (Laible and Thompson 2002; Thompson, Meyer et al. 2006). Parental reasoning encourages children to reflect on the consequences their actions have upon others, a skill that facilitates making the moral/conventional distinction.

In contrast, power assertive parenting techniques, such as strict and aggressive rule enforcement, with little or no consequential or emotional explanations, are effective at inducing short-term moral compliance, but do not facilitate long-term moral development. Power assertive techniques fail to provide adequate information about the nature of the transgressions, and do not encourage children to reflect upon their action because the emotional focus (anger from the parent and fear in the child) is on the transgressing child and their feelings, rather than the consequences to the victim of the transgression (Zahn-Waxler and Chapman 1982; Smetana 2006). Furthermore, the act is prohibited because the caregiver says so, not because the act has negative consequences on the victims of the transgression. Here we see environmental effects on developmental trajectory, confirming the importance of domain-specific content in moral development and showing that moral development is not invariant to exposure. Children's proficiency at differentiating moral and conventional violations is, in part, dependent on the information the developing child receives through parenting practices (Smetana 2006:139).

Cross-cultural studies show that people across a broad range of ages differentiate moral from conventional norms. Although very little of this cross-cultural research has focused on pre-schoolers' competencies (Yau and Smetana 2003), one study by Yau and Smetana looked at category distinctions made by pre-school Chinese children in Hong Kong. They found that in comparison to US children, a much larger percentage (almost all) of the Chinese preschoolers they tested gave justifications for moral transgressions in terms of welfare and fairness. The authors write that it appears "young Chinese children have a

more sophisticated understanding of the intrinsic wrongness of moral transgressions than do comparatively aged US children...” (2003). They suggest this difference could be due to Chinese disciplinary practices that focus on eliciting guilt and shame through the narrative retelling of transgressions committed by the child.^{6 7} This is evidence that informational exposure differences in the form of parental disciplinary practices affect the rate of moral acquisition.

We also see quantitative differences with respect to the context in which domain-specific transgressions occur. The majority of moral transgressions and moral conflicts happen in sibling and peer relationships (Smetana 1989, 2006; Dunn and Slomkowski 1992). These interactions provide first-hand experience of the effects (particularly emotional) of moral transgressions and a means by which to actively explore the moral domain. When a child unfairly takes a toy from another, they experience the negative emotional consequences that come with that type of transgression. The primary source of conflict within the conventional domain, on the other hand, happens with authority figures such as parents, and these conflicts are not typically victim orientated. Because learning is frequency dependent, the contexts in which the different types of transgressions occur is an important means by which to differentiate between the two domains.

Finally, it is not solely through first hand experience as perpetrator, victim or observer that children obtain all domain-specific information. Second-hand information plays a role in providing emotionally salient exemplars of moral and non-moral behaviours and their consequences. From a very young age children experience fables, nursery rhymes, and children’s stories that positively emphasise moral behaviours while negatively emphasising immoral behaviours. The narrative structure of story telling allows the child to see the causal structure of the socio-moral events and their consequences.

From what we have seen so far, the child’s external environment is full of resources which direct the child to the differences between moral and non-moral transgressions, thereby structuring their learning environments and reducing the gap between learning

6. It should be noted that the authors only suggest this is only a possible reason that requires further research (Yau and Smetana 2003).

7. Although the inducement of shame is found to negatively correlate with moral development, guilt is found to correlate with moral development (Yau and Smetana 2003). Note also, this is not to claim that young children necessarily experience guilt or shame, but that the parenting practices focus on guilt and shame.

target and achieved competence. It might be objected that the paradigm cases of moral transgressions, such as those involving harm and welfare, are obviously emotional in nature. What about differentiations between more abstract and less obviously emotion generating transgressions? For example, Susan Dwyer gives two transgression examples that are not so obviously emotionally salient, claiming that "...it's worth noting that 'You ought to keep your promises' has precisely the same form as 'You ought to put the fork on the left.'" (2006:240). Dwyer, unfortunately, does not provide evidence that children can and do make this specific distinction at a sufficiently young age to warrant support of the POMS argument. As we have seen above, ontogenic time of expression is relevant if that competence is to be held as evidence for nativism.

These two prohibitions have the same syntactic structure, but they do not have the same consequential structure. Children have numerous examples and first-hand experiences of promise-keeping as well as failures to keep promises and the consequences of both. When a father promises to take his child to the zoo and subsequently fails to do so, the child has good evidence of both what a broken promise is, and what it feels like to be the recipient of a broken promise. Children also learn that promise-keeping is embedded in a web of social interactions, preferences and desires. Learning that "you ought to keep promises" is not just about learning a particular norm; it also involves learning the various relations between agents, their desires and preferences, and the consequences of promise-breaking on those relationships. One keeps promises not just because there are norms about promise-keeping, but because one has, among other things, a desire to not hurt another, to maintain another's trust, or perhaps to please or impress others. These factors are all relevant to the identification of promise-keeping as a moral norm. Violations of a social regularity such as fork placement norms on the other hand are unlikely to be treated in the same way. It may have the consequence of annoying or upsetting the enforcer of that regularity and children identify it as such, but in most cases it is unlikely that parents' emotional and reasoning responses will be like that of a moral transgression.⁸ Again, we see the consequences in the two domains as different, and it is these consequences that mark the qualitative difference between moral and conventional violations (see also Prinz 2008b).

8. Some children do view some conventional transgressions as having "second-order moral consequences", justified by not wanting to hurt another's feelings. Should the authority be sufficiently upset at incorrect fork placement, then child may view the transgression as harming the authority in this way (Yau, Smetana et al. 2009).

It is apparent that early competence at making the moral/conventional distinction involves drawing inferences from features of the acts as opposed to specific knowledge of the norms pertaining to those acts (Turiel 1998; Smetana 2006). From this, we can see how the features of the two prohibitions are quite different. Internal and external resources draw attention to those features and structure and support the learning experiences in the two domains. It is not merely through observation of regularities in the moral environment that children become disposed to moral education. Children commit, and are the victims of, transgressions, and in doing so they actively participate in the learning experience and their environment actively participates with them. It is the *total* experience that provides the requisite information for moral acquisition. The channels of moral information are more directed and salient than many authors promoting the POMS argument are willing to concede.

2.4.3 The complexity of the learning target and tacit moral knowledge

It is apparent that internal and external resources contribute to moral learning, making the learning target more salient, easier to identify and therefore easier to learn. However, while it may be the case that moral norms and their violations are clearly signposted as distinct from the non-moral, whether or not morality is learnable is another issue. John Mikhail writes:

to explain the observable data we must attribute unconscious knowledge and complex mental operations to the child that go well beyond anything she has been taught.... These concepts and principles which underlie them are as far removed from experience as the hierarchical tree structures and recursive rules of linguistic grammars. It is implausible to think they are acquired by means of explicit verbal instruction or examples in the child's environment. (2008b:355)

There are two related aspects with respect to the learning target and the problem of acquisition as highlighted by Mikhail. The first is what I will term the *Problem of Complexity*. The complexity of the learning target and the task of information extraction is

such that we are unable to extract the relevant moral information from the environment. We can, for example, be taught to identify a mathematician working on complex mathematical problems. However, the task of extracting the relevant information required for acquiring the mathematic skill of, for example, differential calculus is difficult, if not impossible, without detailed and explicit instruction as well as relevant background information such as various principles of mathematics. Similarly, due to the complex nature of the task, it may be impossible to extract the relevant principles of morality without the appropriate instruction. Because we do not have this detailed and explicit instruction, we cannot possibly learn the underlying structural principles of morality. Mikhail writes: “for it is the complexity, or lack thereof, of the output of the acquisition model which determines whether or not the stimulus is impoverished with respect to that output” (2008b:353).

The second reason for believing that moral competence cannot be learned is the claim that *moral knowledge is tacit knowledge*. Marc Hauser, for example, claims that “the key idea driving my argument for the moral faculty, is that much of our knowledge of morality is ... intuitive, based on unconscious and inaccessible principles for guiding moral judgements of permissibility” (2006b:125). There are two reasons why one might think this is the case. Firstly, people often fail to justify many of their moral behaviours and cannot articulate the principles that underwrite those judgements, so it is unlikely that we come to those principles through reason alone. Secondly, these principles are not explicitly presented to us in our environments, so it cannot be the case that they have been learned (“Parents can’t teach what they don’t know” (2006b:423)). If we can neither reason nor learn certain moral knowledge, then that knowledge must be innate.

Although the issues of complexity and tacit principles come apart, their connection is implicit in much of the nativist’s argument. The complexity of the task gives more credence to the nativist’s claims of ‘intuitive, unconscious principles’. If the tacit principle to be learned is simple, then implicit learning without explicit instruction is more plausible. We can, for example, plausibly implicitly learn that ‘hitting is wrong’ from salient exemplars in our environment. On the other hand, it is not as obvious how we can implicitly learn such principles as the doctrine of double effect, which holds that ‘it is sometimes permissible to cause harm if that harm is an unintended side-effect of bringing about a good, but it is impermissible to cause harm if it was intended as a means

to bring about the that good'. The complex nature of some tacit principles makes implicit learning less likely, and hence provides a stronger case for the moral nativist.

I will address each of these aspects in turn. We will see that the task of moral acquisition is typically not that complex, and is often not specifically moral at all. It will also become apparent that we have reason to think complex tacit moral principles can be learned without explicit instruction. Furthermore, some examples of tacit moral principles used by moral nativists can be explained without recourse to innate, tacit moral principles at all.

The problem of complexity

The child's learning environment is awash with moral and non-moral norms. The nativist claims that the learning task of extracting the correct principles and corresponding judgements appears to be so difficult that it cannot be explained by learning. We have already seen that early proficiency at the moral/conventional distinction does not require moral rule recovery: the focus is on the features of the acts, rather than the norms prohibiting those acts. So the moral nativist needs to be careful not to overstate the complexity of the learning target in the early years of moral development. The young child does not require an understanding of the principles that underwrite harm prohibitions to accurately identify instances of harm and be motivated not to cause harm. Natural aversions of the young child can lead the child to act in accordance with a moral rule without accepting, internalising or understanding that rule. The learning child is not necessarily performing rule extraction from the environment because competence at the task does not require mastery of a general rule.

As we saw in the previous section, the complexity of rule extraction is reduced substantially by the internal and external resources available to the moral learner. This then leaves the task of rule extraction itself. Yet, importantly, it is not specifically a moral problem as it pertains to all forms of rule extraction from one's environment, be that etiquette, sports, games, or social norms, to name but a few. Children can learn many of these early in life and in many cases without explicit instruction. We know therefore that the ability for rule extraction itself arrives early. Rule extraction in general may often be very complex, and the capacity may be underwritten by an innate rule learning faculty

(Sripada and Stich 2005), but because it is a process we find across many different domains it cannot be considered a constituent part of an innate moral faculty. Rule extraction is a problem to be addressed independent of a nativist moral faculty and hence is not captured by the POMS argument. This objection is specifically problematic for defenders of a rule or principle-based moral faculty.

What about the later stages of moral acquisition? Perhaps here we see complex mental operations beyond anything that could have been taught. In defence of the POMS argument, John Mikhail uses psychological research to argue such a case. According to Mikhail, in complex trolley bus problems, eight-year-old children have shown sensitivity to the doctrine of double effect and intentional battery, beyond anything that they could have learned from their environment. He writes: “In the case of trolley bus problems, for example, children must represent and evaluate these novel fact patterns in terms of properties like ends, means, side effects, and prima facie wrongs such as battery, even where the stimulus contains no evidence of these properties. These concepts and the principles which underlie them are as far removed from experience as the hierarchical tree structures and recursive rules of linguistic grammars. *It is implausible to think that they are acquired by means of explicit verbal instruction or examples in the child’s environment*” (2008b:355, italics added).

A proper analysis of Mikhail’s example requires the separation of those properties that are specific to the moral domain and those that are not. The complexity of the moral acquisition task is not as great as Mikhail maintains, as the ability to identify and evaluate ends, means, and side effects are not specific to the moral domain. Much of the relevant information therefore is recoverable from the environment with the aid of these capacities, independent of positing an innate moral faculty. From Mikhail’s list, battery is the only property that is specific to the moral domain. So, assuming that the notion of battery really is a requirement for making distinctions in the trolley bus examples, the question becomes whether or not a child can learn, *by eight years of age*, the notion of ‘intentionally harming without consent’. There seems no reason why concepts such as battery cannot be learned from a child’s environment over such a timeframe.

So far I have argued that early moral competence need not be a case of rule extraction and that, in cases where it is, the problem of rule extraction is not specific to the moral

domain. Furthermore, many cases of moral acquisition that appear to be complex and beyond the information available are not as complex as it would first appear. I have not yet mentioned whether agents are aware of, or can cognitively access, aspects of their moral knowledge. This brings us to an additional argument put forward by the moral nativists: we have tacit knowledge of certain moral principles that operate over representational structures such as means, ends and side effects. Because we cannot readily identify or articulate those principles and they are not explicitly presented to us in our environments, it is unlikely, the nativist concludes, that we could have learned these principles. Therefore, they are likely to be innate.

Tacit moral knowledge

Many moral nativists have been drawn to their conclusions from the observation that agents make what are perceived to be complex intuitive moral judgements without an explicit understanding of why they make those judgements. For example, the observation that agents typically make judgements in accordance with the *doctrine of double effect* but cannot explicate the principle, is cited by many as one such source of evidence (see Dwyer 1999, 2006; Cushman, Young et al. 2006; Hauser 2006b; Mikhail 2007, 2008b; Harman 2008).⁹ Gilbert Harman writes that “[s]upposing there is widespread implicit acceptance of [the principle of double effect], it would seem that ordinary people have no explicit instruction of it, and it would appear not to be transmitted by explicit instruction. If there is no other obvious way for the principle to be learned, the hypothesis suggests itself that the principle is somehow innate...” (2008:346). And Marc Hauser claims that an innate moral faculty explains the “principle of double effect, a principle that is operative but not expressed when people are asked to justify their responses” (2006b).

It is clear that this argument from unconscious tacit moral principles does not apply to many of the basic moral principles that guide our lives. Throughout development the learning child witnesses many clear and cognitively accessible examples of principles like fairness and equality through their participations in childhood activities. Caregivers also provide explicit instruction and justifications by appealing to moral concepts and

⁹ Dwyer (2009) has recently claimed that the principles of moral grammar cannot even be expressed in moral principles such as doctrine of double effect. Moral grammar is deeper and more abstract. On this point, Dwyer differs from some previous work by moral grammarians like Hauser and Mikhail on the doctrine of double effect.

principles, such as “why having more ice-cream than your brother is unfair”. Unlike the ‘subjacency principle’ in linguistics, many of the basic principles that underwrite moral judgement such as harm, welfare and fairness are encoded in ordinary language and salient to both transgressor and victim. People can also typically give justifications for most of their moral judgements that appeal explicitly to these moral concepts and associated principles.

With respect to complex moral knowledge, such as the case of the doctrine of double effect, the question is whether or not complex tacit knowledge can be learned without explicit instruction. If it can, then we have some reason to doubt the nativist’s argument. There are many examples of learned knowledge that are tacit. Most people who can ride a bike cannot explain in any great detail the coordinated actions and various skills involved in doing so. They cannot, for example, articulate how one calculates the angle relative to speed with which to lean into corners. We may be taught how to ride a bike, but these required competencies are never explicitly explained. Bike-riding knowledge is one example of tacit, learned knowledge.

In their book *Mind Over Machines*, Hubert Dreyfus and Stuart Dreyfus (1986) detail how we can gain advanced intuitive, unconscious, pattern-recognition based skills through learning. They cite the example of Grandmaster chess players, who can play multiple high-level, high-speed chess games simultaneously with very little time to reflect on the moves they make. Their moves are intuitive and show minimal reliance on conscious reasoning. Beginner chess players, on the other hand, make decisions based on conscious reasoning and explicit applications of the rules of chess which they have been taught. With experience, the learning chess player starts to show less reliance on conscious explicit rules and more on maxims and strategic play. As they progress towards the level of expert, the chess player operates heavily by intuition, without relying on conscious considerations of the rules of chess to make their judgements. Their strategic goals are more general and they can identify very large numbers of board positions and their profitability without reflection. By the time he reaches expert, the chess player is a master at intuitive, unconscious pattern recognition. These are patterns that he may not be explicitly aware of, and importantly, patterns that have been learned through repeated play and simulation of various game positions. Chess players learn by doing and not always through explicit instruction. Expert chess is a clear example of how complex tacit

knowledge can be learned. It is plausible that the acquisition of tacit moral knowledge can proceed in an analogous way (Sterelny forthcoming). We learn and implement explicitly taught simple moral concepts and principles, graduating through development to more general, exemplar-based pattern recognition skills. High-level moral competence under such a picture is intuitive pattern recognition rather than the application of specific principles like the doctrine of double effect.

We also have reason to doubt that moral principles are inaccessible as moral nativists like Hauser claim. Kim Sterelny (forthcoming) points out that the principles of morality can be made explicit in natural language and understood by the folk moraliser (see also Prinz 2008b). Once we know of those moral principles, we can use that explicit knowledge to guide future judgements in novel and more complex moral dilemmas. This is not something that we do in the case of language; even once linguistic experts learn the syntactic structure of language they do not use that conscious knowledge to guide their parsing. Furthermore, Sterelny also makes the point that our reflections about moral dilemmas alter our intuitive moral behaviours; through debate and reflective moralising people can reject previously held principles and adopt new ones.¹⁰ Adopting a consequentialist ethic that rejects the doctrine of double effect is, under the moral grammarian's picture, a rejection of a structural principle of universal moral grammar. Yet we can still morally converse with, debate and understand the different moral positions of the consequentialist. We could not do the same with anyone who rejected structural principles of language. (In chapter 5 we will see that the purpose of moral debate is to allow the flexible and successful group navigation of the social world. This would be severely limited if morality were as rigid as some nativists maintain.) Some of our moral principles may be tacit, but they are neither inaccessible, unable to be made conscious, nor unchanging in the way the syntactic principles of language are.

If it is true that complex tacit principles can be learned without the aid of explicit instruction, then it is possible that moral principles like the doctrine of double effect can also be learned without explicit instruction. Yet, it may be that we need not invoke complex tacit moral principles at all to explain apparently complex moral behaviours.

10. Note how this fits with the evolutionary picture drawn in chapter 1 involving punishment and group conformity. One of the adaptive benefits of moral cognition was the influence and control of other moral views, to argue, to debate and punish to make others change their moral outlook. (See also, chapters 5 and 6).

Jesse Prinz (2008c) argues that the contrasting results of trolley bus examples can be explained as the product of conflicting moral norms rather than singular, tacit, complex innate moral principle such as the doctrine of double effect, as proposed by Hauser, Mikhail et al. Prinz points out that in the trolley bus cases, we have two conflicting norms: a norm for helping, and a norm that prohibits killing. In the case where we are asked to pull a lever and divert a trolley bus down an alternative route, saving five innocent people but in the process killing one, we have norms of helping (saving the five) and norms against killing (killing one). But in this case the killing of one person by diverting the trolley is not a paradigm instance of prohibited killing, such as, for example, shooting someone. Nor, Prinz claims, is it an emotionally intense form of killing. Therefore, norms of helping trump the less salient and less obviously applicable norm against killing. In contrast, killing a man by pushing him off a bridge into the path of an on-coming trolley bus, thereby preventing the death of five others, is generally deemed impermissible. Prinz argues that this is because pushing a man to his death in front of a trolley bus is more emotionally salient and is also a paradigm case of forbidden killing. In such a case, norms prohibiting killing outweigh norms for helping. Other factors such as proximity to action and the violent nature of the death will also manipulate our emotional responses to the various acts and play a significant role in our judgments of moral permissibility in such cases. What is key to Prinz's argument against moral nativism is that these norms of helping and norms prohibiting killing can be learned (2007, 2008c). So, although moral principles are involved, one does not need to posit a singular, tacit, innate principle of double effect to explain the observed phenomena. (See also Shaun Nichols (2005) who suggests separate utilitarian and deontological psychological systems, rather than universal moral grammar, explain judgments in accordance with doctrine of double effect.)

Other evidence suggests that observations such as the doctrine of double effect may in fact be an artefact of domain-general causal reasoning. Waldmann and Dieterich (2007) examined whether causal reasoning had any bearing on the outcome of the observed phenomena. What they found is that the point of causal intervention has a bearing on people's moral permissibility judgements. The domain-general processes of causal psychology result in people preferentially focusing on one particular causal path to the exclusion of others, thereby biasing appraisals of the various moral dilemmas. In this case, task performance can be explained, not by appealing to a singular, innate, tacit

moral principle, but by understanding the observed phenomena as an artefact of our more domain-general processes of causal reasoning. (See Box 2.0 for more detailed explanation.)

Both of these explanations are plausible and have some evidential support. I hazard a guess that the results in the trolley bus problems are due to a combination of both accounts above, as well as various other factors (see for example Greene, Sommerville et al. 2001). The importance of these models is that they show plausible non-nativist explanations of moral phenomena, weakening the argument that we need to posit complex, innate tacit moral principles to explain phenomena such as the doctrine of double effect.

Box 2.1 Causal psychology, moral problems and the doctrine of double effect

It is claimed that the tacit moral intuitions we have in various moral dilemmas provide evidence for an innate underlying structure of moral cognition (see for example Mikhail 2002; Hauser 2006b; Dwyer 2009). One example of evidence used in support of these claims is people's responses to the famous trolley bus moral thought experiment. Subjects are asked to imagine a scenario in which a trolley bus is travelling towards a group of five people on a track, all of whom will be killed if the trolley continues on its path. The death of these five people can be averted if the trolley is diverted on to a side-track but there is a one person on the side-track who will be killed. When presented this moral dilemma, people will typically judge that it is morally acceptable to divert the trolley and kill one person to save the lives of five. In contrast to these judgements are those in the spare-parts surgeon moral thought experiment. In this scenario, a surgeon has five patients who will all die if they do not receive organ transplants. Unfortunately, there are no donated organs available. The doctor has an opportunity to take the life of a healthy individual, without anyone knowing, and use his organs to save the lives of the five people who require organ transplants. In this thought experiment, it is typically deemed impermissible to kill one patient to harvest their organs to save the lives of five other patients. In these cases, the outcomes are identical, but the causal paths are different. What then, could generate the different permissibility judgements in scenarios like these?

One hypothesis is that all normal people have an innate set of moral principles which generate the various judgements that we make. John Mikhail, for example, claims

that “the moral intuitions generated... can be best explained by postulating intuitive knowledge of specific moral principles, including the prohibition of intentional battery and the principle of double effect” (2002:11). The *doctrine of double effect* (DDE) states that it is sometimes permissible to cause merely foreseen harm if that harm is a side-effect of bringing about a good, but it is impermissible to cause that harm if it was intended as a means to bring about that good. Although almost all people, including children, will often act in ways which are consistent with the DDE, very few are aware of the principle, let alone able to articulate it, and children are neither explicitly nor implicitly taught it. Therefore, it is claimed, the DDE is likely to be innate. An innate, distinctly moral reason explains why these cases are treated differently.

Research by Waldmann and Dieterich gives us good evidential reasons to question this claim. Drawing upon models from causal psychology, they looked at whether “domain-general features such as the structure of causal models in reasoning and learning ... affect moral judgment” (2007:248). These permissibility judgements are not due to any specifically moral feature of human cognition but are an artefact of more domain-general causal reasoning which focus attention on particular causal paths.

The authors make a distinction between agent-interventions and patient-interventions: agents are those that cause harm, whereas patients are those who are the recipients of harm. What they found is that the type of causal intervention is a primary factor which influences many of the moral permissibility judgements that we make by focusing attention on specific causal paths within the dilemma. Throwing a bomb out of a restaurant to save nine lives but killing one person outside standing on a patio is an example of agent-intervention. In this case, people’s focus of attention is on the agent of harm (i.e., the bomb) and therefore the evaluations that we make focus on the causal effects between throwing the bomb or not throwing the bomb. In contrast, throwing a person onto a bomb to save nine lives is an example of patient-intervention, whereby the patient subject to harm is the person thrown on the bomb. The evaluative focus is on the consequences for the patient to the exclusion of background effects, such as the consequences for other people in the room. The authors dub this attentional focus on the act of intervention as *intervention myopia*.

The authors then tested whether the observations above were actually due to people’s reluctance to use people as a means, as in the case of the DDE. They found that this was not the case and concluded that “the point of intervention, rather than whether or not people were used as means, seems to have been responsible for the effect”

(2007:251). Their explanations do not rely on innate moral principles such as DDE, but instead on the more general features of domain-general causal reasoning. These types of models show that there are plausible alternative explanations of the evidence that the moral nativists appeal to in support of their position. These explanations do not rely on an innate moral faculty.

Before moving on it is worth recapping what has been shown above. In support of the POMS, moral nativists have claimed that the task of moral acquisition is an especially difficult one which cannot be explained by moral learning, and morality must therefore be innate. I have argued that moral learning is in many cases either not particularly complex (in the case of early moral learning), or the complexity of the task is not specific to morality. Furthermore, moral nativists have claimed that much of our moral knowledge is tacit. People cannot express the underlying principles involved in their moral judgements, nor are people explicitly taught these principles. Because it is not obvious how they can be learned, or acquired by reason, they must be innate. However, I have demonstrated that there are many analogous cases where tacit knowledge is learned without explicit instruction; we therefore have some reason to believe that the same could occur in the moral case. I have also presented examples of how one can explain apparently tacit, innate moral knowledge without appealing to innate moral principles. We have good reasons to doubt the moral nativists' claims about both the complexity of the learning task and tacit moral principles in support of moral nativism. The gap between input and output in the POMS argument is not as the moral nativists claim.

2.5 Authority independence and the moral learner

So far we have seen a number of arguments against the moral nativist POMS argument. Much of the focus has been on the resources that enable the learning child to differentiate moral from conventional transgressions. Very little has been said about the specific dimensions with which children and adults make these distinctions. I will now look at how children make what are called *criterion judgements*, focusing in particular on the dependence relation between moral and conventional transgressions and authority. Some moral nativists have claimed that drawing the authority-dependence relation is especially complex and that competence in drawing this dimension is evidence of innate moral

structures (see for example, Joyce (2006)).

Adults and children identify both moral and conventional transgressions as wrong, and, at a finer grain of analysis, distinguish them as wrong along several dimensions. Moral norms, as compared to conventional norms, are typically judged to be more serious, more generalisable, less dependent on authority, and differ in the justifications that are given for compliance (Turiel 1983; Turiel 1998). The fact that judgements are made along these criteria has been cited by some as evidence in support of the POMS argument as there is insufficient evidence and the task is too complex for the learning child to make such distinctions. Hence, the information required to master the distinction must be innate. What we will see is the resources that allow children to identify the moral as distinct from the conventional, also impart the various properties of the criterion judgements (See Nichols 2004, 2005; Prinz 2008b). I will briefly address the dimensions of seriousness, generalisability, and justifiability before focusing on authority-dependence.

2.5.1 Seriousness, generalisability and justification

When we make moral appraisals we typically classify them as being more *serious* than conventional appraisals. We can account for this classification in terms of the negative emotional responses that canonical moral transgressions (such as those pertaining to harm and welfare) elicit in the child. We also saw above that parental response to moral transgressions exhibit more negative emotions. Emotional responses from both parent and child are more serious in the moral domain (Smetana 1999; Nichols 2002a; Prinz 2008b). Moral transgressions that are judged to be wrong are classified as *generalisable* if they are judged to be wrong across different social contexts, be that in another home, school, cultural group or country. One relevant factor is that affective consequences remain, irrespective of the context in which the transgression occurs. To the young child, hitting causes suffering in all instances, irrespective of its context. Conventional transgressions on the other hand do not involve an obvious victim or suffering that can be generalised. Children also view authority as context dependent and not generalisable (see below), and give domain-specific *justifications* as to why transgressions are prohibited. Moral justifications involve appeals to the prevention of harm and the promotion of people's welfare, fairness and rights (Turiel 1998). They typically do not involve appeals to authority. Again, we can account for this because the emotional consequences of

moral violations are different to those of conventional violations (Nichols 2004; Prinz 2008b). The justifications for appraising a conventional transgression are typically in terms of consequences to social acceptability and organisation, directives from authority and customs, because those are what conventional norm violations transgress (Turiel 1998).

Still, although children can justify the transgressions as being dependent or independent of authority, this does not address the issue of how children can learn the relevant dependence relations. In the case of moral transgressions, for example, they may be emotionally salient, but they are also typically prohibited by authority. How then can one learn that moral norms are authority-independent without witnessing this independence? (Joyce 2006).

2.5.2 Domain specific relations with authority

One of the more peculiar features of the moral/conventional distinction is each domain's particular relation with authority. Moral prohibitions are typically seen as being independent of authority.¹¹ Research shows that children will reject an adult's authority if that adult makes immoral requests (Smetana 2006). In contrast, whether or not a conventional transgression is viewed as wrong is typically seen as being more dependent on authority figures. For example, although it is typically judged wrong for a boy to wear a dress to school, it ceases to be wrong if a teacher sanctions the act.

Richard Joyce has claimed that the capacity to draw the dependence relation within each domain is evidence in support of innate moral knowledge. He argues that in order to understand that conventional norms are *dependent* on authority "one would have to observe a correlation between the relevant authority changing its mind to permit the boy to wear a dress and that action no longer counting as a transgression." And in the case of moral norms, "in order to infer an *independence* relation one would have either to (A) observe the relevant authority change its opinion about an act of harming while one

11. This is a different claim than saying that moral judgements have moral authority – the claim that moral judgements have authority over an agent irrespective of their interests. Authority independence means only that an action's moral status will hold irrespective of whether or not an authority figure sanctions it, which is still consistent with the claim that moral judgements are subject to one's own preferences.

notes that the act nevertheless continues to count as a transgression or (B) observe a previously condemned act of harming cease to count as a transgression while one notes that the relevant authority's opinion on the matter has not altered" (2006:138-139). The observational evidence required to draw the domain relevant dependence relations are complex and difficult, if not impossible, to come by (especially for a three year old child). The information required is therefore likely to be innate.

This picture misdescribes the learning situation, however. For the young moral learner, drawing the dependence relation is not a case of rule extraction from the environment followed by determining the relevant dependence relation with authority. Moral norms are not norms that also happen to be authority-independent; they are moral norms *because* they are instantiated by authority-independent events such as harming. They are prohibitions because of the consequences their violations have, not because an authority figure tells the child they are prohibited. Similarly, conventional transgressions are not norms that happen to also be authority-dependent; they are conventional norms *because* they are instantiated by features of the world that are determined by authority, such as social regularities. They are prohibited because authority prohibits them. (Note that the authority need not be a single authority figure, authority can also come from groups.) We can demonstrate this by looking at what information is available to the learning child and how that information contributes to making the distinction between authority-dependent and authority-independent norms. Although it may be authority figures who are involved in teaching both types of transgressions, the informational differences available to the child are sufficient to make the distinction. I will look at four aspects relevant to explaining the relationship between norms and authority: the contribution of emotions, the contribution of caregiver instruction, the frequency of transgression, and the young child's conception of authority.

Emotions

As we have seen above, the emotional consequences of moral transgressions allow the learning child to see them as distinct from conventional transgressions. In canonical moral transgressions, these emotional consequences will occur irrespective of directives from authority (Nichols 2005; Prinz 2008b). The child can therefore determine the wrongness of the act independent of authority. This explains why when young children

witness events that are neither clearly moral or conventional, such as a bystander crying, the young child will typically judge the events as moral because they infer that the bystander was somehow harmed by the event (Smetana 1989). We also find that authority-independence is a property of other, non-moral emotional responses. Research by Shaun Nichols shows that non-moral disgust transgressions were ranked by children as less authority-dependent (and more generalisable) than conventional type transgressions. Disgusting acts are therefore perceived to be disgusting regardless of what anyone else says (Nichols 2002a, 2004).

In cases where an authority sanctions the harming of others, for the child the more emotionally salient consequence (harming others) will trump the other consequence such as the violation of the authority's proscriptions, thereby causing the child to appraise the action as wrong. This is not to say that considerations of harm and welfare are the only relevant factors enabling the child to determine what is morally wrong. What it does show, however, is that affect contributes significantly to the establishment of authority-independence as a property of moral norms. (I will address the issue of the authority-dependent nature of conventional norms below.)

Parental transgression responses

As we saw earlier, parents and caregivers specifically appeal to victim orientated consequences in the case of moral norms, and social regularity and authority in relation to conventional transgressions. Again, how the victim feels is typically independent of authority; children are being directed to identify those transgressions that cause harm as wrong because, first and foremost, they cause harm (Smetana 1997; Prinz 2008b). The young child may be taught this fact by an authority figure, but it is not authority that determines whether and event causes harm.

Frequency of transgression

Research shows that children experience the majority of moral transgressions with peers and siblings (Smetana 1989). Children compete with each other, hurt each other, take each other's possessions and have playground arguments. These interactions are reciprocal, involving both transgressor and victim, and are often resolved in the absence

of authority. In contrast, it was found that conventional transgression occurred evenly between peers and siblings, and caregivers. Importantly, then, the frequency of the child's experience with relations to authority in each domain is asymmetrical – a higher frequency of moral transgressions and their resolution is experienced independent of authority figures. Combined with the domain-specific responses to transgressions detailed above, we have clear differentiation between the domains in terms of their relations with authority.

So far, we have a story about how and why moral norms are perceived as authority-independent due to authority-independent emotional content, parenting practices directing them towards this content, and the frequency of transgression. Because emotional responses trump directives from authority, authority makes a diminished contribution when assessing emotionally salient moral transgressions. Young children need not infer any relationship between harm and authority to determine that an act is morally wrong. The dependence relation is not as complex as Joyce makes it to be. Still, this does not address Joyce's issue with respect to learning the authority-dependent nature of conventional norms, as it fails to explain how children can infer this relation without seeing the relevant authority change the norm. How do children know that if teachers change their mind about dress wearing that it is ok for boys to wear a dress to school, without actually witnessing this change of mind? To understand the classification of conventional norms as authority-dependent we need to look at the child's concept of authority and its relationship to social regularity.

Conceptions of authority

In many cases, the determinant of social regularities, such as conventional norms, will be authority, and, importantly, children view the legitimacy of authority as being dependent on context. Research shows that young children see authority figures as having specific jurisdictions. For example, 4-6 year olds see a school principal's authority as largely limited to the context of the school (Laupa 1994; Laupa and Turiel 1986), and there is plenty of evidence from the child's environment to establish this fact. In the case of schooling, children have ample evidence that teachers and school principals determine the social regularities of the school: they are told when they can and cannot talk, when they can eat, when they can play, and when they can go to the toilet; they get in trouble

when they do not obey the teacher; parents tell children to “do as you teacher tells you at school”. Additionally, teachers will tell a child “you will have to ask your parents that” if it falls outside their jurisdiction. Children experience teacher authority as context dependent and, in the child’s eye, teachers and school principles determine the social regularities of the school. This is true of other areas of life in which authority determines the rules, and children experience context dependent authority from a very young age.

Children therefore have specific ideas about which authority is relevant to which social regularities: “The findings show that preschoolers have understandings of the social context in which authority relations are embedded, conceptualising authorities not solely with respect to their adult status, but to their role within the social context of the school” (Laupa 1994:1). The child (tacitly) knows that the consequences of wearing a dress to school violates a social regularity (all children know that boys do not wear dresses to school!), and they know the relevant authority who determines social regularities at school. When the relevant authority changes the social regularity, then children will change their appraisal in relation to that new social regularity. Wearing a dress to school is a transgression because it violates the social regularity as determined by the authority at school, and it is not a transgression when it does not. Children do not need to observe specific instantiations of the particular authority-dependence relations of specific norms to draw this inference; they need only a general conception of the relationship between the relevant authority and the jurisdiction of that authority with respect to social regularities. We can conclude therefore, that the dependence relation in the case of conventional norms is not as complex as Joyce proposes.

Re-characterising the dependence relation

We can now re-characterise the relationship between transgression and authority. Children’s judgements of the dependence relation are not a case of rule extraction from the environment, followed by determining the relevant dependence relation with authority. Moral norms are not norms that also happen to be authority-independent; typically they are moral norms *because* they are instantiated by authority-independent events (such as harming). Similarly, conventional transgressions are not norms that happen to also be authority-dependent; they are conventional norms *because* they are instantiated by authority-dependent features (such as teachers’ dictates). In the moral

case, the young child's sensibilities to harm trump considerations of authority and, in the conventional case, there simply are no relevant authority-independent consequences to trump authority. I propose that we ought to see this in the justification that children provide for making such judgments. In the case of moral transgressions they will appeal to the harm and welfare of the victim, "it is wrong *because* it hurts them", not "it is wrong *and* it hurts them" (where 'hurting' is identified via an authority-independent emotional response). In the case of conventional transgressions, children will appeal to social regularity and authority: "It is wrong *because* the teacher told you not to do it", not "it is wrong *and* the teacher told you not to do it."

We can see these ideas more clearly when we consider the case of a young child learning that hitting is wrong. Children can quickly learn that harming events—being hit or scraping one's knee, for example—hurt irrespective of what people say. They also learn that the events that cause harms are wrong – they are told by their caregivers that hitting is wrong because it hurts others; they experience and witness the negative effects of being hit. They are sensitive to the harmful consequences of hitting and cannot conceive of harmless hitting occurring. When asked if it is ok to hit someone when an authority figure tells them it is, the child will still conceive of the hitting event harming, and therefore judge it as wrong. It is the harmful consequences of hitting that have been moralised by the learning child, not the norms from authority forbidding hitting. Although authority figures may teach a child that hitting is wrong because it harms, they do not determine whether hitting causes harm. In the case of conventional transgressions, there simply are not the relevant authority-independent consequences to over-ride the dictates of authority because the consequences are determined solely in relation to authority. It is wrong to wear a dress to school because an authority determines the rules about dress wearing at schools, no other reason. Therefore, children can conceive of the relevant authority telling them it is ok to where a dress to school and it no longer being wrong to do so. Children witness, and are involved in, conventional norm negotiations all the time, such as when playing games. Harming events, however, are non-negotiable.

Through development the child will learn the many subtle aspects of moral and conventional characterisations. Such a picture does not mean that authority has no influence or can alter moral behaviour or moral judgements; Nazi concentration camp officers and the Milgram Experiments in the 60's show people can, under certain

circumstances, obey authority in ways that contradict their normal moral beliefs and behaviours. Nor does it mean that authority does not play an essential role in moral learning. Authority figures guide moral learning, giving children the tools to make more nuanced judgements (Prinz 2008b). In the very early years, harm is sufficient for moral categorisation. As children develop, they can be taught relevant concepts such as “victim” and “intention” to supplement notions of harm. These more nuanced concepts then become relevant in establishing moral violations and their exceptions. For example, contact sports such as rugby, boxing, and gridiron are, by and large, excluded from moral appraisal because, although there is harm, there is no victim. In cases where a player harms another by breaking the rules, there is a victim and hence it is deemed morally wrong. Learning that some harm is not immoral is a case of learning exceptions to the early default position of categorising all harms as wrong. Although moral learning will often be complex, we can see that the internal and external resources available to the child allow him to identify the criteria that differentiate prototypical moral transgressions from conventional transgressions.

2.6 Conclusion

In this chapter we have seen that moral nativists make specific claims about the innate structure of the moral mind and the role that the environment plays in moral development. One argument in support of their view is that empiricist accounts of moral learning are inadequate to explain the move from informational exposure to mature moral competence, because the child’s environment is impoverished with respect to the information required for moral learning. This POMS argument combines a developmental claim and a task analysis claim. We have seen that the trajectory of moral developmental is consistent with a domain-general learning account. With respect to task analysis, I have shown that the (non-moral) internal and external resources available to the moral learning child radically narrow the gap between informational input and performance output. I have also argued that moral acquisition is not as complex as proponents of the POMS maintain. Finally, we looked at, and rejected, one specific argument in defence of the POMS – that criterion judgements (in particular the authority-dependence relation) can be accounted for without innate moral information. We have good reason, therefore, to doubt the poverty of the moral stimulus argument and the moral nativist’s positing of an innate moral faculty.

In chapter 1, I detailed the argument that morality is an adaptation. Yet in this chapter I have argued against an innate, domain-specific moral faculty. This raises questions as to how morality can be an evolutionary adaptation. In the following chapters I will argue that our moral cognitive systems are not bound to the head of the moral thinker nor even the boundaries of their body. Moral cognition is embodied and extended. This is important from an evolutionary point of view because it enables us to establish the proper parts of the moral cognitive system and hence the target of selection. In the final chapter I will detail how this non-nativist adaptationist picture fits within an evolutionary framework, namely a gene-culture dual inheritance model of the evolution of moral cognition.

2.7 Summary

In the previous chapter we saw that morality has a plausible evolutionary history; it is an adaptation. Being an adaptation does not necessarily mean that morality is innate, however. In this chapter I have argued against one particular argument in support of the claim that morality is innate. The main points are as follows:

- Poverty of the stimulus arguments have recently been applied to the moral domain in support of an innate moral faculty.
- The poverty of the moral stimulus (POMS) argument rests on two claims: a developmental claim and a task analysis claim.
- Against the POMS argument, I have argued that:
 - the development of moral cognition is consistent with an empiricist account of moral cognition;
 - both the internal and external resources available to the moral learner are sufficient for a domain-general account of moral learning;

- the moral learning task is either not an explicitly moral problem, or when it is, is not that complex; and
- we can account for children's grasp of the moral/conventional distinction without appealing to an innate moral faculty.

Having outlined why we have good reason to doubt that morality is underwritten by an innate, domain-specific moral faculty, I will now lay out the terrain for subsequent chapters, detailing traditional individualistic approaches to cognition and how some prominent nativist accounts of moral cognition are individualistic. I will then propose that moral cognition is best seen as incorporating bodily and highly structured external resources as part of the moral cognitive system. This has implications for the type of evolutionary account I will give in chapter 6.

3

Cognition: Embodied and Extended

If the object of study for a true cognitive science is cognitive systems, then of central explanatory importance is what we shall include as proper parts, and the locational bounds, of those systems. Is one's cognitive system confined to the head, or does the cognitive system extend beyond the bounds of the brain and associated neural structures? Further, what is the architecture of the cognitive system? Traditional approaches to cognition in the cognitive sciences and philosophy of mind have typically seen cognitive states as being realised by internal states of the agent – more generally “in the head” of the cogniser. The architecture of cognition has generally involved computational processes operating over amodal representational structures. Such a view relegates the role of perception and motor action to being peripheral inputs and outputs to the cognitive system.

There is a growing body of empirical and theoretical work questioning this traditionalist account, which raises questions as to how wide we should cast our explanatory net. Does a complete account of the inner complexities and structure of the agent's inner workings provide us with a complete account of cognition? What are the constituent parts of cognition? To understand cognition we must establish the scope of what we ought to include in our explanation. If our target is moral cognition, then we must establish the supervenience base of that cognition. There are two reasons why this is important for the larger evolutionary project. Firstly, establishing the proper parts of the moral cognitive system provides us with the explanatory framework within which to understand the

mechanisms of moral cognition. Secondly, it enables us to establish, from an evolutionary point of view, the mechanisms that were subject to natural selection.

In this chapter I will outline one traditional account of cognition and the particular commitments it makes with regards to the locational bounds and internal architecture of cognition. I will then outline how two recent approaches to cognition, which are relevant to our larger explanatory project, make the same commitments as the traditional account. Finally I will explain why we have reason to doubt this approach to cognition. This will set up the next few chapters, which will describe how moral cognition, contrary to the traditional approach, is both embodied and extended.

3.1 Traditional accounts of cognition

Most recent accounts of moral cognition have, explicitly or implicitly, been individualist theories of some kind. They fit with traditional approaches to cognitive science and psychology that place cognition firmly inside the head. The dominant methodological and metaphysical approach to studying cognition has been, and still is, *individualism*.¹ Individualism is a methodological doctrine: cognition is explained through the study of the internal states of individuals. It is a metaphysical doctrine in that the individualist approach views cognitive phenomena as instantiated by the internal structure of individual agents, and by this internal structure alone. Thus the structure of the mind is an *internal* structure, and mental states and their structures are characterised and individuated independently of that which is external to those structures.

Individualism about psychology is the thesis that psychological states should be construed without reference to anything beyond the boundary of the individual who has those states. Put loosely, it is the view that for the purposes of scientifically understanding the mind, the individual is the boundary of cognition. (Wilson 2004:398)

Drawing the boundary of cognition in terms of the individual raises obvious issues with

1. Not to be confused with the individualism/anti-individualism debate over whether or not intentional mental states are individuated solely by the intrinsic properties of an individual or whether those states are individuated in terms of intrinsic properties and their relations to the external world. (Putnam 1975; Burge 1979).

respect to what constitutes the “individual”, and therefore where exactly the boundaries of cognition lie. Following the methodology of much of traditional cognitive science, I will take the bounds of cognition as typically posited by individualism in the cognitive sciences to be confined within the brain, or what Andy Clark refers to as being “brainbound” (2008).² Here we see an important contrast between the brain (and certain associated neural systems) and the non-neural body and external world. In more general terms, the human cognitive system is “confined to the head of the thinker” (Chemero and Silberstein 2008:3). Confining the cognitive system in this way means that the (non-neural) body and everything beyond is excluded from the cognitive system. This is not, of course, to claim that traditional accounts of cognition deny that one’s body and environment play important roles in, or that they are irrelevant to, the processes of cognition. But as we will see, these approaches attribute a very specific role to both the body and world, and the ways in which they interact with the mind – namely, they function as the inputs and outputs to the cognitive system but are not constitutive parts of high-level cognitive processes. For what is central to the individualist’s claim is that the actual processes of cognition itself (i.e. the proper parts of the cognitive system) are divorced from both the body and the environment:

The world is (just) a source of inputs and arena for outputs; the body is just an organ for receiving inputs and effecting outputs (actions); the task of early processing is to render the inputs as an inner world-model of sufficient thickness to allow the bulk of problem-solving activity to be defined over the inner model alone.
(Clark 2004:35)

To study cognition is to study what goes on in the head.

The dominant *individualist* views of cognition, since the advent of modern cognitive science, have been *computational* theories of mind, according to which cognition takes place via the *inner* manipulations of *internal* mental representations.³ There have been

2. Brainbound refers to the entire central nervous system (brain and spinal chord).

3. Note that not all individualist accounts of cognition are computational. For example, some have proposed that cognition is individualistic and best explained “in terms of large scale neural dynamics” as opposed to computation (see Chemero and Silberstein (2008) for a brief discussion). Nor are all computational accounts of cognition individualistic. But paradigmatic cases of computational theories of cognition typically are individualistic and vice versa. The point I will be pursuing in the following chapters is that what counts as a moral cognitive system (computational or otherwise) cannot be confined to an individualistic methodology nor metaphysics.

numerous accounts of how this computational architecture of the mind is actualised. They include, but are in no way exhausted by, formal logical approaches, rule based systems, concept based systems, image based systems and various connectionist systems to name but a few (for a brief overview see Thagard 2008). Each proposes different accounts of what constitute representations and/or the systematic processes which act upon those representations. What they have in common is that they all converge on a particular explanation of how the mind works, namely through *computational procedures operating upon internal representational structures*: “thinking is essentially having and manipulating representations” (Bem and Looren de Jong 2006:280).

Take, for example, the two prominent individualistic accounts of cognition: the Computational Theory of Mind (most prominently associated with Jerry Fodor⁴ and associates (Fodor 1975; Pylyshyn 1984)) and connectionism (see for example Rumelhart and McClelland (1986)). Both approaches are computational. Cognition is best understood as computational algorithms acting upon representational structures (although some connectionist approaches claim to be neither computational nor representational). One of the main points of difference between these two accounts is what constitute the representational structures upon which the computations occur. (Chemero and Silberstein 2008; Thagard 2005; Bem and Looren de Jong 2006).

The Fodorian style computationalist, for example, takes representations to be symbolic, formal language-like structures where cognising involves the manipulation of these structures. The claims of the Fodorian are heavily influenced by Chomskyan linguistics’ picture of an innate, species universal, generative grammar. Fodor, for example, holds that the mind has an innate, species universal, language-like mental architecture that accounts for the systematic structure and productivity of thought – the language of thought hypothesis (1975). (For similarities with the Principles and Parameters account of moral cognition, see section 3.2.1 below). Connectionism on the other hand views representations as activation patterns in neural networks. By applying particular rules of learning, the connectionist network is able to acquire exemplar or prototype based

4. To be fair to Fodor, he is at pains to explain that he does not think that a Computational Theory of Mind (CMT) accounts for all, or even most of our cognitive processes: “I certainly don’t suppose that it could comprise more than a fragment of a full and satisfactory cognitive psychology” (2000). This cannot be said for many other proponents of CTM, (especially those with an evolutionary psychological bent) whom Fodor targets in his book *The Mind Doesn’t Work that Way: The Scope and Limits of Computational Psychology* (2000).

conceptual representations. Cognitions (such as those involved in reasoning and inference making) operate by way of pattern completion processes acting over these representations. (See Churchland (1996) for suggestions as to how this type of approach might work in the moral domain.)

In each case, cognition consists in the computational processing (algorithmic manipulations or pattern activation) of representations (symbols or sub-symbolic activation patterns).⁵ What typifies these perspectives is that the mind possesses *rich, internal (symbolic or sub-symbolic) structure*, and the essence of cognition involves the manipulation of this structure. The model of the mind is that of an inner mental world populated by representations upon which the problem-solving activity is enacted. Problems present in the environment are represented in the head. Cognition then solves the problem via the representational manipulations in the head, and produces the relevant output to which the body acts.

The computationalist theories of mind typically view the representational structures over which cognitions operate as being amodal – that is, separate from perception and motor systems which interface with the world. Perceptual inputs are transduced to amodal symbols over which cognitions occur, the output of which is transduced to motor response. In this way, modal sensory-motor systems (involving representational structures or otherwise) are seen as separate from high-level cognising, whereby the perceptual and motor systems provide the inputs and receive outputs from that cognitive process. Sensory-motor operations are peripheral to cognition proper. (See for example Pylyshyn 1984).

The above discussion has identified a number of positions (brainbound individualism, computationalism, and amodality), which need not be taken as a package deal. Where one draws the bounds of cognition need not commit one to any particular mental architecture nor to any particular view of the structure of representations. Moreover, how one views mental architecture need not commit one to any specific position on the bounds of cognition nor the modal nature of representations (if there are even any

5. There are debates in philosophy of mind as whether cognition is identical to computational processes, or whether cognition is realised by computational processes. For present purposes, I am neutral on these matters.

representations at all!). Taking all three together, however, does establish the traditional view of cognition: realised in the brain and involving computational processes acting over amodal representations. This tripartite account is what I will thus refer to as individualism.

3.1.1 A mental world of inner complexity

Individualistic accounts of cognition focus on representations and processes acting over those representations rather than physical interaction, and in doing so make a specific commitment to inner complexity. As I have outlined above, the individualist's approach to understanding the human cognitive system is to look "inside", for the cognitive system is bound within the brain and nervous system. Of central importance to individualistic explanations is internal structure and organisation. The most prominent individualistic accounts of cognition posit mental representations to account for this structure.

Mental representations are theoretical entities; they are posited by the sciences of the mind because they have rich explanatory value in that they account for the relationship between mind and world. They do this by doing what their name suggests, namely *representing*. Although accounts differ as to how representations actually go about representing the world, they are minimally construed as "causal surrogates for distal features of the environment" (Chemero and Silberstein 2008). In this sense, they bring aspects of the world inside, and in doing so allow cognitive processing *about* the world to occur in a manner that is *independent* or *isolated* from the world.⁶ Yet if this picture of the mind is correct, and our cognitive system is instantiated by the brain and neural systems, then it is not just the environment for which mental representations can be causal surrogates; they can also be causal surrogates for the non-neural body, its configurations and its movements.

As explained above, for computationalists such as the Fodorian and the connectionist this internal complexity is typically implemented by way of either symbolic or sub-symbolic structures. They often differ as to whether or not, and to what degree, these internal structures are innate or acquired. Fodor once famously (if not notoriously) held

6. Hence Clark dubs this form of inner world problem solving "isolationism" (Clark 2004).

the view that most of our concepts are innate (1975; 1981). On the other hand, connectionists working with neural networks often place much emphasis on the powerful associationist learning abilities of such networks. Although they differ somewhat as to the exact nature of the computational processes that act upon those structures, both accounts of the mind understand cognition as abstracting away from the physical and social environments in which cognitive processes are taking place (Wilson forthcoming; Marr 1982). If cognition is the proximate cause of behaviour, then this behaviour is due in part to the internal complexity of the cognitive system – be that complexity learned or otherwise.

What is explanatorily important from the individualist's perspective is that adaptive problems encountered in the world become problems to be solved inside the head. External complexity of the problem domain becomes brainbound internal complexity. We build complex inner models of complex outer worlds. To explain cognition and resultant behaviour is to look inwards.

3.1.2 The body and world

Since the individualist believes the cognitive system resides in the head, any explanations of cognition proper are themselves going to be confined to the head. Such reliance on internal models or representations in explanation *reduces* the importance and role of the body and environment (Haugeland 1998b). Not only does this make a specific commitment to the internal complexity of the mind, but in doing so these individualistic accounts relegate the environment and body as being external to the cognitive system. The role of the body and world is to operate as the inputs and outputs to cognitive process, with the environment providing the adaptive problems to be solved and information for their solution (i.e., the inputs to the cognitive process). The body mediates those informational inputs (via perceptual apparatus), and enacts the action orientated outputs (via motor structures) to the environment. The environment is typically viewed as the problem domain and our perceptual apparatus as a “peripheral channel through which the problem is initially posed, and incidental facts are supplied” (Haugeland 1998b:220).

Although the individualistic perspective does not deny that the environment and body both play important roles in cognition, it is clear that they do make a particular commitment to the nature of their role. The environment populates the contents of our minds over multiple timescales: the environment to which aspects of our minds are adapted, and those life-time environments within which cognition takes place. In each case, the individualist's mind is furnished via some well defined, stable channels; be that through genetic channels such as inherited innate representational content (such as our ability to focus on faces soon after birth), or over one's life history via perceptual inputs from the world around us. The body in this case provides a stable structure through which the mind can receive inputs from and interact with a variable world. The computational algorithms are constructed and operate with certain assumptions about the structure of the body and its properties that enable those algorithms and heuristics of the cognitive system to work reliably and successfully. In sum, we have environmental inputs via stable perceptual channels, cognition operating on representations of those inputs, and output through stable bodily structures with which to interact with the world.

So far we have seen that traditional accounts of cognition make specific commitments relating to the bounds and architecture of cognition. This view also makes specific commitments to the internal complexity of internal cognitive structures and the roles attributed to body and world. We will now look at the ways in which two prominent accounts of cognition are individualistic. Both have relevance to our larger explanatory project, the evolution of moral cognition.

3.2 Traditional cognition and the moral mind

Individualism is pervasive. It is inherently assumed in many, if not most, of our psychological and biological sciences (Wilson 2004). Moral cognition is not immune from this dogma. I will briefly look at how two accounts of cognition are either explicitly, or tacitly, individualistic in nature. The first is a recent approach to moral cognition; namely the principles and parameters account of moral cognition recently presented by Marc Hauser (2006a, 2006b) and John Mikhail (2007, 2008a). The second is a more general

account of cognition proposed by modular evolutionary psychology⁷ (Barkow, Cosmides et al. 1992; Wright 1994; Pinker 1997; Buss 1999). Both approaches make specific individualistic commitments. Cognition is internal and the driver of complex behaviour is internal. These approaches ignore, or at least tacitly relegate, the role of the body and world and the dynamic interplay between brain, body and world as being external to one's cognitive system. Both also involve computations operating over representational structures. I will look at the two approaches in turn.

3.2.1 Individualism I: the Hauser-Mikhail model of the moral mind

Marc Hauser (2006a, 2006b) John Mikhail (2007, 2008a) Susan Dwyer (1999, 2006, 2009) and Gilbert Harman (2000, 2008) have proposed accounts of moral cognition that are explicitly influenced by Chomskyan accounts of linguistics and the computational theory of mind (as seen in chapter 2).⁸ The internal structure of the moral mind, like Chomskyan accounts of language cognition and Fodorian accounts of cognition in general, relies on formal internal structures representing external states of affairs which are acted upon by computational manipulations. The Hauser-Mikhail⁹ model of the moral mind posits internal complexity in the form of a moral grammar, the structure of which is universal, abstract, and context independent.

Moral grammar is *universal* in that it is claimed that all normal human beings have this innately (genetically) encoded moral grammar which is required for moral acquisition, normal moral development and moral competence. The underlying principles of the moral grammar determine the space of possible moral variation and which variations take shape will depend on the various moral parameters in our developmental environments. Moral grammar is *abstract* in that the underlying principles are abstract structural descriptions of moral states of affairs that include such parameters as intention, harm, ends, means and side-effects to name but a few (see for example Mikhail (2002, 2008a)

7. As I am using the term here, I will understand 'evolutionary psychology' to be the research programme that makes specific commitment to innate cognitive modules selected for in the Environment of Evolutionary Adaptedness (see for example Barkow, Cosmides et al. 1992; Pinker 1997; Buss 1999). I am not referring to other psychology research programmes that are evolutionary in nature.

8. Mikhail, for example, has a recent paper entitled *Moral Cognition and Computational Theory* (Mikhail 2008a).

9. I will exclude Dwyer and Harman from the name of this model for brevity's sake.

and Dwyer (2006), for detailed exposition of this structure). And it is *context independent* in that the underlying structure of the innately specified principles are themselves invariant to situational effects.

The moral grammarians also claim that our internal universal moral grammar explains the structure and production of our society-wide moral systems: “[t]he universal moral grammar is a theory about the principles that enable children to build a large but finite range of distinctive moral system” (Hauser 2006b:421). Although the underlying structural principles of moral grammar are fixed, the systematic differences between cultures are due to the various parametric variations as set by the culture. Moral grammar can explain both the systematicity and productivity of our moral systems. In this way the “universal moral grammar provides a toolkit for building possible moral systems” (Hauser 2006a:215).

According to the Hauser-Mikhail account, what is centrally relevant to understanding moral cognition (and moral systems) is explicating the internal structure of the mind by cashing out the underlying moral grammar through the analysis of the relevant input conditions and resulting output responses (for detailed application of this methodology see Mikhail (2008a)). This model of cognition is explicitly computational. Take for example the following passages from Mikhail, which make clear references to a computational account of cognition:

the human brain contains a computationally complex “moral grammar”... analogous in certain respects to the mental grammars operative in other domains, such as language, vision, music, and face recognition. (2007:l)

the problem of descriptive adequacy in the theory of moral cognition may be divided into at least three parts: (1) the problem of describing the computational principles (“deontic rules”) operative in the exercise of moral judgment, (2) the problem of describing the unconscious mental representations (“structural descriptions”) over which those computational operations are defined, and (3) the problem of describing the chain of inferences (“conversion rules”) by which the stimulus is converted into an appropriate structural description. (2007:4)

The role of the body in each of these accounts is that as described in one of Andy Clark's critiques of traditional cognitive science, whereby "the body is just an organ for receiving inputs and effecting outputs (actions)" (Clark 1998:36). Hauser claims, for example, that emotions occur *after* the moral computational process (i.e., the making of a moral judgement) has been engaged by making the distinction between "emotions playing a role in our moral judgements and emotions playing a role in our moral behaviour, or what we actually do" (2006b:195). Hauser claims that, by and large, the latter exemplifies the role of emotion in moral cognition such that emotion plays no causal role with respect to the permissibility of moral judgements.¹⁰ One of the centrally important differences between the moral grammarians' model of moral cognition and other emotionally grounded accounts is that emotions typically follow from the judgement-making process rather than having an essential causal role. Hauser hypothesises that "[e]motions may only function to modulate what we actually do as distinct from what we comprehended or perceive as morally permissible" (2006b:46) ... and "our emotions are not part of the dedicated and specialized components of the moral faculty" (2006b:53). Emotional engagement occurs after the moral computations have been made and operate as a way of controlling and directing our subsequent moral behaviour. Here our moral computational system (i.e., the moral faculty) is employed *before* our bodily (emotional)¹¹ processes are enacted – the role of emotion is to (when required) enact the outputs of the moral faculty in the environment. Under such a view emotions may be central to moral performance but not moral competence. I will later argue that emotional information processing involves the body, and that moral cognition involves emotions, hence the body is part of moral cognition (see chapter 4).

The role that the Hauser-Mikhail model ascribes to the environment and body is that of the problem domain and effectors. The environment presents us with a problem, such as in the case of a moral violation. The information from the environment is in turn presented to our moral mind through our perceptual apparatus. This triggers an analysis

10. It should be noted that Hauser does not necessarily think that this is *always* the case. He concedes that it is possible that emotions may play a causal role in *particular* classes of moral judgements such as "personal/high conflict/other-serving dilemmas" (Hauser 2006a:218). Hence Hauser has recently entertained the possibility of a plurality of moral psychological processes that include emotion driven ones (Hauser, Young et al. 2008b). More recently, Hauser and colleagues suggest that emotion could provide input to the moral circuits, rather than modulate the operation of those circuits (Huebner, Dwyer et al. 2009) – cf. the role of the body in traditional cognition.

11. The assumption I am making is that emotional processes include bodily effects or that their processing involves bodily effects. See chapter 4 for discussion.

of the information with respect to the causes and consequences of the action or event, from which a moral judgement is made. This judgement itself produces outputs which are (often) enacted through our bodies. This is clearly what Clark describes as a “linear processing cycle: perceive, compute and act” (Clark 1999:346). The appropriate moral behaviour is caused from the inside; we are moral in the ways that we are *because* we have algorithms that specify our moral judgements and behaviours.

The moral grammarians, akin to traditional views of cognition, can also be read as taking the relevant representational structures of the universal moral grammar as being amodal. For example, Hauser writes that “the principles of [moral grammar] operate on experiences that are independent of their sensory origins, including imagined or perceived visual scenes, auditory events, and all forms of language—spoken, signed, and written” (Hauser 2006b:53). And Mikhail states that “[t]hese concepts and principles which underlie [our unconscious moral knowledge] are as far removed from experience as the hierarchical tree structures and recursive rules of linguistic grammars” (2008b:355). Amodality is central to their version of the poverty of the moral stimulus argument and hence a core feature of their model. Somatosensory and sensory-motor systems are quite separate from the abstract structures of the universal moral grammar.

Whilst the above is only a brief overview of the moral grammarians’ position, the central themes are clear: The central causal determinants of our moral cognitions are internal brainbound cognitive mechanisms that are realised independently of body and external environment; bodily and environmental resources are not considered part of the moral faculty (qua moral cognitive system); the structure of our moral systems and the scope of cultural moral variation are constrained by the underlying principles of our innate moral grammar (see discussion of evolutionary psychology below); the moral cognitive system is firmly in the head. As we will see, this position does not capture the full gamut of moral cognitive processes and hence does not adequately explain moral cognition. An individualist moral cognition fails to account for the indispensable role that body and environment play in providing solutions to adaptive moral problems.

3.2.2 Individualism II: evolutionary psychology

The Hauser-Mikhail model of the moral mind is a special case of a second, more general type of individualism, evolutionary psychology. The human mind, according to evolutionary psychologists, consists of a collection of genetically endowed, species-universal, specialised cognitive mechanisms, or *Darwinian algorithms*. Those mechanisms are computational in nature. The way that natural selection accordingly shaped our species' psychology was by selecting those Darwinian algorithms that solved specific problems in our evolutionary history. The architecture of the mind according to the evolutionary psychologist is computational, innate (genetically encoded), almost entirely modular, and encapsulated (Barkow, Cosmides et al. 1992; Buss 1999). Evolutionary psychology is, by and large, individualistic (Wilson 2004).

The main aim of evolutionary psychology as a scientific enterprise is to explain these algorithmic mechanisms (i.e., the proximate causes of our behaviour) in terms of the selective pressures (ultimate causes) that hominin encountered in the Pleistocene. As Cosmides et al. explain:

By understanding the selection pressures that our hominid ancestors faced—by understanding what kind of adaptive problems they had to solve—one should be able to gain some insight into the design of the information-processing mechanisms that evolved to solve these problems. (1992:9).

The important explanatory consequence of this methodology was that “[e]volutionary psychology “went inside,” shifting the explanatory focus from behavioural patterns to psychological mechanisms” (Sterelny 2003:234). For example, it is often claimed by evolutionary psychologists that we have evolved specialised cognitive mechanisms for cheater detection. In our evolutionary environments, where cooperative interactions were central to survival, the threat of partner defection posed a real risk. Being able to identify trustworthy partners and cooperating conditionally has obvious fitness advantages over being an unconditional cooperator. Evolutionary psychology proposes that we have accordingly adapted cognitive mechanisms to enable specific reasoning in the social domain. Evidence for this, it is claimed, can be found in our superior ability to reason in this domain (as opposed to the domain of logical reasoning) as shown by the Wason

selection task (Cosmides and Tooby 1992).

In the moral sphere, many evolutionary psychologists claim, among other things, that evidence suggests we have cognitive adaptations for many moral behaviours¹² such as those that regulate sexual behaviours towards family members (Lieberman, Tooby et al. 2003; Lieberman, Tooby et al. 2007; Lieberman 2008), a social contract mechanism (Cosmides and Tooby 1992)¹³, and moral emotions such as guilt (for a populist take see Wright 1994). Our particular behaviours in response to different moral dilemmas occur because we evolved computational algorithms designed specifically to respond to such dilemmas. Internal cognitive complexity gives rise to the external complex moral action.

The evolutionary psychologist's individualism moves beyond explaining individual behaviours, to explaining the generation of culture as well (Wilson 2004). Two of the leading proponents of evolutionary psychology, Cosmides and Tooby (1992:115-116) claim for example that much of human culture is "evoked culture", not transmitted from generation to generation but instead evoked from our universal human psychology. Similar environments provide similar informational-inputs to our universal cognitive mechanisms and therefore produce similar behavioural outputs which manifest at a group level as culture. Similarities between cultures occur due to the manifestation of those behaviours produced in similar environments. Much of the cultural variation that we see occurs because the cognitive information processing mechanisms receive different informational inputs when in different environments, thereby evoking "different representational and behavioural outputs", and hence different culture. Cosmides and Tooby liken this to a juke box that will play different songs depending on local preferences. (See Tooby and Cosmides 1989; Cosmides and Tooby 1992; Boyd and Richerson 2004.) Internal psychology determines the structure and shape of culture:

Group-level cultural and social phenomena, while they have some emergent properties, are the *consequence of the operation of evolved psychological (and morphological)*

12. I read the moral grammarians as being a special case of evolutionary psychology. Hauser, for one, makes direct appeals to the evolutionary origins of the universal moral grammar (2006b).

13. Cosmides and Tooby write that many of the computational requirements are "so particular to adaptive problems that arise in social exchange that they could only be implemented by a computational system whose design was functionally specialized for this function" (1992:53).

mechanisms functioning in individuals who evolved to live in groups. (Tooby and Cosmides 1989:45, italics added)

Evolutionary psychology views cultural phenomena as the aggregation of individual Darwinian algorithms giving rise to individual social level behaviours (Wilson 2000, 2004). Culture operates as the cause (setting the various parameters¹⁴) and effects (aggregation of behaviours) of an internal psychology, but they are in no way constitutive of that psychology. Again, we see the input → computation → output model of cognition and its aggregative generation of culture.

Hence, the explanation for both individual behaviours and the cultural manifestation of those behaviours is individualistic: “The cognitive machinery is organism bounded, and the properties of organism important for the evolution of both cognition and culture supervene on what is inside those organisms” (Wilson 2004:416). Again, cognition is in the head and those brainbound psychological mechanisms are the central determinants that “create, shape, and maintain culture” (Tooby and Cosmides 1989:45).

Having explained the traditional view of cognition in the cognitive sciences, and outlined how one recent account of moral cognition and how evolutionary psychology is consistent with this view, I will now look at alternative accounts of cognition which do not adhere to the traditional picture. These alternative views emphasise the role of both the body and the world as part of our cognitive systems. In the following chapters I will argue that moral cognition, contra to both the Hauser-Mikhail model and evolutionary psychology, is best captured by these alternative views.

3.3 Explanations beyond the brain

There is no doubt that cognition is a complex phenomena and that the cognitive machinery inside the head is very complex indeed. It is not the aim here to question that *much* of cognition is confined to the brain and neural systems. What is at question,

14. “By feeding modern conditions (economic, cultural, social, etc.) as parameters into these algorithms, modern behavior can then be both predicted and understood – not simply as adaptive or maladaptive, but as a consequence of the structure of the mechanisms that regulate behavior and the modern conditions that are their input” (Tooby and Cosmides 1989:37).

however, is whether or not the human (and especially in this case the moral) cognitive system can (always) be viewed as a *distinct* subsystem within the greater brain-body-world (whole) system (Clark 1997; Haugeland 1998b). Can we make a principled distinction by which we draw the boundary of cognition at our neural structures? Or, does cognition “leak into the body and world” (Clark 2008).

There are several ways in which physically external resources can become part of our cognitive system. Often these will run together. Firstly, external informational resources can play a similar role to inner informational resources such that there is no cognitively significant distinction that can be drawn between the two (this is the ‘parity thesis’ which we will encounter below.) The second is the idea of a dynamic interplay between brainbound processes, body and world. This is in contrast to the traditional picture of a linear flow of perception → cognition → action. Finally, external resources can become part of our cognitive system when the cognitive representational complexity that suffices to explain adaptive responses to a complex world is not wholly brainbound. That is, there is not enough in the head to explain intelligence.

In contrast to the moral individualism discussed above, I argue that we cannot confine our studies of moral cognition to the head. Explaining moral cognition will involve the application of one or more of these three means of cognitive extension. In other words, much of moral cognition is best explained by paying attention to the role of external information structures and our interactions with those structures, rather than from the “inside”. Very often the body and world are not mere inputs and outputs to the moral cognitive system but play an *active* role in the adaptive problem solving routines we employ to navigate our moral worlds. We cannot have a complete explanation of moral cognition, its acquisition, deployment and evolution without accounting for the dynamic interplay between brain, body, and world.

3.3.1 Cognition beyond the head

The temptation to give internalised explanations of cognition is understandable, if only because our phenomenological experience appears to be from the “inside”. But, the brain centred view, to use Andy Clark’s phrase, is “not true to the natural facts” because it fails to appropriately account for the role of external resources in our cognitive systems. Such

views have “an unmistakable tendency to marginalise such factors: to dwell on inner complexity whilst simplifying or ignoring the complex inner-outer interplays that characterize the bulk of basic biological problem-solving” (Clark 1998:35). Additionally, there is often an assumption within the science of the mind that behavioural complexity is to be causally related to internal cognitive complexity (Haugeland 1998b; Barrett, Henzi et al. 2008). As above, moral cognising (such as the making of moral judgements, tracking of other agents’ moral behaviours, having moral beliefs, taking moral action etc) is often considered to be essentially linked to the complex internal make-up of the moral mind. Yet not all complex behaviours are necessarily internally driven. John Haugeland draws on an example from Herbert Simon to bring attention to this type of bodily and environmental marginalisation, while at the same time highlight the role of agent-environment interplay and the importance of external factors. He asks us to imagine an ant walking across a “wind- and wave- moulded beach” (Simon as quoted in Haugeland 1998a:209). The movement of the ant will be in “response” to the contours of the sand and pebbles which make up the beach surface, weaving and winding over and around obstacles as they come to pass. The trajectory will be “irregular, complex, and hard to describe” (Simon as quoted in Haugeland 1998a:209).

If we take the explanatory target to be the ant’s behaviour as it walks across the beach, there are two explanations that we can proffer. One is by way of an individualistic account, remembering that the methodology of an individualistic explanation is to explicate *internal structure* because cognitive systems are “confined to the head of the thinker”. The internalist picture of the ant’s complex behaviour is that it is determined by the computational algorithms acting over internal representations (i.e., the causal surrogate for distal features in the environment, in this case the contours of the sand, the size, shape and placement of the pebbles etc). What guides the complex behaviour in this case is the mechanisms of the internal (brainbound) cognitive system.¹⁵

In contrast to the internalist’s brain centred account of cognition is one in which the cognitive system includes not only the brain, but the body and world and their tripartite interaction. The complexity of the behaviour is explained in terms of this interaction, not

15. It is worth noting the strong parallels with many current accounts of moral cognition, in that what guides the agent’s moral response is an internal set of predefined rules, principles, or responses – behavioural complexity is essentially determined from the inside.

solely by the internal states of the agent's brain. In the case of the ant walking on the beach, the complexity of the behaviour is due to the real-time interaction (close coupling) between ant and beach; the complexity of the beach surface matters for our explanation (Haugeland 1998b). The explanation cannot involve solely the ant, but requires an account of the dynamic, real-time interaction between both ant and environment. "If ... there is constant close coupling between the ant and the details of the beach surface, and if this coupling is crucial in determining the actual path, then, *for purposes of understanding that path*, the ant and the beach must be regarded more as an integrated unit than as a pair of distinct components" (Haugeland 1998b:217, italics added). In this way we treat the details of the body and world and their interaction as part of the explanation for the system. Note also, by shifting the focus onto real-time agent-environment interaction we reduce the heavy reliance on internal representational structures. The ant need not have a full representation of the beach because the real-time interaction uses the *world as its own best model* with problems to be solved in real-time as they arise (Brooks 1991). We see more pronounced examples of cognitive integration in the real-time, adaptively complex skilled use of tools explained below.

By treating the body and world as the inputs and outputs of the cognitive system, the individualist explanation fails to capture an important facet of cognition, namely that much of the complexity of our cognitive system resides not in the internal make-up of the brainbound system but in the dynamic and contingent interplay between brain, body, and world. Complexity is often external to our neurological structures, and importantly, this complexity is often part of our cognitive system. This is not to say that all and every cognitive process involves such dynamic interplay. One of the trends in evolution has been to internalise cognitive resources in a manner which allows adaptive cognitive behaviours to operate in multiple, varying environments – internalised cognitive processes can be both portable and adaptive. Kim Sterelny suggests that one factor driving the move to an internalised cognition is that external resources are disruptable and exploitable. By internalising those resources we protect them from corruption (2003, 2004 #401; Clark responds in 2008). Humans, most notably, can decouple their cognitive processes from their environments. Cognising of this kind involves decoupled representations, which are "registrations of the environment that are relevant to many possible aspects of the environment but functionally specific [i.e., tightly coupled] to none" (Sterelny 2003:50). So unlike the beach travelling ant, human cognising about

beaches and beach associated behaviours can occur independently from real-time interaction of the beach. Some cognitions and the states that realise them are therefore going to be more internalist in flavour than others. (See Clark's (1997) discussion on 'representation-hungry problems'.) There are some cases in which our explanation is not enhanced by "beyond the brain" explanations. But the important point is that sometimes more complete explanations will involve recourse to the interplay and close coupling between mutually influencing internal and external resources. Much of moral cognition, I will later claim, falls into this category.

It is worth looking at a few examples of extended cognition to put this into a human context. Take, for example, the oft cited example of the traditional way in which people use a pen and paper when engaging in such activities as calculating long division or multiplication (Clark 1997, 2008; Giere and Moffatt 2003; Wilson and Clark 2009). When trying to work out long divisions that are too complex to solve purely in the head, we often rely on pen and paper to complete the process. In doing so, we are using written symbols (numbers) as external representations. Through the use of the pen and paper we can manipulate these representations. The process by which we complete the calculation involves the coordinated integration of brain, body, and world. It is the *entire* system that completes the cognitive task in this manner, not the isolated brain, and in this way cognition is *distributed* over all three interacting components.

Kirsh and Maglio (1994; see also Maglio, Matlock et al. 1999; Clark and Chalmers 1998; Clark 2008) present compelling accounts of how we offload computation processing to external artefacts to facilitate cognitive tasks. One of their key studies is the computer game Tetris. Playing the computer game Tetris involves the manipulation of falling geometric shapes. The aim of the game is to manipulate them in such a way that they can be arranged to form complete patterns (rows) across the bottom of the screen. This is done by rotating and moving the shapes across the screen as they fall. The game is often played at high speed and involves real-time, fast perceptual and cognitive performance. One would expect that as players gained experience they would be more precise in their play, that is, they ought to make fewer moves to complete the tasks required. But, contrary to expectations, Kirsh and Maglio found that the more experienced players actually rotate the shapes more than inexperienced players do; they do not always make the least number of moves to complete the task. The use of extra rotations actually

increases with skill level. They use, what Kirsh and Maglio call, “extra extraneous actions” to achieve the task.

Kirsh and Maglio conclude that these extra rotations alter the player’s epistemic environment. Rotations on the screen change the information available to the player and help the skilled player make decisions about where is best to place the shapes to complete the patterns. Rather than work out the required rotations in the head, the more experienced players manipulate the shapes on screen, simplifying perceptual computations. By incorporating the manipulation of external artefacts into their cognitive problem solving routines, skilled Tetris players offload perceptual computation onto their environment. Kirsh and Maglio call this type of utilisation of the external environment ‘epistemic actions’. They improve cognition by reducing memory load, the number of steps in the computational task, and the probability of error. They contrast this with ‘pragmatic’ actions which aim to bring agents closer with their physical goals. The use of epistemic actions reduces internal cognitive load by altering the information environment, ensuring we can access information as and when we need it. In this way, physical manipulations of the environment allow cognitive processing to be offloaded on to our environments. As Kirsh and Maglio tell us, “better players use the world better” (1994:513).

The individualist may respond that although the agent organises her workspace to simplify problems and enhance performance, the internal structures of the human agent are still the central organising structures. Human brains have explanatory primacy because they are the active ingredient. It is no doubt true that, in such cases, the human brain is the central organising structure. Yet, the complete explanation necessarily includes the constant but dynamic, mutually influencing information feedback loops between agent and Tetris console. The cognitive load is spread over the agent–environment structure. As Clark and Chalmers (1998) explain, if this extended cognitive task were performed within the brain of the agent, we would rightly define it as a cognitive task (see also chapter 5). The problem solving itself cannot be readily apportioned to computations operating on inner representational structures alone (for detailed discussion see Clark 2008). The cognitive system extends into the environment and in this way is said to be *extended* cognition.

Maglio and Kirsh and colleagues (1999) explore similar uses of external resources as integrated cognitive tools in their studies of people playing the game of scrabble. Scrabble players typically reorganise the sequence of their letters to more easily identify potential words hidden amongst their scrabble tiles. In doing so, the players perform a cognitive task by altering their environments. People reorganise physical space to alter their epistemic access to information. Players do not have to reorganise the scrabble letters in their head because they can physically reorganise them in the world. They therefore reduce the demands of internal cognitive processing and increase successful task accomplishment by offloading some of the work to the environment. The nature of the task is changed from a cognitively demanding internal process into a perceptual one, reducing memory demands and allowing the agent to focus on real-time perceptual based pattern completion processes. In this way, we incorporate the physical world into our adaptive problem-solving routines such that the cognitive system involves the interaction of brain, body and world. It is the entire system that achieves the cognitive task, not the isolated brain and associated neural system. We often use external resources in similar ways to transform the nature of the computational problems we encounter (Clark 1997, 2001, 2008).

In each of the cases above, focusing on the internal workings of the head will not provide the best explanation for the phenomena to be explained. The best explanation is that the physical manipulations of the Tetris shapes and the scrabble tiles, and the coupling between agent, pen, and paper when doing long division, are integrated cognitive resources. Similarly, the best explanation of the path travelled and behaviour of the ant involves the close, real-time coupling of the ant and the beach. The claim then, is that the traditional theoretical frameworks in which the cognitive sciences operate are insufficient to explain the cognitive system as the individualist perspective excludes explanatorily significant cognitive elements which are external to the neurobiological system. (Clark 1997; Wilson 2000; Chemero and Silberstein 2008).

In the following chapters I will demonstrate that moral cognition is a prime instance of this beyond the brain model as it is partly constituted by both bodily and environmental resources. The question that will be addressed is “how wide is the locational scope of our explanation and what do we include in that explanation?” In the moral case, I believe that

there is a very compelling argument that *most* of our moral cognition is body based – we use our bodies as part of our moral cognitive system (moral cognition is embodied). There is also a strong case that *some* of our moral cognition is extended beyond our bodily boundaries (moral cognition is extended). Our explanations of moral cognition will therefore involve the integration of factors external to the brain and associated neural system. The understanding of moral cognition requires attention to body and environment structures, for they are as important to the system as brainbound internal structures.

3.4 Cognition is embodied and extended

The theses of *embodied* and *extended* cognition are in contrast to the brainbound individualism presented earlier. Central to these approaches is the idea that the cognitive system is not bound by the spatial limits of the brain and neural system, and the body and world play a more *intimate* role in the processes of cognition (Haugeland 1998b). Properties associated with the physical structures through which problem solving is implemented are coopted to undertake much of the problem solving itself. Embodiment focuses on the way that the body (and its interaction with the world) shapes and structures cognition and how the body can be said to be a constituent part of cognition. Extended cognition on the other hand focuses on the ways in which environmental resources (such as artefacts and representations) shape and structure cognition, and therefore how cognition is extended into the world.

Embodied and extended theories of cognition come under the broad theoretical and empirical research programme most commonly referred to as *Embodied Cognition*. The Embodied Cognition programme itself is young and diverse, and as yet there is no cohesive or unified position on what constitutes embodied cognition (Wilson 2002; Ziemke 2003). As such the terminology itself is often used in different ways to signify different things. Margaret Wilson (2002) for example has identified six different notions of “embodiment”, from those that place specific emphasis on interactions with the environment such as “cognition being situated in the world”, through to those that focus on the role of bodily experience such as “off-line cognition being body based”. What is clear though is that according to embodied cognitive science, it is not enough to ‘look

inside' solely at the mechanisms of the brain and associated neural system to fully understand what is going on in cognition. For the mechanisms of the brain are (often) deeply causally integrated with what is taking place in the body and environment, to the extent that they are an integral part of the problem-solving (i.e., cognitive) system. Adaptive success is achieved through the dynamic and reciprocal influences of the interactions between mind, body, and world, not (solely) the internal manipulations of brainbound processes. In Figure 3.1 we can see the changing understanding of the interactions between the nervous system, body and environment. Early understanding is clearly individualist, with the focus on the nervous system as the generator of adaptive behaviours and the sensory-motor systems acting as inputs and outputs. We move away from the individualist/internalist view towards one that involves more dynamic and reciprocal feedback from body and world. Finally, we see a fully coupled and integrated, dynamic brain-body-world system.

Figure 3.1 Changing understanding of the interactions between the nervous system, the body, and the environment

A) Focus on the nervous system as the generator of adaptive behaviours.

B) Sensory and motor inputs are processed and transformed by the body.

C) Nervous system and the body seen as complementary providing constraints and opportunities for action.

D) The functioning of the nervous system is affected by feedback from both motor inputs and environment.

E) Nervous system embedded in the body which is embedded in the world. The relations between each are closely coupled and dynamic. Adaptive behaviours emerge from the interaction of all three, mutually influencing components. (Chiel and Beer 1997:554, image as in original, text abbreviated from original).

Both embodied and extended cognitive research programmes take the physical context within which cognition occurs to be central to the structures and workings of cognitive systems. To be sure, the two programmes are not unrelated. For it is clearly the case that the ways in which our bodies interact with the world depends on the structure of that world, and the ways in which we utilise environmental resources depends heavily on the nature of our bodily engagement with the world. Brain, body, and world are intimately integrated and cognition emerges from this intimacy (Haugeland 1998b). In the next two chapters I will look at these issues in more detail and address the ways in which moral cognition is both embodied and extended, paying particular attention to how a) moral cognition is (often) directed and constrained by the nature of our bodies; our bodies structure cognition and do cognitive work, and b) moral cognition is (often) distributed throughout our physical and social environments: cognition is extended beyond our bodies.

We saw earlier in the chapter that traditional individualistic accounts of cognition emphasise the manipulations of symbolic structures that are context-independent, as well as action-independent (Clark 2004). In contrast, one of the central themes of the Embodied Cognition research programme is that (much of) cognition emerges from the physical interaction between agents and their environments such that “the way people represent and understand the world around them is directly linked to perception and action” (Pecher and Zwaan 2005:3). But this emphasis on perception and action does not mean we deny the existence of inner representational structures (although see Thelen and Smith 1994). Cognition is not just the appropriate perception-action links. As such, an embodied cognitive science need not reject a computational / representational theory of the mind (see for example Clark 1997, 2008; Wilson and Clark 2009). However, it does transform the nature of many of those representational structures and their role in adaptive processes (Smith and Semin 2004). In embodied cognition, the role and nature of many representations changes from being abstract and decontextualised to being action orientated and situated. Representational structures are seen as action determining (Clark 1997). In support of this idea is the growing body of evidence from psychology and cognitive science detailing the ways in which sensory-motor systems are involved in mental representations (Barsalou 2008; Pecher and Zwaan 2005). Much of this literature focuses on the ways in which cognition is partly (or wholly) constituted by sensory-motor states and in which active bodily engagement with the world structures and alters our

cognitive processes. Cognition is action orientated. We will see in the following chapter how moral cognition is embodied in this way.

With this in mind, it is worth highlighting more generally how moral cognition is specifically action orientated:

- *Morality is normative*: morality guides actions. It guides one's own behaviours as well as guiding others' behaviours. Morality in this way prescribes *action*.
- *Morality is motivational*: It is not merely normative, but has a particular motivational oomph. There is an intimate (be that either necessary or contingent) connection between morality and motivation. Morality motivates us to perform *actions*.
- *The subject matter of morality is action*: Moral judgements are made typically about the actions of agents or their intention for action. Moral violations are typically actions of some kind or another.
- *We learn morality through active engagement in the world*: We commit moral violations and are punished for them, we witness others being punished, and we learn to punish others and ourselves. Much of moral learning is through *action*.

Action is clearly central to moral cognition, which provides reason to believe that it is a fitting subject for study under the embodied cognitive research paradigm.

3.5 Embodiment, extension and evolution

It is worth emphasising why this is important with respect to our larger explanatory project, the evolution of moral cognition. If moral cognition is an evolutionary adaptation then we need to establish where the bounds of moral cognitive systems lie and the role of the components within the system in order to establish the targets of selection. The evolutionary psychologist has specified that the central determinants of cognitive phenotypes are computational processes residing specifically in the head. Selection is for internal computation mechanisms (i.e., Darwinian algorithms) that give

rise to individual moralities and group moral systems. Phenotypes are brainbound psychological processes (e.g., the moral faculty). But, if moral cognition is not individualistic, then we have to understand how wide the system is to know what structures natural selection is acting upon. As Dawkins makes clear in his book, *The Extended Phenotype*, the spider's web is in a "very real sense ... a temporary functional extension of her body" (1982:198), and should rightly be considered part of the phenotypic expression of the spider's genes. In the same way, if cognition is the target of selection and is both bodily and environmentally extended, then these mechanisms of extension are part of one's cognitive phenotype. They too will be targets of selection. Furthermore, understanding the relationship between internal and external resources is important for the relationship between cognition and culture. If the relationship is bidirectional and mutually influencing then the aggregative view of culture from the individualist's perspective is mistaken (Wilson 2000). I will address these issues in the following chapters.

3.6 Conclusion

In this chapter we have looked at how traditional approaches to cognition are typically individualistic in nature. The structure of the mind is internal and typically seen as independent of the external non-neural body and world. I also detailed how one very recent and prominent account of moral cognition and the research paradigm of evolutionary psychology are both inherently individualistic. In contrast to individualism, I examined the idea that resources external to the brain and organisms can become part of our wider cognitive systems. In this way cognition extends beyond the bounds of our brain and neural systems. In the following two chapters we will look at how, in the case of moral cognition, the body and world play problem solving roles, becoming constituent parts of our problem-solving apparatus, and hence part of our moral cognitive system.

3.7 Summary

The previous chapter provided evidence to suggest that moral cognition is not underwritten by an innate moral faculty. In this chapter we have looked both at traditional accounts of cognition (of which the nativist account in chapter 2 is an example) and an alternative account. The main points of interest are as follows:

Traditional accounts of cognition are individualistic, involving brainbound computations operating over rich, symbolic or sub-symbolic (amodal) representational structures.

- Individualism makes specific commitments to inner complexity and the roles of body and world in cognition.
- The Hauser-Mikhail model of moral cognition and evolutionary psychology are individualistic.
- Some recent accounts of cognition maintain that cognitive systems extend into both the body and the world.
- Embodied cognition claims that our body and its interaction with the world shape, constrain and are part of cognition.
- Extended cognition claims that environmental resources in the world are often constituent of our cognitive systems.

The following chapters will demonstrate how moral cognition fits into this alternative account of embodied and extended cognition, outlining concrete examples of both embodied and extended moral cognition. We will subsequently see, in chapter 6, how this has important consequences for an evolutionary account of non-nativist moral cognition.

4

The Embodiment of Moral Cognition

Embodied and extended views of cognition challenge many of the traditional notions of cognition, including its locational bounds and the structures over which cognitive processes operate. In this chapter I will focus on embodied cognition. After presenting examples of how cognition can be embodied, I will look at evidence for the embodiment of emotional processing and other processes central to moral cognition. We will see that our modality-specific neural systems and body structures are constituent parts of moral cognition.

4.1 The role of the body in cognition

We experience the world through our bodies and, importantly, also affect the world around us through our bodies. In this sense, the body mediates our brain's interactions with the world, and the world's with our brain. As uncontroversial as these interactions may seem, their role in cognition is not. We saw above how traditional accounts of cognition view the body as providing the inputs about the world to the computational process as well as acting as the effector for our cognitive systems to engage action in the world. Cognition is seen as the mechanism(s) that operate between perception and action (Clark 1997).

The embodied cognition view is that perception and action are central to, and very often

part of, cognition (Clark 1997, 1998, 1999; Wilson 2000; Wilson and Clark 2009; Barsalou 2008). Our bodily states play important and often essential roles in the operations of our cognitive systems, as much of that cognition is partly constituted by our sensory-motor systems, body structures and bodily action. There is a growing body of empirical evidence detailing how the body plays centrally important roles in cognitive systems, in domains as diverse as religious knowledge (Barsalou, Barbey et al. 2005), language comprehension (Zwaan, Stanfield et al. 2002), metaphor and folk metaphysics (Lakoff and Johnson 1980, 1999), theory of mind (Gallagher 2005; Gallese 2008), and social cognition (Semin and Smith 2002; Smith and Semin 2007).

The term embodiment is ambiguous and has been used in numerous different way within the embodiment literature (Wilson 2002; Ziemke 2003). I will focus on two prominent interpretations. The first is that body structures and states are constituent parts of our adaptive problem-solving routines. The second is that neural based modality-specific systems play important roles in cognition. (When referring to ‘embodiment’ I will be referring to these two types of embodiment.) Whilst these two understandings of embodiment are intimately related, their explanatory focus is different: the former emphasises the role that the body plays in processing information and transforming brain-body-world interactions, the latter on the modality-specific sensory-motor systems that mediate much of these interactions and how they are involved in both on-line (coupled) and off-line (decoupled) cognitive processes. In both cases, and what is central to an embodiment perspective, the *role of the body is paramount*. Most, if not all, of cognition is either partly or wholly constituted by the body’s physical characteristics, abilities, and its activities, as well as associated body-based sensory-motor and affective systems.

I will now look at some examples of embodiment with specific emphasis on each of these two types. With these examples in mind, I will then examine the ways in which the body is used in the processing of emotional information. This is centrally important to moral cognition because emotional information is a core element of moral cognition.

4.2 The embodiment of cognition

There is an increasingly strong case to believe that our bodily structures are deeply integrated parts of cognition. In contrast to the linear processing cycle (perception–cognition–action) of more traditional approaches to cognition, embodied explanations emphasise the ways in which output in turn affects inputs. This process is bidirectional and dynamic, operating as a form of reciprocal causation (Clark 1997). The role of the body and associated sensory-motor structures become central components in the cognitive system: “action generates perceptual feedback that, in turn, generates further action, so that outward behaviour becomes an important co-contributor to the process, including neural processes, that generate further behavioural response” (Barrett, Henzi et al. 2008:139). As such, the ways in which our bodies interact with the world and the various represented and non-represented properties of our bodies and our sensory-motor systems become explanatorily important in understanding cognitive processes.

4.2.1 The body is part of cognition

A useful example of the first type of embodiment is the case of auditory cognition, where the dynamic interplay between brain and body is seen as essential to performing the cognitive task. Furthermore, we also see the inclusion of non-representational components as part of cognition. Larry Shapiro explains:

Generally, larger distances between ears provide greater auditory acuity. But also important is the density of the matter between the ears because sounds of varying frequencies will behave differently when travelling through a given medium. The auditory system incorporates facts about ear distance and head density in its processing, but not in a way that requires their symbolic representation. There is no need to represent the distance between ears because it is the distance itself – not its representations – that creates the opportunity for greater auditory acuity. By analogy, a spring scale needn’t represent the compression of the spring in order to determine the weight of an object. The compression of the spring tells us directly the weight of the object. (2007:340-341)

Deeper analysis gives us an understanding of just how auditory cognising is embodied. The role of the auditory system is twofold: to help identify ‘what’ the acoustic stimulus is,

and to locate 'where' the stimulus came from in the external environment. 'Where' information processing involves the use of the torso, head and the physical structure of the ear (pinna) to filter sound signals (Recanzone 2002). Differences in the amplitudes between different frequencies are generated depending on where the sound is located in relation to the agent. These interaural differences between auditory inputs via the ears (and other bodily structures) provide essential cues in normal sound location processing. Many owl species, for example, have asymmetrically placed ears with one ear higher than the other to enable greater locational acuity for sounds above or below them. Sounds emanating from locations below eye level are louder in the left ear than the right, and the opposite is true for sound emanating from locations above eye level. At eye level the volume will be the same. This allows greater auditory locational ability in the darkness when visibility is very poor and even in cases of complete darkness (Volman and Konishi 1990; Knudsen and Konishi 1979), a clear example of body structures structuring cognitive processing.

Interaural differences, however, are not the only factors important for an explanation of auditory localisation, and any explanatory model of auditory cognition that focuses on interaural differences alone will be incomplete (Aytekin, Moss et al. 2008). Auditory locational cognition also involves incorporating the consequences of motor actions and sensory information coordinated with the sampling of auditory space. Voluntary motor changes have sensory consequences. If one moves in a particular way relative to the sound source, auditory input will change. Through voluntary sensory-motor exploration of the spatial and auditory environment, the brain and nervous system can learn the contingent relations between spatial and auditory sensory inputs (O'Regan and Noë 2001). Once these contingencies are acquired, they can be used in auditory localization. We can see the necessity of this type of cross-modal sensory-motor coordination when we think of the way in which bodily movements need to be coordinated with acoustic cues as one walks past a stereo speaker. Any model of auditory locational cognition that relies solely on acoustic cues alone will not be able to differentiate between an agent walking past the speaker from the speaker moving past the agent. In addition to processing interaural differences, navigating auditory space involves integrating and coordinating the movements of head and body, as well as feedback from other sensory modalities. (For a computational model of this type of sensory-motor integration in the learning of sound localisation, see Aytekin and Moss et al. (2008).)

We see evidence that auditory locational cognising involves these kind of cross-modal feedback loops from the fact that audition is subject to cross-modal illusions. One of the most famous is the 'ventriloquist effect', whereby one experiences the illusory location of a sound when there is an apparent visible source of that sound, as in the case of the ventriloquist's puppet (Casey 2008). We navigate much of auditory space by integrating cross-modal information sources. Under normal circumstances they will synchronise with one another (because they typically have the same source), enabling successful location of the sound source. But when the input relations between the sound and vision are inconsistent, we experience an illusion. After a period of exposure, agents will actually recalibrate their integration of the sensory inputs to adopt the new pattern between the different modalities (De Gelder and Bertelson 2003). Information in one modality affects our perceptions in another. Cross-modal integration of this sort involves brain-body-world feedback cycles constantly sampling auditory and visual space and adjusting our perceptions in those modalities accordingly.

The properties of our bodily structures and associated actions are very much a part of the auditory cognitive system and are necessary for successful navigation of the auditory domain. Bodily structure and movement constrain auditory cognition as well as provide affordances (opportunities for action, i.e., sampling auditory space). This ability arises from the entire coordinated system, not from any one specific or isolatable component. As such, a complete explanation of the auditory cognitive system necessarily involves an account of between-ear distances, head densities, and the like, as well as the coordinated effects of head and body movements and the role of other sensory modalities. Auditory spatial knowledge and auditory information processing is therefore grounded in our bodily interactions with the world: auditory cognition is embodied.

One important point to be taken from the Shapiro passage cited earlier is the way in which non-representational components are part of the explanation of the wider cognitive system – it is not the case that the physical structures (head density, pinna structure etc) are necessarily represented internally. They are non-representational components of the cognitive system that play important information processing roles. Importantly, they reduce the internal computational load placed on a brainbound neurological system by changing the nature of the brain-body-environment interface and

therefore the routines required to sample acoustic space. This is not to deny that computational processing occurs. What is important, in considering them part of the cognitive system, is that these non-representational, non-represented resources are appropriately situated in what Wilson and Clark describe as an “ongoing web of computational activity” (2009). In the case of auditory cognition, those body structures are suitably located within the dynamic information processing feedback loops between brain, body and world. This kind of adaptive problem-solving is not a case of ‘input → cognition → output’, but one of integrated, dynamic, reciprocally influencing feedback cycles between brain, body, and world, involving representational and non-representational structures, across multiple modalities. Bodily structures are a central and integrated part of our problem-solving routines to the extent that those bodily structures are part of the cognitive system.

The second interpretation of embodiment focuses on the ways in which modality-specific neural systems are a central part of the representational architecture of much (some claim all) of our cognitive processing. As we will see, this is relevant to embodied moral cognition. We saw earlier how traditional accounts view higher-cognition as algorithmic processes operating on amodal representational structures. In contrast to this is the view that many representational structures involved in these processes are modal-specific.¹ Embodied cognition claims that our conceptual systems are not functionally separate from our sensory-motor systems.

Some recent accounts of embodied cognition claim that when we experience situations, the sensory-motor states elicited by that situation are captured in modality-specific representational systems (for an overview see Barsalou 2008). When we experience a cat for example, the (online) sensory-motor experience is captured and stored in neural regions as patterns across multiple modalities (such as the visual, olfactory, auditory, proprioceptive, and affective). Later, conceptual processing involves the partial re-enactment of those states in the representation of categories. Subsequent (offline) thinking about cats involves the partial reactivation of those patterns, “in a sense partially “reliving” experience in its sensory, motor, and introspective modalities” (Winkielman, Niedenthal et al. 2008:265). This is what Barsalou and colleagues call ‘simulations’

1. See for example, Zwan (1999), Prinz (2002), Barsalou (2003), Pecher (2005), Gallagher (2005), Niedenthal (2007) and Gallese (2008).

(Barsalou 2008; Niedenthal, Barsalou et al. 2005b), or ‘situated models’ (Zwaan 1999). Simulations, I will claim below, play an important role in moral cognition and moral learning.

Neurological and behavioural studies have provided increasing evidential support for this view. Evidence from lesion studies shows that selective damage to particular sensory-motor regions of the brain has adverse effects in individual categories. For example, damage to the neural regions involved in spatial processing afflicts locational knowledge abilities. Damage to motor areas often produces corresponding deficits in knowledge of manipulable objects such as tools, because manipulable object categories rely on motor areas of the brain for processing information. Damage to the somatosensory cortex corresponds with difficulties in facial expression recognition, as we use those brain regions to simulate facial expressions in others. Damage to somatosensory regions also corresponds to an inability to identify others’ emotional states. Adolphs and Damasio conclude that the “findings are consistent with the idea that we recognize another individual’s emotional state by internally generating somatosensory representations that simulate how the other individual would feel when displaying a certain facial expression” (2000:1). The somatosensory and sensory-motor brain regions are heavily implicated in retrieval of category-specific information, strongly suggesting that modality specific systems are integral aspects of conceptual processing. (For a review of the above findings and discussion see Barsalou 2008).

In conjunction with the neurological evidence, there is a growing body of behavioural evidence implicating modality specific systems in our representations of category membership. When subjects are asked to verify the properties of concepts, for example, a switching cost is observed if those properties are from different modalities. Peecher et al. (2003) studied property verification times for various conceptual categories. They found that when asked to verify a conceptual property in one modality (such as *loud* being a property of BLENDER) subjects were quicker in making that identification if, in the preceding property verification task, they had been asked to identify a conceptual property belonging to the same modality (such as *rustling* being a property of LEAVES). They exhibited slower identification times when the preceding conceptual property belonged to another modality (such as *tart* being a property of CRANBERRIES). In other words, there is a cost to switching between different modality specific properties of

concepts. If different modalities were not involved in conceptual representation (i.e., if concepts were represented amodally) then one would not expect to observe any switching costs (Pecher, Zeelenberg et al. 2003). This is evidence for the claim that modality specific systems are involved in category knowledge (for review see Barsalou 2003; 2008).

Brain scanning techniques such as fMRI support these conclusions, showing that regions of the brain involved in perceptual experience of, and physical interactions with, a particular category become active during subsequent processing of category information independent of actual experience of the category. Thinking about food activates gustatory regions of the brain that are activated when actually experiencing food; thinking about tools and sports activates motor centres of the brain; thinking about emotions activates limbic regions associated with emotional experience. This is suggestive evidence that areas of the brain used in perception and action are also used in the representation of conceptual knowledge in off-line (decoupled) processing (Zwaan 1999; Barsalou 2008). Category representations are (at least partially) encoded in modality specific neural regions. Importantly, this means that high-level cognition is not divorced from those systems involved in our active engagement with the world: “The way people represent and understand the world is directly linked to perception and action” (Pecher and Zwaan 2005:3).

This kind of explanation is important for the purposes of understanding moral cognition. It suggests that when we engage in off-line moral thinking we simulate situations, thereby engaging sensory-motor systems. When we think about moral dilemmas such as trolley bus problems, spare part surgeons, or the moral betrayal of a friend, we are partially reactivating those regions of the brain that we would employ *as if* we were actually there, in a sense, vicariously simulating the experience. Of course, simulation does not mean that we necessarily consciously visualise or re-create the experience (although in most moral dilemmas I suggest we do), any more than simulation theory in folk psychology necessitates that we consciously recreate other people’s first person experiences. There is a difference between simulation and imagination, but in both cases we employ cognitive resources that we normally use on-line, in off-line cognitive tasks.

There is some debate as to the extent of embodiment and whether embodied conceptual systems can characterise *all* conceptual knowledge and conceptual processing (see for example Machery 2007, 2006). However, this need not concern us here. I claim only that much of our moral cognition is embodied. Some of our more abstract thought processes may not involve sensory-motor systems.²

In the following pages we will see that embodiment has some important consequences for moral cognition. But first I will focus on emotional embodiment. There are two reasons for this. Firstly, emotions are an excellent study of embodiment. Our emotional systems (typically) involve bodily effects (see Box 4.1). The perception and use of emotional information involves body states and associated neural structures. There is a rapidly growing body of empirical work detailing the role of body-based effects in emotional processing, but comparatively little specifically addressing the various roles of the body in moral cognising. There is, however, research detailing the effects of emotions on moral cognising. This brings us to the second reason: emotions are central to much (if not all) of moral cognition. In chapter 2 we saw that emotions were a central tool in early moral learning. There is also strong support for the view that emotions are heavily involved in many aspects of moral cognition.³ If processing emotional information involves body states, and if moral cognition involves processing emotional information, then we have good reason to think that moral cognition also involves the use of body states and associated neural mechanisms.⁴ The embodiment of emotion therefore has particular relevance to embodied moral cognition.

2. See Barsalou (2005) for ways in which abstract concepts (such as mathematical concepts) might be constituted by modality-specific systems.

3. See for example: Greene et al. (2001; 2002), Haidt et al. (2001, 2003; 2005), Nichols (2002a, 2004), Joyce (2006), Prinz (2006, 2007) and Moll et al. (2008)

4. One recent example of a body-based (embodied) account of moral cognition specifically focusing on the role of emotions is that from Jesse Prinz (2005, 2006; 2007). Prinz has detailed an account of how moral concepts such as “good” and “bad” are grounded in bodily based emotional states. Prinz argues that emotions are perceptions of ones own bodily changes in relation to our perceptions of the world. Moral concepts are dispositions to experience differing emotions in different situations. What unifies moral concepts across various situations are the emotions that they elicit in us.

Box 4.1 Theories of emotion and embodiment

There is by no means a consensus on what exactly emotions are. There are numerous theories and much of the debate has centred on whether emotions are “cognitive-appraisals”, “body-based”, or some hybrid of the two. Cognitive-appraisal accounts hold that what is central to emotions is that they involve particular types of cognitions, namely propositional attitudes (de Sousa 2008). While various cognitive-appraisal theories differ as to whether somatic components are necessary parts of emotions, they share the view that cognitive aspects of emotions are distinguishable from somatic or bodily aspects – what Prinz (2004) calls the *disembodiment hypothesis*. Emotions are disembodied and our bodily reactions typically occur after the cognitive-appraisal emotional system has been engaged. When we judge that a situation is fearful, the cognitive judgement itself is central to the emotion, and any bodily reactions contingently follow from that appraisal.

In contrast, body-based accounts emphasise the role of body states as constituent parts of emotions. Emotions are perceptions of changes in our body states (see for example Damasio 1994 (2005)). Hybrid theories of emotions bring together various aspects of the two theories (see for example Morris (2002), Thagard (2005; 2008) and Prinz (2004)).

On the face of it, body-based and hybrid accounts of emotions appear to be more amenable to embodied accounts of cognition because they view the body as being (partly) constitutive of emotions. If this is true, and if emotions are part of our cognitive system, then the body is also part of our cognitive system. Cognitive-appraisal theories of emotion on the other hand do not obviously lend themselves to embodied theories of cognition. Yet we can make the claim that cognitive-appraisal theories of emotions are consistent with embodied accounts of cognition. The connection between cognitive-appraisal and bodily effects may be a contingent one, but the fact is that when we experience emotion we *typically* exhibit bodily effects, such as increased heart rate and heightened sensory awareness. In other words, emotions (typically) have bodily effects whether or not those effects are in some way or other constituent of the emotion (as in body-based accounts) or derivative effects of the emotions (as in cognitive-appraisal accounts). From an embodiment perspective, what is important is the role these body-based effects play in the brain, body, world feedback loops. So, even from the cognitive-

appraisal account of emotions where body states are not necessarily constituent of emotion, what is important is that those body states are involved in the brain-body feedback loops. The bodily effects of emotions play an essential element in the normal functioning of emotions in organising our lives. Those bodily effects, whether part of the emotion proper or not, are part of our wider cognitive systems. We look for this evidence in the influences the body has in both the processing of emotional information and in our wider cognitive systems (Niedenthal, Barsalou et al. 2005b).

4.3 Emotional embodiment

There are two aspects of emotional embodiment that I would like to look at. The first is evidence that emotional processing (i.e., the perception and use of emotional information) involves the use of the body (see Niedenthal, Barsalou et al. 2005b). Here we see that body states and associated modality-specific neural systems are intimately related to the perception and use of emotional information. The second is the way in which those body-based effects are integrated into our wider cognitive, and in particular moral, systems. If cognition involves reciprocal feedback from the body, then those bodily manifestations associated with emotional processing will have cognitively significant causal effects. Note that it is not enough that they have causal effects on one's cognitive system, but they have the *right kind of causal effects*. Seeing my cat may cause me to think of cats in general, or to consider the last time I fed him, but that does not mean that my cat is part of my cognitive system. The right kind of causal effects are those in which the processing of information is deeply integrated into our problem solving routines. Emotions elicited both in oneself and our sensitivity to the emotional displays of others directly change the way in which we perceive, interact with, and ultimately understand many of the adaptive problems that confront us.

4.3.1 Emotional processing involves body states

We are all familiar with many of the bodily accompaniments of emotional activity. When we experience anger we may clench our fists, raise our voice, and our face becomes flushed with blood. We may even be compelled to exact violence. When we experience disgust, we will often react by screwing up our face and turning away from the offending

medium; we may sometimes experience an involuntary gag reflex and on occasion vomit. It would appear that emotional episodes often involve body states, although it is not always the case that bodily manifestations of emotional episodes are themselves externally observable. Sometimes we manage to suppress or inhibit those bodily expressions, such as in the case of culturally specific "display rules" which can override the pattern of emotional expression. For example, when Japanese subjects were shown unpleasant film footage they were found to mask their previously displayed facial expressions of negative emotions when in the presence of authority figures, whereas American subjects did not mask their expressions (Ekman, Liebert et al. 1972). In the case of self referential emotional episodes, such as guilt and shame we may not have observable bodily manifestations at all.⁵ But it is certainly the case that there is something quite distinctive about what it *feels* like to have guilt or shame. Believing that you have done something wrong will, for normal people, elicit states quite distinctive from having say, the belief that 'a square is a polygon that has four equal sides'. Much of this distinction is body-based. Emotional episodes are associated with typical body effects, even if those effects are non-observable. The question is whether those bodily effects are part of the processing of emotional information.

Body states are resources with which we can understand others. Witnessing someone stub their toe elicits strong motor responses in the observer, to the effect that it is near impossible not to wince or react in some bodily way. This 'vicarious experiencing' suggests that much of our representational knowledge is stored in our modality-specific systems, such as our sensory-motor systems. Our body-based systems also enable the interpretation of emotional states of others. When we experience emotional events ourselves, our "sensory-motor and affective states that are triggered during the encounter with an emotion-eliciting stimulus... are captured and stored in modality-specific association areas" (Winkielman, Niedenthal et al. 2008:265). Those modality-specific association areas are reactivated on encountering similar states in ourselves or others. Witnessing someone stubbing their toe activates those sensory-motor systems and

5. Although Robert Frank (1988), for one, thinks that there are externally observable bodily manifestations of guilt: diverting ones eyes from another's gaze and blushing (pp.126-127). Frank thinks that these are signalling mechanisms, they signal that one is moral because moral people feel guilt, whereas immoral and amoral people don't. Therefore it is better to deal with someone who shows they have the capacity to feel guilt (and hence more likely a moral person) than with someone who doesn't: "being known to have the capacity to feel guilt would make someone an attractive partner for ventures that require trust." (1988:132). For an argument against the claim that guilt has observable bodily expressions see Prinz (2004). (See also Keltner 1996.)

thereby the toe stubbing knowledge stored in those association areas. Thinking about emotions and interpreting others emotional episodes similarly involves the partial reactivation of those modality-specific, sensory-motor systems (Niedenthal, Barsalou et al. 2005a; Winkielman, Niedenthal et al. 2008). In this way our own experiences of emotional stimuli help to construct the relevant knowledge structures used in the future understanding of the emotional states of other agents.

There is evidence for a close link between others' body states and our understanding of their emotional states (Niedenthal, Barsalou et al. 2005a; Niedenthal, Barsalou et al. 2005b; Niedenthal 2007; Winkielman, Niedenthal et al. 2008). Charles Darwin's (1872) famous cross-cultural research on the facial expressions of emotions provides evidence for the connection between bodily states and emotional processing. Darwin collated the results of questionnaires sent to Englishmen located throughout the world, asking them to report on the emotional expressions of the indigenous people. From the results, he concluded that "the same states of mind is expressed throughout the world with remarkable uniformity ..." (Darwin 1872; as reported in Ekman 1999:302, see also Ekman for discussion). People, according to Darwin's results, can reliably identify emotions from facial (bodily) displays. Although the scientific methodology of Darwin's experiment was questionable, the results strongly suggest that we use the bodily states of others to process emotional information. In more recent times and with more scientific rigour, Paul Ekman and many others have conducted numerous experiments further investigating the relationships between emotions and facial expression. In what has become a rather famous experiment, Ekman and colleagues presented the preliterate Fore people from Papua New Guinea with emotion-inducing vignettes. He then presented them with a series of pictures displaying facial expressions associated with different emotions. The subjects consistently picked the same facial expressions corresponding to emotional states in the vignettes as those picked by Western subjects (Ekman and Friesen 1971; Ekman 1999; Keltner and Ekman 2000). Ekman also asked the Fore people to display how their own faces would look should what happened in the vignettes also happen to them. He filmed their facial expressions and showed them to American university students who correctly identified their facial expression with the emotion that was intended (Ekman and Friesen 1971). Here we have *prima facie* evidence that we use perception of the body states of others to process emotional information about them.

But visually recognising someone's emotional state from their body state is not enough to establish that one's *own* body states are used in processing emotional information. All it tells us is that we have knowledge of what facial expressions (body states) match which emotion. Further research does however show our own body states are used to interpret the emotional states of others. When we witness the body states of others, we often unconsciously mimic those states and in doing so vicariously experience those same states ourselves. (Often referred to as the Chameleon Effect (Chartrand and Bargh 1999)). Seeing happy people smiling, for example, induces positive feelings, whereas seeing people frowning induces negative feelings. In this way, mimicking others' facial expressions plays a role in the processing of emotional stimuli and our responses to them. When asked to identify the emotional expressions displayed in photographs of other people, it has been found that subjects not only mimic the facial expressions that they are asked to identify, but that the more they mimic them the more accurate they were at identifying the expression (Wallbott 1991; Niedenthal, Barsalou et al. 2005a). Furthermore, when mimicking is inhibited, rapid identification of emotional states is often diminished (Stel and Knippenberg 2008).⁶ This strongly suggests that physical body states (in this case, facial muscles) play a role in the identification of the emotional states of others. Like in the case of hearing, it is the smile-routine, not its amodal representation, that plays a role. It is not a straightforward 'perception → cognition → action' sequence involving computations over amodal representations, but is best seen as involving a brain-body feedback loop: when we witness facial expressions we automatically and unconsciously mimic those facial expressions, and in doing so we get information feedback from our motor and somatosensory systems about the other person's emotional state. Our bodies are involved in processing information in the perception and understanding of others' emotional states. (Niedenthal, Barsalou et al. 2005a; Niedenthal, Barsalou et al. 2005b; Niedenthal 2007; Winkielman, Niedenthal et al. 2008).

Further evidence from neuroscience that the inhibition of somatosensory structures influences the perception and understanding of emotional stimuli supports this view (Winkielman, Niedenthal et al. 2008). Research shows that damage to the somatosensory

6. Interestingly this effect is noticed in women and not men. The authors suggest this is because women have been shown to be more facially expressive than men and therefore facial feedback plays a more significant role in emotional information processing in women (Stel and Knippenberg 2008).

systems produces deficits in the processing of emotional information (Damasio, Grabowski et al. 2000; Adolphs 2002; Niedenthal, Barsalou et al. 2005b). Patients who had suffered lesions to the somatosensory-related cortices displayed impaired abilities in identifying typical facial expressions of six basic emotions: happiness, surprise, fear, anger, disgust, and sadness (Adolphs, Damasio et al. 2000; Adolphs 2002). From their studies Ralph Adolphs, Hannah Damasio and colleagues conclude that “recognizing emotion in another person engages somatosensory representations that may simulate how one would feel if making the facial expression shown in the stimulus” (2000:2683). Emotional knowledge is (in part) stored via somatosensory representational structures, i.e., modal representations as opposed to amodal representations. It is the reactivation of these structures that delivers the requisite knowledge to interpret the emotional states of others. As was described earlier, during perception, action and introspection the brain captures modality specific states. When required the brain recalls that knowledge through the partial reactivation of those states (see Niedenthal 2007).

The evidence outlined above indicates that both actual body states and somatosensory and motor states are used in the processing of emotional information – the cognitive system involves feedback from these states in our interactions with the social world.⁷ (For more evidence and discussion see Niedenthal, Barsalou et al. 2005a; Niedenthal, Barsalou et al. 2005b; Niedenthal 2007; Winkielman, Niedenthal et al. 2008). The many ways in which bodily structures and modality-specific systems are part of cognition undermines the traditional account of cognition presented earlier. I will now look at the claim that emotional processing is integrated into cognitive systems, notably moral cognition.

4.3.2 Emotional processing is integrated into our wider cognitive systems

The perception and understanding of emotional stimuli involves active engagement of bodily and modal-specific resources that play an important role in providing information about the stimulus. Furthermore, embodied emotional processing is integrated into the adaptive problem-solving routines involved in the navigation of our physical and social

7. Whether or not amodal representations are involved is an open question, but this evidence strongly indicates that modal specific structures are used in the processing of emotional information.

environments. These body-based mechanisms do this by altering our perceptions of the world, providing information about stimulus and motivating action.

Antonio Damasio presents compelling evidence that emotions play critical roles in our judgements and decision-making processes (Damasio 1994 (2005); see also Ramachandran and Blakeslee 1998). When particular emotional capacities are absent or inhibited, decision-making abilities are correspondingly compromised. Damasio's famous case study is that of Phineas Gage, a railroad worker who suffered a traumatic brain injury to his frontal lobes. Gage was transformed from what was, by all accounts, a normal, likeable individual to one with "nefarious personality traits ... he no longer showed respect for social convention; ethics in the broad sense of the term, were violated; the decisions he made did not take into account his best interest ... he was given to invent tales ... [and there] was no evidence of concern about his future, no sign of forethought" (Damasio 1994 (2005):11). Damage to the language and motor areas of the brain were spared in the accident, allowing Gage to maintain many other aspects of normal neuropsychological function. The damage was mainly confined to the ventromedial prefrontal cortices, which are involved in emotional processing and control. Emotions are required, concluded Damasio and colleagues, for normal reasoning, rational decision-making and behaviour (Damasio 1994 (2005); Damasio, Tranel et al. 1990). "The action of biological drives, body states, and emotions may be an indispensable foundation for rationality.... Rationality is probably shaped and modulated by body signals, even as it performs the most sublime distinctions and acts accordingly" (Damasio 1994 (2005):200). Here we have evidence that emotions are deeply integrated cognitive resources providing adaptive problem-solving opportunities that are not clearly available through other, non-emotionally driven, means.

We can see more evidence of this from the situational manipulation of emotions in normal patients and how they influence and alter the judgements we make about the world. Lerner and Keltner (2001) detailed some of the ways in which emotions can alter evaluative judgements. Looking at the impact fear and anger have on risk perception, they found that they have the opposite effect on some of the evaluations we make. Experimentally induced fear caused subjects to make pessimistic judgements and choices with respect to risk perception, appraisals of certainty, and appraisals of control. In the case of experimentally induced anger, subjects made relatively optimistic judgements and

choices. Evidence that specific emotions shape our judgements and the choices we make, again suggests that emotions are integrated into our cognitive processes. To reach this conclusion, however, we need further evidence that the bodily aspects of emotions are causally efficacious in our cognitions, rather than mere action-inducing outputs (cf. the ‘post-moral faculty’ role of emotions in the Hauser-Mikhail model of moral cognition). In other words, it is not enough to show that emotions matter; they must matter *because* they are embodied.

One way that bodily aspects do matter is by imparting an “affective tonality” (Gallagher 2005:201) to situations or stimuli.⁸ In this way emotions impart a particular ‘what it is likeness’ to experience.⁹ Emotional experience has a phenomenology, and this phenomenology plays a causal role in our problem-solving routines by changing our experience of the world and thereby our actions within the world. The absence or disruption of this tonality changes the way in which we think about, and act within, the world. Sufferers of Capgras’ syndrome, for example, display a reduced emotional response to a specific category, faces, while still maintaining a normal emotional life in most other respects. For the majority of Capgras sufferers, the syndrome manifests in regarding those who are close to them as impostors. Neuroscientist V.M. Ramachandran describes the response of a patient with Capgras’ syndrome when asked about the patient’s father: “No, no, doctor. That guy isn’t my father. He just looks like him. He’s—what do you call it?—an impostor, I guess. But I don’t think he means any harm” (1998:159). One way to interpret these results is that, as Ramachandran argues, Capgras’ syndrome delusions are caused by a disconnection between the areas of the brain associated with facial recognition and emotional processing. Capgras patients fail to experience “the “glow” or sense of familiarity associated with a known face and set of memories”, even though their visual recognition system is in tact. Sufferers show lower sympathetic nervous system activation (as shown by the Skin Conductance Response test (Ellis, Young et al. 1997)) in these familiarity tests – their body-based systems do not respond to the stimulus. However, Capgras patients can make normal emotional

8. Similarly, Damasio’s somatic marker hypothesis maintains that bodily feelings are connected representations of situations or stimuli (1994 (2005)).

9. There is some debate as to whether or not one can have unconscious emotions. I am only arguing though, that the phenomenology of emotions is *one of the many* ways in which emotions alter our cognitions. It may be the case that some emotions are indeed unconscious, but I think it is fair to claim that ‘most’ of our emotional reactions typically have a phenomenology and that this phenomenology influences the way we perceive and interact with the world – especially in the moral case.

categorisations in non-visual domains. They do, for example, get the emotional “glow” and correctly identify familiar people on the telephone. Sufferers identify the impostor as looking identical to the person they know; they recognise the person. But they do not *feel* familiar, so they reason that the person must be an impostor (Ramachandran and Blakeslee 1998:170). It appears that Capgras sufferers fail to make the connection between visual stimulus and bodily response.

Cases such as this indicate that when we do not have access to the correct body-based feedback there is a change in our problem-solving abilities. Here we see bodily states and somatosensory regions providing information about stimuli or situations. Feeling fear, for example, typically makes people judge events as being more “uncertain”. Feeling sadness at the death of a friend provides us with information about the value of the relationship between oneself and the friend and subsequent loss. Feeling sorrow at someone crying informs us of the emotional state of another. Jesse Prinz furthers this argument by claiming that the primary role of emotions is to carry information about relations to one’s well-being, or “core relational themes”. Emotions in this sense provide us with information about our relations with the world (Prinz 2004). Viewing emotions through this lens, we can see that in the case of the Capgras sufferers, the disconnect with embodied responses results in a failure to provide important information about the relationship between the sufferer and those close to him. The body-based ‘feel’ of emotions provide us with information about our relationships with the world, information that is central to our wider cognitive processing required for successful navigation of our social world (Damasio 1994 (2005)). Body-based aspects of emotional processing are integrated into our adaptive problem-solving routines.

So far we have seen that processing emotional information involves body states and modality-specific systems. We can also expand this to moral cognition. If the processing of emotional information is used in moral cognising, we have reason to believe that body states play some role in moral cognition. I will now outline some evidence that emotional states and the processing of emotional information is used in moral cognitions. Changes in motor and somatosensory resources alter our moral perception, judgements, and behaviours. As with the embodiment of emotional processing above, behavioural, neuroimaging and lesion studies indicate that bodily states and sensory-motor systems are (partly) constitutive of moral cognition.

Wheatley and Haidt (2005), for example, show evidence that inducing emotion alters the moral judgements that people make. Research participants were hypnotised to experience feelings of disgust in response to particular emotionally neutral words such as “often” and “take”. After being presented with fictional stories (both moral and non-moral) which contained the emotionally associated words, the participants were asked to rate the behaviours of individuals in the stories along the lines of *moral wrongness* and *disgust*. The results showed that participants who had been hypnotised to feel negative emotional response with respect to the particular words were more likely to rank events as “disgusting” or “morally wrong” than non-hypnotised subjects. These results show that manipulations of people’s affective states can alter the severity of the moral judgements people make. Similar evidence is also found in studies that do not involve hypnosis. In these cases, people who are induced to feel disgust (such as sitting in a dirty room) will subsequently judge morally ambiguous actions as immoral ones. (S. Schnall et al., unpublished, reported in Clore and Huntsinger 2007:394). (For more examples and discussion see Schnall and Haidt et al. (2008), Levin (1975), also discussion in Joyce (2006) and Prinz (2006).)

This is evidence that disgust responses (and hence embodiment effects) are involved in processing moral information (disgust is one of the facial patterns that people mimic). But these kinds of results are not confined to the emotion of disgust. Recent work by psychologists Piercarlo Valdesolo and David DeSteno (2006), involved inducing positive affect in participants by showing them a five minute comedy clip. A control group was shown an affectively neutral five minute documentary about a small Spanish village. The participants were then asked to judge the appropriateness of the protagonist’s actions in both the footbridge (personal) and trolley bus (impersonal) moral dilemmas.¹⁰ Subjects who had viewed the positive affect film clip were more likely to judge pushing an agent off a footbridge (and thereby killing him) to save the lives of five others to be morally permissible than the control group who had watched the neutral film clip. There were no differences in performance between the two groups in the less emotional impersonal dilemma. The authors conclude that these “findings demonstrate that the causal efficacy of emotion in guiding moral judgement does not reside solely in responses evoked by the

10. For the original footbridge and trolley dilemma thought experiments upon which Valdesolo and DeSteno’s experiments are based, see Judith Thomson (1986).

considered dilemma, but also resides in the affective characteristics of the environment” (2006:477).

Affective/emotional states causally affect the moral judgements that we make. As in the case studies on emotional cognition above, the causal pathway is from *induced affective state* to *judgement*, not the other way round. In each case, the emotional inducement affected judgment as well as behaviour, strongly suggesting that emotional processing is involved in the moral cognitive process itself, not solely the behaviour motivating output of the process. There is also a growing body of evidence from neuroscience detailing the contribution of emotion to moral cognition (Greene, Sommerville et al. 2001; Moll, Zahn et al. 2005; Moll, De Oliveira-Souza et al. 2008). Lesion studies support this view. As in the case of Phineas Gage, modern day patients who have selective deficits in the ventromedial cortex (associated with emotional processing) show deficits in particular types of moral cognition, such as an inability to draw the same moral judgements on emotionally salient moral dilemmas as normal patients do (Ciaramelli, Muccioli et al. 2007) (See section 4.5 below).

Recalling the information-bearing role that body states of affect/emotion play in brain-body feedback loops, we can now see how affect changes the way in which we perceive moral stimulus by altering the information that we process about that stimulus (as we saw earlier). This occurs even when the affective state is elicited independently of the moral stimulus. Recall also the ventriloquist effect observed in auditory locational processing, whereby we falsely attribute the sound source because of information from the visual perception. It is quite possible that something similar is happening in the cases of induced disgust above, whereby we attribute the source of affective inducement to a mistaken stimulus.

What is clear is that moral cognition often involves body-based emotional processing and when that processing is altered, disrupted or absent, we see corresponding alterations in moral cognition. The processing of emotional information involves the use of body states. Emotions elicited in ourselves and others are intimately linked to moral cognition – our motivations, the judgements we make, the types of morally relevant information we attend to; moral cognising involves the processing of emotional information about ourselves and others. We can conclude therefore, that moral cognition is embodied. In

addition to emotional processing there are many other processes involved in moral cognising which are also embodied. I will detail some of those processes now.

4.4 Embodying more than emotions

Body-based effects play a role in non-emotion related cognition. Apparently innocuous body movements can affect the evaluations we make. When people hold a pencil in between their teeth, thereby forcing them to exhibit a facial pattern similar to that of smiling, they report finding cartoons funnier than those whose smile is blocked by holding a pencil between their lips (Niedenthal 2007). The furrowing of the brow brings about feelings of mental effort. When subjects shake their head they show a higher recognition of negative words. Conversely, when subjects nod their heads they show higher recognition of positive words (Wells and Petty 1980). Something as simple as the type of arm movement that people make when they are shown neutral images (such as Chinese ideograms) effects the evaluations that they make about those images (Cacioppo, Priester et al. 1993; Priester, Cacioppo et al. 1996). (See also discussion in Smith 2008) Often subtle body movements alter our cognising about the world. Even if these effects are small, as we will see in chapter 6.5, they can also have important bearing on moral evolution.

There is also a wealth of evidence from studies on priming effects. The incidental activation of knowledge structures show how many of those structures are based in our sensory-motor systems. In one experiment by Bargh, Chen, and Burrows (1996) subjects were presented with groups of randomly ordered words and were asked to construct sentences out of them. Some of the groups contained words that were associated with elderly stereotypes (such as “wrinkled”, “retired”, “lonely”, “grey”, and “forgetful”) but excluded references to slowness. Participants who were subject to the elderly priming condition exhibited a slower walking speed after the experiment ‘ended’ than those who were subject to the neutral condition. Unconscious activation of the elderly stereotype produced embodiment effects associated with that stereotype, showing the embodiment effects of social information processing, even in cases where the social stimuli were absent. Results like this are very suggestive that the activation of knowledge structures involves the activation of motor regions associated with particular knowledge – knowledge structures for particular categories are grounded in motor systems and are not

purely amodal representational structures.

This is relevant from a moral point of view because many of these embodiment effects are important for understanding others. From a very early age we are attuned to the bodily actions of other agents. Bodily engagement, such as intention-signalling via hands, eyes, and gaze, as well as processes such as emotional contagion, provide a ready source of inter-subjective understanding. Shaun Gallagher presents evidence that we undertake perception-based understanding of others, in what he calls “body reading” (as opposed to “mind reading”) (2005:227). From a very early age infants imitate, they practice and improve their matching between their own body states and those they perceive in others. As Gallagher explains, the actions of the infant and the perceived actions of others are encoded in the same multimodal system. Our sensory-motor systems match those of others when we witness the actions of others (Chartrand and Bargh 1999; Barrett, Henzi et al. 2008). This matching forms the basis of inter-subjective understanding from the level of perceptual and motor systems, diminishing the need for a mentalistic “understanding of the intentions and dispositions of other persons” (Gallagher 2005:227). At this primitive level of understanding, we don’t need to infer the mental states behind the actions, because we derive meaning from our perceptions of the actions themselves.

The idea of using our own body states to understand others has some strong parallels with ‘simulation theory’ found in the theory of mind literature. Simulation theory maintains that some of our mindreading tasks can be achieved through the use of our cognitive system to simulate the activity of the target agent whose mind we are reading (Nichols and Stich 2003). Since our minds are relevantly similar to others, we can use the mechanisms of our own minds to understand the minds of others. One of the suggested advantages of such an approach is that it reduces the need to have an internally represented, rich body of knowledge about the mind, such as that posited by other theories of mind reading. The informational requirements are reduced by utilising machinery already available for other cognising. This is not to claim that ‘theory of mind’ is itself just simulation; many think the evidence points towards some hybrid meld between theory-theory and simulation theory (Nichols and Stich 2003; for a slightly different take see Gallagher 2005). But, both the evidence that children mimic others from a very early age, and the roles that embodiment effects play in cognitive processing,

suggest that body-based and sensory-motor mirroring have some role to play in the development of intersubjective understanding. (See also Gallese 2008.)

There is also an important role for gesture in communication and thought. Susan Goldin-Meadow details how gesture operates as a form of communication by providing a secondary representational format with which the listener can understand the communicator. She suggests that the mimetic and analog nature of gesture allows the representation of different types of information from the discrete nature of language. Embodied communication allows the communication of different types of information to that which can be conveyed in speech. Furthermore, Goldin-Meadow details empirical work on ways in which gesture functions as a means to facilitate thought and reduce brainbound cognitive load. When gesture is inhibited, for example, people perform more slowly on cognitive tasks (Goldin-Meadow 1999; Gallagher 2005). Think of the way that pointing at an object reduces the cognitive demands of spoken language, reducing the need to articulate to another to look in a certain direction and which object to look at and so forth. Gesture is a body-based cognitive tool to convey meaning, draw attention, and facilitate thought. Much of communication and thought (including that of moral knowledge – see below) is embodied in this way.

Although I have outlined just a small portion of the evidence detailing how many of our cognitive functions intimately involve body states, we can start to get a picture of how these processes function in the construction of embodied moral knowledge. Much of our moral development and execution of moral cognitive processes will involve body-based resources. When we moralise acts of harming, for example, we process emotional information pertaining to the victims', as well as our own, responses. The body-based displays of victim and caregivers are resources for the communication and understanding of the negative consequences of inflicting harm and the motivation to avoid such actions. The learning child uses their body-based processes and sensory-motor mirroring systems to interpret the intentions of perpetrator and victim, constructing moral knowledge by linking actions, consequences, agent responses, and the prohibited nature of certain acts. This knowledge is stored, in part, across affective and sensory-motor systems. Although the learning process may be slow, the construction of this knowledge can start early, prior to even to a command of language, as those various capacities come online. The application of this knowledge to moral states of affairs will involve the partial reactivation

of those cross-modal knowledge structures. Moral cognition is embodied cognition. Below I will examine why these body-based views are important for moral cognition, and the relevance of this thesis to the larger project – the evolution of moral cognition. But before doing so, I would like to address an important implication of the embodied view, namely the ways in which we reason through embodied simulations.

4.5 Moral simulations of moral situations

One feature of embodied cognition is that our modality-specific neural structures are central parts of our representational architecture and cognitions. This is in contrast with more traditional approaches which view representations as amodal and their associated cognitions as functionally separate from the modality-specific systems. The embodied approach presented earlier also claims that on-line experiences are captured in modality-specific states and are re-activated during off-line processing – what proponents of embodied cognition call “simulations” (Barsalou 2008) or “situation-models” (Zwaan 1999). This has important consequences for reasoning about moral dilemmas. Much of the time when we engage in off-line moral reasoning or moral-problem solving, we engage in the simulation of situations.

The idea of simulating moral scenarios in moral reasoning is intuitively plausible. Ethics text books are full of hypothetical moral scenarios and counter scenarios through which we can simulate the different sides of various debates. When Peter Singer (1993) asks us whether we would ruin an expensive suit to save a drowning boy we can picture ourselves in exactly that situation: coming across a small boy drowning in a pond while walking along in an expensive suit. In this way the abstract moral dilemmas are presented as a quasi-perceptually based problem to be solved, allowing our pattern-completing brains to operate independently of real-world stimulus.

Take for example the popular trolley bus problems. Thinking about pushing a fat man off a bridge into the path of an oncoming trolley bus will involve simulating that situation: a fat man, a bridge, a trolley bus, and the five people that may be saved. It will also involve simulating our actions over time and the consequences of our actions. We will simulate what it is like to push a fat man off a bridge, or at least imagine approaching the fat man to do so. When we think of the spare parts surgeon case, we imagine being a

surgeon, or at least observing one, in a room with a patient and a scalpel in the hand ready to take the life of the innocent patient to save the lives of others. And importantly in these cases, we imagine what it feels like. There is a strong link between imagination and affective/emotional states. Imagining dangerous events can induce fear (Vrana, Cuthbert et al. 1989), and imagining sexual acts can induce sexual arousal. Our evaluations and decisions will therefore depend on, and be influenced by, the various sensory-motor knowledge structures and processes that we use in our simulations, such as those associated with visual and other perceptual regions, causal reasoning systems, belief attribution systems, and limbic systems. Neuroscience lends some support to this idea. Disparate and functionally different brain regions are involved when we reason about moral dilemmas (Greene and Haidt 2002; Moll, De Oliveira-Souza et al. 2008; Moll, Zahn et al. 2005; Young, Cushman et al. 2007). By engaging these different regions and associated knowledge structures we simulate “being there”, and it is through this embodied simulation we can reason about the moral dilemma. Simulation solves the problem by using many of the same resources we would employ “as if” we were actually experiencing the situation (Gallese 2008).

There are a number of reasons this view is important from a moral perspective:

- 1) Simulations reduce need for complex, abstract, principles by offloading much of the computational work to our sensory-motor systems and body structures.¹¹
- 2) Simulations explain contextual influences.
- 3) Simulations have implications for moral learning.

I will look at each of these in turn.

11. It should be noted that simulations are contested idea. This includes debate over the representational resources required for simulations and the types of processes that operate over them.

Simulations reduce the need for complex abstract moral principles

Simulations reduce the need for complex, abstract, schematically structured, disembodied "moral grammars" that determine the judgements we make. Instead, we reason about moral dilemmas by simulating and solving the problems in real-time, as they arise. Think, for example, how one reasons about novel hypothetical situations such as the following: "While out on a walk in a safari park in Africa you come across a herd of rhino. As you move in for a closer look to take some photographs, inexplicably the rhino start to charge. Do you a) climb the large tree next to you, b) run away and hope they don't catch you, or c) stand still"? Although this is a scenario that (like trolley bus problems and spare parts surgeons) is foreign to almost all, I conjecture that we would almost all come to the same solution to the problem, without the need for any "when being chased by a large angry animal, climb a tree" principle to guide our actions. We can simulate the scenario and solve the problem in real time. This will involve bringing together various knowledge sources such as a 'what rhinos look like', 'what trees look like', 'fear of large, fast moving objects', 'a desire to not be hurt', 'knowledge of rhinos' tree climbing abilities', 'causal and consequential reasoning' and 'past experiences'. Simulating dilemmas transforms the problem into a quasi-perceptual one by offloading some work onto sensory-motor systems.

Simulations will involve contextual influences

When people simulate they adopt a particular perspective in the simulation. People tend to look up when thinking about skyscrapers and birds and look down when thinking about the bottom of a canyon, and they describe the relevant properties from that perspective (Barsalou 2003). One feature that research found is that when people simulate "being there" they also "infer introspective states likely to arise during these interactions" (2003:542). This means that our judgements will depend on the ways in which the scenarios are simulated, the ways they are presented, the knowledge we have available and how that knowledge is incorporated in those simulations. Particular aspects of a situation are made salient through simulation. This helps explain why moral judgements are subject to some context dependent effects. We see the influence on our moral cognitions of such factors as emotions and mood (Haidt 2001; Greene, Nystrom

et al. 2004; Valdesolo and DeSteno 2006), proximity to action (Prinz 2006, 2007), causal reasoning (Waldmann and Dieterich 2007), anticipation of consequences (Ciaramelli, Muccioli et al. 2007), conflicting moral principles (Nichols 2005; Prinz 2008c), framing effects (Sinnott-Armstrong 2008; Nadelhoffer and Feltz 2008), the actor-observer bias (Nadelhoffer and Feltz 2008) and group membership (Wohl and Reeder 2004). When we simulate what it is like to "be there", the various knowledge structures and reasoning processes that we employ will bias and alter the simulations we make. Angry and sad people are known to process social information differently and this effects the social judgements they make (Bodenhausen et. al. 1994). In heightened states of happiness we pay less consideration to the harm of an immoral act (Valdesolo and DeSteno 2006). Nadelhoffer and Feltz (2008) show that our moral judgements differ according to whether we imagine (simulate) ourselves as the protagonist or an observer. We can see these type of effects in how books and movies manipulate sympathy in this way; point of view is crucial to where one's induced sympathies lie. In some cases, knowledge structures will override others because they are more salient or motivating (recall Prinz' explanation for the doctrine of double effect and the fact that some acts of killing are more prototypical instances of killing than others). Research by Valdesolo (2006) and Haidt et al. (2005), presented earlier, shows triggered emotional states (incidental to the moral stimulus) bias moral judgement. From a simulation perspective, we can see that these states are included in the simulations that one makes, thereby directly biasing the simulated experience and hence the judgement made. To understand the judgements we make, and the moral reasoning we employ, we must understand the effects embodied knowledge has on our reasoning and decision making, the setup and framing effects of the dilemma, the perspectives we take and other biases. Our judgements are not (solely) due to context independent, abstract moral operatives.

Research by Ciaramelli et al. (2007) gives credence to this view. They compared patients with ventromedial damage (a region associated with social and emotional processing (Damasio 1994 (2005); Greene, Sommerville et al. 2001; Moll, Zahn et al. 2005)) to healthy subjects across a range of moral dilemmas. These included personal moral, impersonal moral, and non-moral dilemmas. The difference between the two groups was not in the non-moral dilemmas (suggesting no deficits in normal decision making) nor in impersonal moral dilemmas. But patients who had suffered damage to their ventromedial prefrontal (VM) cortex have a higher than normal tendency to approve emotive personal

moral violations. The authors claim this is because the VM patients' deficit means they fail to "anticipate the emotional, self-focused, long-term consequences of their choices" (2007:90). Because they cannot predict particular consequences of their actions, they fail to see how they will feel about their actions in the future (such as pushing a man to his death). Impersonal actions don't elicit emotional involvement to the same degree (pulling a switch in the impersonal case rather than pushing a man to his death in the personal case). In contrast, for normal people the emotional consequences of each act will be different. Hence the VM patients tend to treat personal moral dilemmas more like impersonal moral dilemmas.

Interpreting these results from the perspective of simulation, we can see how, without the appropriate emotional processing capacities the, VM patients fail to simulate the problem in the manner that normal subjects do. Certain emotional knowledge structures are absent from their simulations and therefore their predictions lack normal emotional components (such as empathy, regret and guilt). The lack of these emotional knowledge structures inhibits sensitivity to the differences between the simulation of personal and impersonal moral violations. Jesse Prinz (2007) argues similarly that psychopaths fail to differentiate between moral and non-moral violations because they do not have the same moral concepts as normal people. According to Prinz, moral concepts are essentially emotionally-laden (embodied), and because psychopaths have emotional deficits, they fail to have normal moral concepts. The amoral psychopath fails to have access to the relevant modality-specific brain regions (i.e., those associated with certain types of emotional processing) required to make the conceptual distinction between the moral and non-moral norms. Such emotional blindness, according to Prinz, results in moral blindness. Recall also that autistic children have difficulty in identifying moral transgressions involving lying. In such cases, the inability to simulate the relevant beliefs of other agents 'blinds' autistic children from making normal moral appraisals in these instances.

Simulations in moral learning

It is plausible that many of the non-moral knowledge and capacities that contribute to moral cognising are innate. But, as we have seen in chapter 2, we have reason to believe that moral knowledge is not innate. This places special importance on moral learning

environments in constructing those moral knowledge structures (see chapters 5 and 6). The simulation aspect of moral cognising has important implications for moral learning, because when we learn, we often simulate. This is especially true when we think of the ways in which we can gain knowledge from others, independent of actually experiencing an event. A clear example of this is through story telling. When we hear or read a story we construct a simulated world involving the characters and relevant details. We may be first person participants or third person observers to these simulated events. Through simulations (or situated models) we can experience the story and the events that take place (Zwaan 1999). Via simulations, children use their own sensory-motor and affective systems to experience the narrative *as if* they were there (Gallese 2008). The relevance to moral learning is clear when we consider that children's fairytales and nursery rhymes are full of moral events. These provide exemplars of moral narratives, detailing the causes and effects of moral actions as well as the consequences upon the transgressors. By creating a simulated world we can experience or witness the events that lead to moral violations, learning the various causal narratives of moral transgressions and their consequences. This plays an important role in passing on moral knowledge by establishing the relevant relations between, for example, sensory-motor and affective knowledge, belief attribution and intention that we can recruit later in our interactions with the moral world.

4.6 Implications for moral cognition

In the previous chapter we saw one account of moral cognition (that of the moral grammarians) that proposes an abstract (amodal), complex, internally represented, computational component of moral cognition. This component is autonomous (modular and encapsulated) and its operations do not regularly depend on interactions with, and feedback from, embodied resources. Further, this functionally autonomous component supposedly plays the driving role in moral judgement. In contrast, in this chapter I have developed an embodied view of (moral) cognition. The body and modality-specific structures are constituent, and functional, parts of cognition. Body-based systems shape both online as well as decoupled cognitive processes. Moral cognition (often) involves modality-specific structures and body states.

Embodiment effects can be seen in operation from two different perspectives: 1) agent

oriented approaches that focus on the roles that the body and sensory-motor structures play in an individual agent's cognitive system and mental life, and 2) at the social level where the focus is on how embodied effects and cues contribute to relational functioning between agents. Embodiment effects couple agents to each other and their normative world, influencing a social relations through emotional attachment, group coordination and prosocial behaviours (Smith 2008).

From the perspective of the individual, information processing roles are offloaded onto body-based structures, reducing the dependence on internal bodies of rich, complex, symbolic, amodal representational knowledge. Our bodies, according to the embodied view, do some aspects of moral cognising for us. Body-based responses also focus attention towards emotionally salient stimuli, filtering out unnecessary information, changing the priority of information processing and overriding other thought processes (Lerner and Keltner 2000).¹² In this way embodiment effects fundamentally alter our individual mental engagements with the world and our evaluations of that world. They couple us to features of the environment that are important to us, such as the harm and welfare of others that, as we saw in chapter 1, promote adaptive success. We can, for example, use our bodies from an early age to help us understand what harm feels like and what it must feel like for others to experience it. Our body-based experiences provide opportunity for individual understanding and generation of moral knowledge by structuring our experiences of the world.

Individual embodiment effects also predispose individuals to respond in particular ways to stimuli they encounter. Fear does not just focus attention on big lion-like creatures and prioritise cognitions about such creatures; it also motivates us to run or hide from them. In the moral domain, empathetic responses not only allow us to detect and focus on the suffering of others, but also motivate us to relieve that suffering. Here we see how embodiment provides the basis for moral cognitive processes to become collaborative through the development of joint goals and norms of behaviour. Embodiment effects play a central role in the establishment of social functioning (Barrett, Henzi et al. 2008; Barsalou 2008; Smith 2008). This is important from the moral point of view, because

12. By thinking of emotions in this way we can see how the feedback from the body operates as an interesting case of 'epistemic action' that reduces the processing demands on internal computational mechanisms. Recalling that Kirsh and Maglio define epistemic actions as physical actions which extract information and alter our information processing routines, making them "easier, faster, or more reliable" (Kirsh and Maglio 1994:3).

morality is inherently a social enterprise, and navigating the moral world is central to one's adaptive success. The mechanisms that operate upon the individual and the relational functioning between individuals are therefore of central importance in any explanation of human moral adaptive success. (I will discuss this further in chapter 5.3.)

We see an example of social-level influence in the ways in which we use the body to convey morally relevant knowledge and teach children to coordinate their behaviours with social-wide moral norms. We use our bodies in the learning and teaching of moral concepts, norms, and behaviours. We have already seen in chapter 2 the way that emotions and parenting practices bridge the gap between informational input and behavioural output in the moral domain. This seriously undermines poverty of the stimulus type arguments in favour of moral nativism. (Indeed, embodied approaches to cognition in general should force a reconsideration of all poverty of the stimulus type arguments about the mind.) An embodied approach to cognition transforms the ways in which the learning task should be conceived, from that of a disembodied observer, to one of an active and motivated participant in the learning process involving the engagement of systems across multiple modalities. Consider the case of childhood moral punishment. 'Telling off' will often involve the use of emotional displays of victim and admonisher, prosodic variations such as raised volume and pitch of voice, gestural actions such as finger wagging and pointing to the consequences of actions, hands on hips to signal authority, and sometimes physical restraint and or/violence. Consequential explanations generate simulations of the causes and consequences of transgressions. These embodied communicative tools operate across different modalities to convey morally relevant information. It is through our active bodily experiences of our social worlds that we learn and construct the moral cognitive abilities and problem-solving routines required to navigate the moral world.

Moral culture plays a central role in providing structured environments from which children learn. But importantly our embodied dispositions also structure those moral environments. By immersing ourselves in moral environments that are themselves largely structured about our bodily reactions (Nichols 2002b, 2004), our embodied engagement with those environments promotes a reciprocal causation: moral culture exists because we have the embodied experiences that we do, but in turn we have the moral capacities that we do because we have the moral cultures in which we are immersed. There is an

intimate and reciprocal relationship between the development of our moral selves and our moral worlds. Embodiment takes on a central role in extending the moral mind, not only through the body, but beyond the body, by actively coupling agents to their external world.

Clearly this is important from a Darwinian point of view. Moral cognition evolved as a solution to adaptive problems of cooperation and social coordination – it is the product of natural selection. Yet, in contrast to the evolutionary psychologists' picture of encapsulated, domain-specific, algorithm containing modules that resided in the neural architecture of the head, we instead have a picture of dynamic, reciprocally influencing processes that are spread over brain, body, and as I will argue in the following chapter, world. The target of selection is the patterned and coordinated interactions between these three components. Selection is for the entire system, not just the neural system. And, as we will see in the following chapter, many components of the moral cognitive system reside beyond the organismic bounds of the agent.

4.7 Conclusion

The account of auditory cognition we saw earlier emphasised the way in which it involves reciprocal feedback loops between brain, body and world. The structures involved in this feedback process do some of the work and are part of the cognitive system. We should think of the embodiment of moral cognition in the same way. We are not disembodied and dispassionate moralisers who think about morality in a way that is divorced from our bodies, nor the effects upon our bodies. The feedback loops between brain, body and world are central to the navigation of the moral world. Even moral reasoning is embodied, for it uses the very same somatosensory and sensory-motor structures that we use in real-world moral action.

The fact that moral cognition is embodied means that many of the central determinants and complexity of moral cognition reside beyond the head. Body and environment change and shape our interactions with the problem domain, and hence shape the

adaptive routines that we employ to solve problems in that domain.¹³ Moral problem solving does not depend solely on the neural system. Furthermore, off-line moral problem solving is also embodied. Much of folk moral deliberation, debate and learning involve the simulation of moral situations far removed from real-world moral events. In simulating we bringing together disparate knowledge structures stored across perceptual, affective, motor and introspective states. Simulations transform abstract moral dilemmas into quasi-perceptual problems. Embodiment also facilitates the establishment of society-wide moral systems through the generation of joint goals and norms of behaviour, and plays an important role in coupling agents to their moral worlds (I will further explore these social roles of embodiment in the next chapter). As I have demonstrated, many of the drivers of complex moral behaviours reside outside the brain and neural system (and more specifically amodal representational systems). Through the integration and offloading of cognitive processing to our body-based structures, those structures become constituent parts of that system. The moral cognitive system includes our body.

4.8 Summary

In this chapter we have looked at embodied cognition and the idea that moral cognition is embodied – body-based systems are constituent of moral cognition. This has important ramifications for the way we view moral cognition and evolutionary accounts of cognition.

The main points are as follows:

- Moral cognition involves the use of multi-modal representational structures as well as body structures.
- Offloading of moral cognition onto body-based structures reduces the computational requirements of a brainbound cognitive system.

13. Chiel et al. state we should adopt such a position because the “most important evidence suggesting that the nervous system cannot be the exclusive focus for understanding adaptive behaviour is that it continuously receives and responds to feedback both from the movements that it induces in its own periphery and from the surrounding environment.” (1997:555).

- Off-line (decoupled) moral cognitions are also embodied. When we think morally in the absence of moral stimulus we make use of modality-specific representational structures by way of simulations.
- Embodied moral cognition is situationally dependent.
- Embodiment effects are used in the understanding of others and the formation of joint goals of action.
- We have reason to be sceptical of individualist accounts of moral cognition such as those proposed by the Hauser-Mikhail model of the moral mind.

In the next chapter we will look at ways in which the moral cognition extends beyond the bounds of the body.

5

Extending Moral Cognition

... there is nothing sacred about skull and skin.

(Clark and Chalmers 1998)

In the previous chapter I outlined the ways in which moral cognition is partly constituted by the body, concluding that the moral cognitive system is not confined to the head of the moral cognisor. I will now explore the various ways in which resources *external* to the body become part of our moral cognitive system and how this cognitive extension creates an extended moral space within which we operate. This will provide an understanding of how cognitively-extended moral agents are embedded within moral worlds and how the structure of those worlds enables the extension of moral cognition. This demonstrates the ineliminable role that the environment plays in the realisation of much of moral cognition.

5.1 Beyond the organism

Richard Dawkins (1982) famously ‘extended the phenotype’, arguing that genes have adaptive phenotypic effects that extend beyond the organism boundary. The spider’s web, the beaver’s dam, and the termite’s mound are all examples of phenotypic effects that are targets of selection beyond the physical boundaries of the organism. In these cases it is not the organism that is the ‘bearer of adaptation’, but the gene’s “adaptive effect is outside of the body they inhabit” (Sterelny and Griffiths 1999:75). The

explanation for why the spider has genes that code for the making of webs is that the web and its fly catching effects are an adaptation for the benefit of the spider's genes, even though the web is physically external to the organism. To restrict the gene's expression to those traits bounded by the physical limits of the organism is, Dawkins argues, arbitrary. The organism-bound view of trait expression is inadequate to explain the selection of many gene lineages. By recognising that some genetic traits have adaptive effects outside of the organism, we have a better explanation for the selection of those genes. The positing of extended phenotypes figure in our *best explanation* of the selection of some traits.

Proponents of the extended mind thesis take the notion of *extension beyond the organism* and apply it to cognition.¹ Cognition often depends on and incorporates resources external to the physical bounds of the agent. For example, Clark and Chalmers (1998) argue that external storage devices such as notebooks can, under the right conditions, be integrated into an extended memory system. In such cases, the notebook constitutes part of the agent's mind; the mind literally crosses the physical boundaries between organism and world.

In the previous chapter, we saw how organism-bound moral cognition incorporates the beyond-the-brain body. Body structures are constituent parts of moral cognition. Taking this a step further, we will now see how moral cognition is extended cognition through the inclusion of external environmental/social structures. These operate to enhance and, in some cases, give rise to new, moral cognitive abilities. This is a more substantial claim than saying our moral cognitive processes are *supported* by external structures or that those external structures trigger moral cognitive processes. The claim is that moral cognition is (sometimes) partly constituted by those physically external structures.

There are two important aspects to this view of the agent-world relationship. The first is the way in which moral cognition is extended through the incorporation of physically external structures. As with the reciprocally influencing brain-body dynamic of embodied cognition, moral cognition often involves a reciprocally influencing agent-world dynamic. As moral agents, we are deeply embedded in a world that tracks and controls our moral

1. See for example Clark and Chalmers (1998), Clark (2008), Wilson (2000) and Menary (forthcoming-a).

actions. This world also provides resources with which we can track, control and remember the moral behaviours of other agents. Moral cognition is extended through the integration of these resources into our moral cognitive processes. We are the moral agents we are because we stand in particular relations with our moral worlds.

The second is the creation and nature of the external structure itself. Our moral worlds exist because we, as members of that world, are moral agents. The collective behaviours of moral agents create our moral worlds by establishing and maintaining social-wide moral systems which in turn scaffold both the development and real-time execution of extended moral cognitive processes. Here, the collectively created moral world is a resource which becomes partly constitutive of individually extended cognition. A useful analogy to help understand the relationship between the individual moral agent and their moral group is that of the individual termite and the termite colony. Each individual termite is able to undertake colony-directed behaviours because the colony itself provides the environment and resources required to instantiate those behaviours. Individual nutritional needs, for example, are met through the group cultivation of mycelium. In turn, the colony exists because of the collective behaviours of the individual termites that create the various structures that sustain the colony. In the case of the extended moral agent, their extended moral cognition is manifested because of the existence of an external moral world that supports the cognitive integration of aspects of that moral world. Our moral groups provide the various external resources which we incorporate into our extended moral cognition. In turn, that moral world exists and has the various dynamics and properties because of the interactive behaviours of agents within those groups.² This view places particular importance on the role of our moral worlds in the development and execution of moral cognition. As we will see in the latter stages of this chapter, and again in chapter 6, this has implications for an evolutionary point of view because of the particular importance of the organism-environment relationship. This view is in contrast with evolutionary psychologists and various moral nativists, who claim that over long periods of evolutionary time, external environments populate the internal psychological mechanisms (modules) – moral cognition is ‘on the inside’. Instead, we will see the coevolution of individuals and their moral environments, and the mutually

2. Note, the claim here is not that moral groups are themselves moral cognitive agents, although it is an interesting and important issue, especially with respect to such things as individual and group moral responsibility, I am, for present purposes, agnostic on this point. The claim that I am making is that certain individual moral cognitions are realised in the presence of groups of moral individuals.

sustaining dynamic between them.

I will now look at a number of aspects of moral cognitive extension. The first involves the ways in which embodiment effects couple agents to other agents as well as to their normatively structured world. We live in a mutually coordinating moral community structured around our embodied responses. These embodied responses allow agents to make use of that world as part of their moral cognitions. The second is the ways in which many moral cognitive tasks such as identifying, tracking and remembering others' moral behaviours and reputations are distributed amongst groups. By accessing these external moral knowledge structures, much of our moral cognitive work is offloaded to, and distributed across, our moral community. The third is how moral norms operate in extended moral cognitive systems. Moral norms are public artefacts that we incorporate into our moral cognitive processes. And finally, I will examine how moral debate and group problem solving creates an adaptive 'moral space' within which we operate as extended moral agents. I will address each of these in turn, discussing some sceptical responses along the way.

Before exploring the idea of moral cognitive extension it is worth highlighting the fact that moral cognition requires the use of external resources. As we saw in chapter 1, this often involves the tracking and controlling of other agents and, as such, moral cognition can be said to be outward looking. As we will see later, we include external structures to achieve these outward looking cognitive tasks.

5.2 Moral cognition is also outward looking

Much of moral cognition is to do with self-regulation. Self-referential moral judgements such as shame and guilt, for example, are mechanisms that enforce individual moral adherence and play a central role in much of morality (Joyce 2006). But, it is also typically in one's interests that other group members adhere to moral standards and a very large part of morality deals with other-person moral adherence. For this we have a suite of mechanisms by which we track and control the moral behaviour of other agents. These additional mechanisms are *other directed*. Moral cognition is, in this way, often outward looking.

Anecdotally it seems obvious that we monitor and attempt to control the moral behaviours of others and punish them for their moral infractions. We often, for example, get angry at people who violate moral norms and we feel they deserve to be punished. Think of those who line up outside courthouses to hurl abuse at child murderers. Many of those people have never known the victims, nor perhaps were they even part of the victim's immediate community. Such cases illustrate the fact that people will go out of their way to punish others at a cost to themselves, even when they are not directly affected by the violation itself. There is strong empirical evidence supporting this, and that people acting in this way are not doing so for obviously selfish reasons. Fehr and Fischbacher et al. (2002) studied the behaviours of people in public goods games. They found a significantly large percentage of people reliably punished others for violating norms of cooperation, even when punishing the transgressor incurred additional costs. Further, subsequent studies by Fehr and Fischbacher (2004) found that people reliably punished violations of fair distribution and cooperation norms when they were merely observers of the norm violation, and in no way directly affected (in terms of costs or benefits) by the violation. People punish others, even at a substantial cost to themselves, suggesting they are not acting out of obviously self-interested reasons (see also Bowles and Gintis 2003; Sripada and Stich 2005). Attitudes that motivate punishment, revenge, and retribution are deeply powerful action generators.

What is clear is that as well as being motivated as individuals to adhere to moral norms, we are motivated to ensure that other agents adhere to those same moral norms; we think others ought to be moral and we punish those who are not.³ The monitoring and controlling of others is therefore centrally important to us as moral agents, both psychologically and, as we saw in chapter 1, adaptively. Many of the structures and processes that enable the tracking and controlling of other agents are external to the physical bounds of the moral agent. We incorporate these external structures into our moral cognitions.

Before moving to a more detailed analysis of moral cognitive extension, I will first look at the role of embodiment in establishing the close coupling between agents and their environments.

3. As we saw in chapter 1, punishment need not be physical. Non-cooperation strategies such as ostracism and the threat of ostracism are effective enforcement strategies, placing central importance on reputation and gossip as effective mechanisms of cooperative adherence.

5.3 Embodiment and moral cognitive extension

Extended cognitive systems exist when structures physically external to an organism become coupled to the organism. But, it is not enough that they are merely causally connected to the organism, they must be “coupled” in the appropriate way. Cognitively coupled systems involve mutually influencing causal relations between the components of the system. This “continuous reciprocal causation” (Clark 1997) between components of the system produces the cognitive behaviour in question. Isolated analysis of singular components within the system will fail to adequately explain the behaviour, which depends on the coupled system as a whole. When there is appropriate coupling, the coupled system provides the supervenience base for the cognitive capacities in question. Embodied moral cognition helps to establish the appropriate organism-moral world coupling in a number of ways. Firstly, it plays an epistemic role by focusing attention on morally relevant aspects of the environment. Secondly, it couples agents with other agents. And thirdly, it couples agents with their normative worlds. I will look at each of these in turn.

The body-based information processing described in the previous chapter plays an epistemic role by focusing attention on morally relevant information in our environments. Our embodied information processing and the embodied responses of others direct our attention towards the behaviours and responses of other agents and our moral worlds. We see this in our responses and attitudes towards moral norms, the moral behaviours of other agents, moral reputations and the consequences of actions upon others. This epistemic aid is important for picking out the relevant features of our moral environments. We care, for example, about moral norms. We not only act in accordance with those norms, we try to bring others in line with them – behaviours which are underwritten by embodied responses. I will argue below that moral norms operate as moral cognitive extensions to track and control other agents. When we couple with external normative structures they become resources for cognitive extension. This allows the exploitation of, and cognitive offloading onto, that social structure.

Embodiment effects also help to couple individual agents with other moral agents. Through this agent-agent coupling we see the establishment of joint intention and

coordinated action. In everyday interactions we use our bodies not just to signal to others, but to coordinate with others' actions. We use gesturing as a means of communicating information (Goldin-Meadow 1999; Gallagher 2005). We also use our bodily actions and cues to manipulate others' behaviours and cognitive states. We may gaze at an object both to signal to another that we are looking at the object, and with the intention of getting them to also look at the object. Importantly then, embodied actions do not merely alter our own cognitive processes; they also alter other agents' cognitive processes and behaviours (Smith 2008). Our embodied responses become tools by which we can (consciously or unconsciously) manipulate others.

An example of this type of manipulation can be seen in the interactions of mother-infant dyads (Spurrett and Cowley 2004). Children cry for want of attention, or perhaps to signal fear or pain. They cry in order to satisfy their needs and desires. Embodied displays such as those associated with crying and the corresponding sensitivities of mothers act to link the child and mother, thereby coordinating the mother's actions with the child's needs and desires, and vice versa. The affective and other embodied actions of mother and child are used to alter the cognitive states of each other. This process is bidirectional and reciprocal. Mothers respond to the actions and vocalisations of the child by augmenting their responses to that of the child, imposing their own normative structure and meaning which the child comes to understand and reciprocate. To illustrate this embodied reciprocal manipulation, Spurrett and Cowley detail how a four-month-old child learns to signals to his mother that he wishes to be picked up:

an infant repeatedly vocalises in ways which to its mother, at least, are suggestive of its saying 'up'. Each time she says 'up?', or 'you want to go up?' and after a few repetitions she lifts the child. Prior to the lifting, there is little evidence that the child actually wants to be lifted, or that it has its attention focussed on anything in particular, except perhaps its own experiments in vocal control. When it is lifted, though, it beams widely. Whatever it did want, if anything, it is now, we suggest, one step closer to figuring out how to behave in ways that lead to its being lifted up. (2004:456-457)

The mother provides regularity and structure to the child's environment. As a result, the child can learn to exploit that external structure. As the child develops, she learns to use the embodied actions involved in lifting as gesturing signals for caregivers to pick them

up. Children “raise and flap” their arms – not gestures that adults commonly use – to signal their desire to be picked up. The raising of the arms is a “common posture of infants while they are in fact being held up” and is interpreted by the mother as being an invitation to be lifted (Spurrett and Cowley 2004). Embodied actions such as these alter both the child’s and the mother’s cognitive states. These types of caregiver-child interactions are embodied and reciprocal, each imparting physical and emotional regularity to obtain joint goals.

This example illustrates how coordinated action arises from the joint attuning of the brain-body-world components of dyadic interactions. Importantly, the child is learning the *norms of interaction* and how those norms can be used to support the achievement of one’s goals and desires – in this case the goal of being picked up by a caregiver. Furthermore, Spurrett and Cowley point out that infants are, “in virtue of affective co-ordination, able to function as a kind of cognitive extension of their own caregivers, who focus their attention, regulate their levels of arousal, reinforce and retard patterns in their behaviour, and provide all manner of sources of environmental regularity amenable for infant exploitation” (2004:463). Bodily effects are not just communicative resources; they are also resources by which we can couple with, manipulate and exploit other agent’s cognitions and actions.

In the moral case our embodied interactions work to link actor with perceiver to coordinate other agents’ responses: embodied responses elicit patterns of behaviour in others. We use our own bodily actions to draw attention, align intentions, and invoke joint action towards moral events. Embodied reactions exploit the emotional sensitivities of others, so that the responses of others become part of a coordinated response. When someone is morally slighted they typically react emotionally and bodily. Moral argument is emotionally heated. In such cases our embodied responses can be used to elicit embodied responses in others.

From this, we can see that embodied reactions have effects beyond those of the victim of the moral violation. We saw in chapter 2 the role that embodiment plays in moral categorisation. But the effects of embodiment are involved in more than moral categorisation tasks. In terms of the punishment of norm violators, for example, it is not that people are solely motivated by the intrinsic normative motivations to punish norm

violators (Sripada and Stich 2005), they are also motivated *because* of the direct emotional (embodied) consequences of such norm violations upon the victim of the violation. People do not like to see others in pain, sadness, or distress and are motivated to respond. The consequences of a moral infraction are not isolated to the victim of the infraction; perpetrators may be subject to self-referential guilt and shame, in part, because of the responses of the victim. Additionally, third-party observers are upset and they too are motivated to act (think of the courthouse protestors mentioned earlier). My distress at having my bicycle stolen is transferred to other agents within my immediate group, thereby transferring moral upset to others in my community. In turn this motivates action not only in me but also in others. Should I be indifferent, my community are unlikely to be upset. My suffering begets others' moral anger and upset in ways that coordinate our moral response. Moral information is transferred through embodiment: displays of moral indignation, anger and disgust act as signals to the agent committing the moral infraction, as well as signals to my group.⁴ Via this information spread, my embodied actions have causal consequences for other agents. In turn, *these consequences are patterns of action that we can reliably exploit and manipulate as moral agents* (see below). These response chains allow the establishment of group-wide systems of collective behaviours and in this way embodiment facilitates the construction of a moral community of embodied agents.

But it is also the case that our moral environments exploit our embodied dispositions. Shaun Nichols, for one, has detailed how moral norms, such as those prohibiting harm, “gain greater cultural fitness when they prohibit actions that are likely to elicit negative affect” (Nichols 2004:138, 2002b). According to Nichols, our emotional sensitivities help explain why particular norms are culturally successful because they ‘resonate’ with our emotions and therefore have an advantage in cultural evolution. The embodied nature of our interactions connects us to moral norms enabling us to use that social world, and the social world exploits the embodied nature of those interactions. The relationship between moral agent and their world is therefore dynamic and reciprocal: embodied agents create and structure the moral world, and in turn we have the moral norms, practices, and moral judgments that we do, (in part) because we are embedded within

4. These effects are not only mechanisms of threat and enforcement (partner control) but also displays of moral trustworthiness (partner identification) (Frank 1988, 2001). In this way my embodied responses to moral situations provide cause for others to adhere to moral norms as well as signals for trustful participation in cooperative activity.

those moral worlds. These factors work together to strengthen and maintain the mutual reciprocal influence and intimacy of the coupling between agents and their moral environments.

It is apparent from the examples provided above that embodiment effects play a number of important roles in moral cognitive extension. They provide an epistemic resource which focuses our attention towards particular aspects of moral worlds. They also play a pivotal role in coupling agents with other agents. Through this agent-agent coupling we can form joint intentions and actions, which in turn become resources that can be exploited and manipulated by moral agents. And finally, embodiment enables the construction of normative worlds as a consequence of forming joint intentions and actions within groups. The body is not just the medium through which we engage in the world, it plays an important role in the establishment of interactive, reciprocal dynamics between agents and their moral worlds.

So far, we have seen that embodiment facilitates close coupling between moral agents, as well as the normative structure of their moral world. In the next section we will look at how our moral cognitive systems extend beyond the reach of the physical boundaries of the body by incorporating external resources into our moral problem-solving routines.

5.4 Extending moral cognition

Humans are born to moral worlds and throughout development they learn to interact with those worlds. As moral agents they are embedded within, and reciprocally interact with, their moral group. This group and its associated moral systems allow the extension of moral cognitive systems. What is central are the resources those groups provide for extending the moral mind. They provide the external, cultural structures (moral norms, gossip, tracking, memory etc) that can be integrated with (to become part of) the individual agents' moral cognitive system. By incorporating features of our moral worlds into their moral cognitive routines, the moral agent is able to reduce internal cognitive load and attain cognitive abilities that would not arise independently of the group and its moral structure.

An instructive way of understanding these processes is by looking at the case of extended memory. When certain environmental resources are coupled to the mind in the appropriate way they are partly constitutive of memory. Clark and Chalmers (1998) argue that environmental artefacts such as notebooks can fulfil this role, where the notebook substitutes for the usual role afforded to biological memory. On this view, remembering can involve accessing non-biologically encoded information in addition to biologically encoded information. Clark and Chalmers agree we should accept such a view because biological and non-biological based information encoding often play functionally equivalent roles in remembering: “were it to go on in the head, we would have no hesitation in accepting it as part of the cognitive process, then that of the world is (for that time) part of the cognitive process” (1998:644). The processes of memory are often constituted by encoding, storage and retrieval mechanisms that exist beyond the physical bounds of the individual. These processes are distributed over the organism and physically external aspects of the environment. Importantly, the process of remembering can include the cognitive activities of other individuals (Barnier, Sutton et al. 2008).

Barnier and Sutton et al. detail evidence in support of the social manifestation of certain remembering processes.⁵ Their evidence details how group recall of events involves the sharing, negotiation and reconstruction of experience. These episodes of collaborative group recall influence both the content of the group recollections as well as subsequent individual recall of those events. For example, their studies found that collective group recall of the death of a famous person produced qualitatively different recollections compared to individual recollections. Group recollections tended to show a lack of personal identification with the emotional experiences discussed in relation to the death (e.g., “people were sad that he died” rather than “I was sad that he died”), whereas individual recall more heavily emphasised personal emotional experiences. Furthermore, when individual participants were asked to recall the event at a later time, those individuals involved in collaborative recall produced qualitatively different recall of the event than those who had not taken part in the group. Individuals who partook in group recall adopt the memories of the group. Partaking in collective group recall thus results in the construction of individual memories that are different to those individuals that did not partake in group recall.

5. Here we see a memory specific application of Robert Wilson’s Social Manifestation Thesis: “some psychological states of individuals are manifested only when those individuals form part of a social group of a certain type” (2001:S269-S270).

Further empirical research also tells us that much of memory is a constructive process which involves pooling different information sources to (re-)construct memories of objects, situations or events. Photographs, for example, are physically external resources that we use in recall. But importantly, photographs do not merely trigger established memories, they are often active in the construction of memories. We see suggestive evidence of this from the large body of work detailing how this process is corruptible. People can be easily led to falsely remember past, non-existent events. False narratives and doctored photographs induce people to recall experiences that they never had (Strange, Gerrie et al. 2005). This evidence suggests that we offload to our environments much of the cognitive effort in storing and remembering, which means those environments are partly constitutive of remembering.

These examples show that often resources external to the remembering agent do not merely play a triggering role by activating storage mechanisms internal to the agent (see discussion in Barnier, Sutton et al. 2008).⁶ Some individual remembering can only be realised when the individual stands in particular relation with their social or physical environments. Moral groups operate in a similar way by enabling the realisation of certain moral cognitions. Individual moral agents offload moral cognitive processing to their group. The collective actions of the group provide resources which become constituents of the agent's moral cognition. For example, one of the roles of moral cognition is to identify and track the moral behaviours of other agents, but independently we cannot commit to memory nor have access to the moral history of everyone in our community. But as a group we can collectively remember the behaviours of group members. We can discuss (gossip) with others their moral knowledge. We construct shared moral opinions. Groups allow the realisation of moral cognitions that are beyond the capacities of any one individual. Using others as memory stores, for example, reduces individual cognitive load while at the same time amplifying potential group problem solving abilities through

6. According to the triggering view, external resources are not constitutive of remembering: the process of remembering is an internal, organism bound process. This triggering account of memory cognition is described by Barnier and Sutton et al. as one where the role of physically external resources and phenomena is to initiate recall, or provide inputs to the internal encoding processes (cf. linear processing cycle of traditional accounts of cognition described in chapter 3: 'input → cognition → output'). Although triggering may be true of some of our remembering processes, Barnier et al. reject it as a global thesis. As the evidence shows, in many cases "remembering is a cognitive process that can only be manifested or realized when the individuals engaged in that process form part of a social group of a certain kind" (Barnier, Sutton et al. 2008).

the division of cognitive labour. But the distribution of moral cognition over external moral scaffolds is not isolated to moral memory. Groups also act as a collective stores and negotiating tools for the construction of resources such as moral norms, concepts and the skills with which to navigate our moral worlds. External social mechanisms are an essential support for moral cognition. (In section 5.5.1, we will see that these external mechanisms meet the requirements for cognitive extension.)

Here we can see the significant role that gossip plays in distributing moral cognitive load. Gossip operates as a low cost method by which to disseminate moral information about group members. As Robin Dunbar (1996) argues, most of our everyday talk is about social relationships and the behaviours of others. An important component of this is gossip about the moral deeds (good and bad) and moral reputations of other members of our group (Boehm 2000; Sripada and Stich 2005; Joyce 2006). This moral gossip is important, for we need to know who to trust and who to sanction. Gossip also allows group members to safely attain negative information about “dangerous deviants” in private (Boehm 2000:94). It does a very effective job of distributing the cognitive load required to track large numbers of agents relatively safely. Through the collective creation of socially shared knowledge structures that index the moral reputations of agents within a group, we use other agents as epistemic tools to access information about our social worlds that we would not easily be privy to on our own.

Gossip also increases the potential costs of defection. Our defections are not only known to those that directly witness them, but also to those with whom they gossip, thereby increasing the number of potential punishers. Part of the incentive to maintain adherence to normative standards is that news of one’s infraction will be passed on to others. The low cost transfer of moral information works as a mechanism to maintain moral adherence. Collective information stores amplify the threat of punishment, thereby reducing the cognitive costs of individual punishment strategies. As such, groups provide distributed mechanisms for maintaining normative adherence by reducing the benefits of free-riding and hence the temptation to defect.

We see this kind of distributed mechanism in the way that group-wide norms encouraging the punishment of free-riders reduce the cost of punishment by distributing that cost over the entire group of punishers, through what Boyd and Richerson call

“moralistic punishment” (Boyd and Richerson 2004; Bowles and Gintis 2003, see also chapter 1). If there is a group-wide norm for punishing defectors or non-conforming agents, then the potential threat of punishment can come from any member of the group who adheres to that norm. Norms spreading the willingness to punish and the subsequent threat of punishment reduce the costs of punishment by spreading those costs over all agents, and in turn reducing the *average cost of punishment* for individual agents of the group. Moralistic norms such as these can stabilise a vast range of different behaviours (Boyd and Richerson 2004). As group agents we rely on others to maintain the threat of punishment, or to at least have the willingness to inform those who will enforce punishments. Again, the threat of punishment is distributed amongst the group, reducing cognitive load (tracking, identifying, making judgement, and exacting punishment) while also increasing the risk to defectors (reduced per head cost of punishment and increasing costs of defection). We tacitly rely on moralistic punishment norms to minimise the free-riding by others in our group.

Group interactions provide unique opportunities and resources unavailable to individual agents because groups fundamentally change the problems that confront the individual moral agent, as well as provide the resources and opportunities to solve those problems. The solutions to many of our moral problems are group solutions (specifically those to do with cooperation and coordination) and they require the appropriate structuring and interactions of multiple agents and artefacts. Many of our adaptive problem-solving abilities within the moral domain thus arise through the particular dynamics of agent-group interactions. The relations between the individual moral agent and the group structures within which they are embedded are centrally important for understanding moral cognition. We have seen that the structure of our moral worlds allows the achievement of cognitive processes that would not normally be available to the non-embedded moral agent. Although identifying, tracking, memorising, and punishing are moral cognitive processes available to the non-extended moral cognitive agent, coupling with our moral worlds enables their distribution over physically external structures, thus furthering their reach through low cost means. By placing moral agents within a web of external moral scaffolds, moral cognition is distributed over organism and moral world, individual internal cognitive load is reduced and cognitive reach is increased.

So far, I have explained how we, as moral agents, are embedded within a moral world

which provides various social supports, allowing the manifestation of certain moral cognitive capacities. I will now look at a more specific case, namely how moral norms become part of an extended moral cognition.

5.5 Moral norms and cognitive extension

Norms are, to a rough approximation, rules or principles that specify patterns of action that are required or deemed to be permissible or impermissible (Sripada and Stich 2005). They operate at multiple levels of action; functioning in individual, interpersonal, group, as well as broader cultural/social contexts. As we saw in chapter 1, moral norms are important for our adaptive success. They are adaptive for both individuals and groups because they provide important solutions to cooperation and coordination problems. Not all norms (moral or otherwise) are adaptive. As Boyd and Richerson point out, many clearly are not (“the adaptive value of ritually handling rattlesnakes is hard to fathom” (2004:168)) but, (generally) the normative capacities of agents are. Being able to identify and follow norms, and being part of a group that enforces norms, is typically fitness enhancing (Bowles and Gintis 2003). Moral norms are also part of our extended moral cognition.⁷

I propose that one explanatorily fruitful way to view norms is as a form of public artefact: they are adaptive action-guiding information stores about social behaviour. They provide agents with the resource to successfully navigate their social worlds and for groups to negotiate cooperation and coordination problems. The information that they provide is portable, readily accessible, and reliable. And, much like the case of extended

7. Kim Sterelny (2004) is sceptical of the extended mind thesis on the basis that external cognitive resources are corruptible by other agents because they “operate in a common and contested space”. If this is a problem for external cognitive resources (although see Clark (2008) in response), it is not clear that it is so in the moral case. Although morality operates in a contested space involving negotiation, debate and norm violation, it also works to dampen many of their effects. In the case of moral cognitive extension there are specific mechanisms that maintain the integrity of the external normative structure. Indeed *much of the point of moral cognising is to retain the integrity and regularity of normative structure and associated behaviour*. Norm conformity is supported by, for example, sanction and punishment as well as meta-punishment norms enforcing norms of enforcement. These are driven by a raft of motivational mechanisms such as moralistic emotions (Bowles and Gintis 2003; Boyd and Richerson 2004). These mechanisms provide what Ehrlich and Levin (2005) call “cultural stickiness”, which helps to maintain normative integrity and reliability by retarding normative change, and strengthening conformist transmission and minimise temptation to defect. Norms and normativity are, by definition, sources of regularity. Therefore it is doubtful that norms acting as external cognitive resources are subject to the same corruptibility and outright contestability that other external resources might be.

memory earlier, they come about via constructive processes involving the interactions of multiple agents.

Viewing moral norms in this way, we can see that they are extensions of moral cognitive systems. Firstly, they provide us with a ready information store of how one ought to act in the social world. Moral norms guide our moral behaviours and the judgements that we make about others. Throughout development we learn to use moral norms as information stores which we can readily and reliably access.

Secondly, moral norms provide information about how we should expect others to act. If I learn that my social group does not do X, then I have reliable (although by no means infallible) information that members of my social group will not do X. Moral norms provide agents with expectations about the behaviours of group members. In many cases, I do not need to track all other agents in my social groups as there is good reason to think that *because* they are from my social group they are subject to many of the same normative constraints that I am. Social-wide moral norms are a low cost coordination device which we can exploit as information stores about the behaviours of others.

Thirdly, moral norms are part of our extended cognitive apparatus because we rely on them to circumscribe the moral behaviour of other agents in one's social group by proxy, thereby reducing the cognitive load required of continuous moral vigilance. One of the roles of moral cognition is to proscribe and thus constrain other agents' behaviours. Moral norms do not simply provide us with information about how other agents will act, they partially determine the behaviour of those agents. They are action guiding in addition to being descriptive. As such, moral norms are part of our moral cognitive system, operating to track, manipulate and constrain other agents *as if* a non-extended agent were to do so themselves. One way I can ensure you would not steal money that I have left in my room is by staying in the room with you to watch your every move. The other method is by knowing that you subscribe to the patterns of behaviour that prohibit stealing money. In the latter case, we are offloading much of our moral cognitive work to our environments by exploiting the normative structure of our society. I know that the local shop owner is unlikely to short-change me because I can reliably assume that he is constrained by the norms of our society, irrespective of whether or not his compliance to those norms is moralistic (i.e., he thinks it is morally wrong to short change people) or

self-interested (i.e., he does not want to get caught). This self-monitoring and self-guidance by other norm bound agents reduces the brainbound requirements for continuous individual moral vigilance. The idea is that we are closely coupled to the social-wide patterns of moral behaviour, so that we exploit the normative regularities and structures of our moral worlds (cf. mother-infant dyads presented earlier). Moral norms operate as part of our extended moral cognitive system by doing cognitive work, namely maintaining moral vigilance and constraining other's behaviours.

5.5.1 Supports as Constituents

There may be some scepticism as to whether we can legitimately include physically external, society-wide moral norms as part of our moral cognitive system. This is part of a larger issue about what resources are, and under what conditions ought we consider resources, a legitimate part of our extended cognitive system; without constraints on what is to be considered legitimate parts of cognition, the extended mind faces accusations of triviality. In response to these issues, as well as an attempt to bypass the inherent boundary bias created by the physical boundaries of organisms, Clark and Chalmers propose the Parity Principle:

If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process. (1998:8)

Here the focus is on the functional nature of the “system” rather than the physical boundaries of the organism itself (see also Clark 2008). As I have explained above, moral norms are physically external resources which track and control the behaviours of other agents. Were we to perform the functionally equivalent tasks in our head, we would rightly consider them part of our cognitive processes. Moral norms satisfy the parity principle.

Additionally, Clark and Chalmers (1998; Clark 2008) offer a four-point ‘rough guide’ to inclusion into one’s cognitive system. We have good reason to include moral norms as part of one’s extended moral mind as they satisfy each criterion:

1. That the resource be reliably available and typically invoked.

People readily adhere to moral norms. Knowledge of that fact, and the norms which they abide by, is readily available to members of the community. As we have seen in chapter 1, the point of much of morality is to maintain reliable and predictable behaviours of group members that generate the benefits of cooperative interaction. In addition to guiding our own behaviours, we use moral normative structures as reliable and predictable constraints on the behaviours of others.

2. That any information thus retrieved be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory.

People unconsciously abide by moral norms and typically do not question the vast majority of their society's moral norms. Furthermore, most moral norms are treated as objective fact rather than matters of opinion. Moral norms reliably describe and predict people's behaviours. Although moral debate occurs, it is only over a rather small sample of moral norms.

3. That information contained in the resource should be easily accessible as and when required.

Social-wide patterns of moral behaviour in one's group are readily observable in social interactions. Their adherence is predictable and their violations identifiable. The normative structure of our moral world is accessible and stable, such that we can use it as and when it is required.

4. That the information ... has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement.

The vast majority of our moral norms will be learned through conscious experience/witnessing their violation (including in stories and folklore) and their consequences. People feel strongly about the moral norms of their community and

endorse them – as shown by people’s adherence to, and objection to the violation of, moral norms. Of the vast array of moral norms we abide by, disagreement and violations are rare. Moral norms exist and are successful because of this endorsement.

In sum, moral norms satisfy Clark and Chalmers’ criteria for cognitive inclusion and as such are part of our cognitive systems.

I have been arguing that we integrate many aspects of our moral environments into our moral cognitive processing. They become constituent parts of moral cognition. The important point throughout this discussion is that moral cognitive states are not merely the products of our moral environments; they are partly constituted by those environments.

5.6 Questions of impermanence

One sceptical response to the notion of an extended moral cognition may be based some niggling suspicions about the nature of the coupling between organism and environment in the moral case. We might have suspicions because agents and their social environments are *impermanent* couplings; we often come into contact and break off contact with our groups and environs. The response to this suspicion can be understood by returning to the idea of extended phenotypes introduced at the beginning of the chapter. Dawkins’ extended phenotypes were cases in which the organism’s adaptive trait extended beyond the organism. For example, the adaptive ability of spiders to catch flies and other prey is due to the spider *and* its web, not singularly the spider or the web. The *prey catching system* consists of both spider and its web (see discussion in Menary forthcoming-b). Take away either and the system breaks down. The explanation of the adaptive trait (i.e., fly catching) is to be found in the spider-web coupling. The fact that this coupling is impermanent, and in some cases fleeting, should not, as Menary makes clear, detract from including it in our explanation. Spiders are not continuously coupled with their webs but, when they, are they are a part of the fly-catching system. When they are not, they have the *capacity* to be part of a fly-catching system. There is also an additional reason to consider the coupled system as legitimately individuated in this way. The spider and its web, as a single system, has an evolutionary history of its own. The system itself has been shaped by evolution *as a prey catching system*, and selected because of

the adaptive effects of that singular system.

The analogous reasoning in the moral case is that when we are coupled to our external environments in the correct manner we extend our moral cognitive system. When we are not sufficiently coupled to those external resources, we have the *capacity* to become coupled to them. Much of our moral education involves learning to exercise that capacity through the appropriate coupling between brain, body, and world resources. A great deal of moral learning is learning to be a part of our morally structured world and the ways in which we ought to use that world in our adaptive problem-solving routines. And, as in the case of the spider's fly catching system, our extended moral cognitive system has an evolutionary history of its own (we will focus on these aspects in the next chapter). What we can conclude then, is that many of our adaptive problem-solving routines of moral cognising are locationally wide systems that extended beyond the physical limits of the brain and neural system. Those extended elements are, like the spider's web, constituent parts of that system.

5.7 Group interaction and the construction of moral scaffolds

I have touched upon the idea of the construction of moral environments through embodiment effects in section 5.3. I will now further explore the establishment of these environments.

We externalise many of our moral cognitive processes to influence and be influenced by others. We often debate, we argue, we try to convince others and others try to convince us. Moral reasoning is not purely an individual pursuit, nor is it solely the product of our internal psychologies; it is part of a wider system of interpersonal dialogues and agent-world interaction (cf. the constructive process of group recall and extended remembering). In this way, moral reasoning and moral decision making will often be a "communal and collaborative affair" (Clark 2000). At a group level, moral debate is part of the mechanisms of collaborative problem solving. The give and take of reasons is an important aspect of public moral discourse, and is essential to solve many of the problems that face moral groups. Technological advances, for example, bring with them a raft of moral dilemmas: IVF, battery farming, stem cell research, genetic and cognitive enhancement, embryo cloning, legal and illegal drug use, and contraception to name a

few. Environmental factors such as drought and earthquakes may also cause moral problems of resource distribution. A public dialectic is a form of group rationality, or group problem solving, for coordination and cooperation problems.

The expression of moral disagreement stimulates public moral debate. This intra-group disquiet is itself a problem to be solved through repression or negotiation. Modern day acceptance of homosexuality and the rejection of slavery, for example, can be seen as the product of members within groups challenging moral norms and laws prohibiting homosexuality and accepting of slavery. Group moral norms can change when agreements and compromises are made, or majority coalitions form. Moral conclusions are accepted or sometimes tolerated, or sometimes debate continues. Either way, externalising different moral opinions and perspectives helps to bring about moral change and enables adaptive flexibility. Much of the time, deciding what is morally right and wrong, good and bad, just and unjust is a group process rather than an individual exercise. In cases of individual moral decision making, the chosen course of action is often grafted on top of already publicly negotiated moral concepts, norms and beliefs. Public moral negotiation in this way spreads the cognitive load, whereby much of the moral problem solving is done for us through collective group negotiation. Group dynamics are central to individual moralities.

An especially instructive way of understanding this is through the conceptualisation of what Andy Clark (1996, 2000) calls *moral space*. We operate in a world of moral concepts, beliefs, norms and maxims which mediate our preferences, desires, behaviours, and social interactions. They are part of the social space within which we live. Moral space is a region of social space, it is “a space which is intrinsically multi-personal and whose topology is defined largely by the different – but interacting – needs and desires of multiple agents and groups” (Clark 2000:274). Successful navigation of moral space is one of the adaptive problems that confront heterogeneous, dynamically structured and constantly changing groups. The changing needs and desires of group members are sensitive to both internal and external perturbations, such that the topology of such a space changes over time as different problems are brought to bear upon the moral system. Group cooperative success is, in part, due to the successful navigation of moral space.

Much of the exploration of moral space is contingent. What shapes trajectories through

that space will include various historical and situationally specific moral and non-moral factors that are brought to bear on moral systems. The fact that Islamic cultures have moral prohibitions forbidding the consumption of pork is most likely due to historical contingencies rather than any counterfactually robust feature of morality. Cross-cultural moral variation attests to such contingencies. However, there also appear to be commonalities across different cultures (Rozin, Lowery et al. 1999; Vasquez, Keltner et al. 2001). In such cases, our explorations of moral space will often arrive at the same location. These commonalities are best seen as convergent adaptive responses to many similar topographic features of moral space, responses that are present in every society facing similar cooperative and coordination problems, due in part to the particular psychological and physiological dispositions humans have.

There are numerous mechanisms for creating and navigating moral space. We have already seen the role of embodiment in directing moral thought and action and group collaboration. Another important tool for our adventures in moral space is language (Clark 2000). It allows the creation of publicly negotiated moral concepts and principles that create new regions of moral space and new problems to be solved. According to Clark, language affords us the possibilities to “create and explore” this shared social space. Inter-subjective perspective taking is central to the collaborative efforts of moral problem solving, and language provides a “collaborative medium” through which we can give and take multiple perspectives. The linguistic formulation of normative standards in the form of principles and maxims provide tools for collaborative problem solving:

Summary moral rules and maxims act as flexible and negotiable constraints on collaborative action. Such rules and principles by no means exhaustively reflect our moral knowledge, but they are the expertly constructed guides and signposts that make possible the cooperative exploration of moral space.
(Clark 1996:124)

Through the use of moral linguistic terms and labels we can simplify computationally complex social space for perceptually based processes of pattern recognition, thereby making moral space more visible. The recoding of patterns in linguistic form can then

become resources for thought and reason.⁸ Clark draws an analogy with the way in which the concept of money transformed the abstract organisational space of trade and associated reasoning in such a way as to enable the evolution of the complex financial ecologies we see today: “The image is thus of a potent cascade in which culturally acquired tags and labels make available a new quasi-perceptual space in which biologically basic capacities of pattern-recognition can be used to negotiate new and otherwise cognitively invisible realms” (Clark 2000:276-277).

We can think of how the origination of the concept of fairness transformed our social worlds not only through the ways in which we interact with others, but also the way we view that world. It enabled us to categorise actions as fair and unfair, to discuss the fairness of others and create principles to mediate social transactions and resource distribution. These concepts and principles become part of the creation of social institutions. They alter the ways in which we organise our social, political and cultural worlds and create new problems to be solved (such as norm enforcement) and new regions of moral space to explore, through the development of associated concepts of equality and equity for example.

The relevant point with respect to extended moral cognition is the way in which socially distributed, public linguistic structures create moral space and moral structures (concepts, norms, principles, systems of punishments etc) that guide social interaction. These enable us to access and identify “otherwise invisible” regions of moral space (Clark 2000). We actively participate in these processes and moral cognition is intimately coupled to, and supported by, this web of socially-constructed moral space. Much of our moral cognising cannot therefore be explained via a reductive explanation of the purely internal cognitive mechanisms that reside in our heads. To understand moral cognition is, like Simon’s ant walking on a pebbled beach, to account for the dynamic interplay between brain, body, and world. Much of moral cognitive complexity is physically external complexity that arises through the interaction between agent and the socially constructed moral space within which they are embedded.

To further understand the navigation of moral space as group moral problem solving, it

8. See also Dan Dennett: “Once we have created labels and acquired the habit of attaching them to experienced circumstances, we have created a new class of objects that can themselves become the objects of all the pattern-recognition machinery, association-building machinery, and so forth.” (1996:151).

is worth drawing some parallels with Ed Hutchins's work on the nature of socially distributed cognition (Hutchins 1995; Clark 1997). Hutchins's research explores the ways in which many cognitive processes operate across multiple agents and/or our physical environments. His most famous account of distributed cognition is the way in which naval ships traditionally navigated their way into port. Ship navigation is a cognitive enterprise, but importantly one that involves multiple agents (brain and body) as well as artefacts and resources external to those agents. Navigation emerges from this "heterogeneous assembly"⁹ and the particular ways in which they are structured and the interactions between them. At any one time, no single agent on the ship is fully aware of the entire process by which the ship navigates. The act of navigation cannot be attributed to any one agent within the process. The ways in which the social structure of the ship is organised, and the relationships between agents, play an essential role in successful navigation. Ship navigation in this way is a distributed cognitive process, one that can only be apportioned to the entire system itself (Hutchins 1995). The augmenting and structuring of individual cognitive processes through the use of multiple agents, artefacts and their arrangements gives rise to new problems solving skills, such that groups of individual "pattern completing brains [can] navigate the unfriendly and mathematically demanding seas" (Clark 1997:77).

Looking at moral rules and maxims from this perspective, we can see many are the product of collaborative group rationality rather than the aggregation of internally structured parametric representations of individual moral agents (re the Hauser-Mikhail and evolutionary psychology model of the mind and culture); they are in fact *strategic* group solutions to the collaborative and cooperative problems which groups face. They are dynamic and contextualised solutions to various contingent adaptive problems, not a set of innate formulaic principles through which culture inputs various parameters to arrive at our respective output solutions. The collective creation of moral rules and maxims can then become a cognitive resource individual moral agents can exploit. This is not to claim that individual moral rules and maxims are entirely the product of group-level interaction. Individual decisions are not automatically reflective of group decisions, and individual interests do not always reflect those of the group. The picture is one where groups and the interests (moral or otherwise) of individuals will sometimes conflict; group norms will be challenged and negotiated. Individual moral minds are the

9. Clark's terminology (1997:77).

product of the dynamic interaction between individuals and the groups to which they belong.

As noted before, the fact that cultures appear to have similar moral rules and maxims need not require one to posit innate moral structures in the head. We can look upon universality as cultures arriving at similar solutions in navigating moral space. Recalling Clark's financial analogy explained above, one would be hard pushed to find a culture absent of trading systems and associated concepts, beliefs, norms and behaviours of trade. In more advanced form, money and associated financial ecologies are the product of commonly found solutions to recurring resource distribution problems. Yet we would not presume that the structure of these systems and the determinants of agents' behaviours within them are due to 'innate trading or financial parameters' or something similar.¹⁰ It may well be the case that the disposition to value is innate and forms the basis of trade, monetary and financial systems, but that is quite different from proposing that the underlying structure of those systems is innate. Morality is similar in that the socially constructed moral ecology that we occupy is structured around embodied responses such as sympathy and empathy, and solutions to cooperation and coordination problems.

So far we have seen some of the important ways in which we externalise our moral cognitive processes in the adaptive problem-solving routines that we employ in the navigation of our moral worlds. Through the active engagement of our bodily structures and the moral structures in our environments, we transform the cognitive strategies that we employ, thus enabling us to become the moral agents we are and to successfully navigate our moral worlds as both individuals and groups. We have also seen how this real-time navigation is very much enabled by, and grounded in, embodied processes. The ways in which moral groups are structured and the various solutions to cooperative problems that they find have important fitness effects. Throughout phylogeny and ontogeny we construct the social supports and incorporate the external structures required for moral cognitive extension. This process occurs over generations and is subject to natural selection. By constructing the moral worlds of our own, and future generations, our actions have fitness effects upon those generations. In this way we are said to be moral niche constructors.

10. Although Cosmides and Tooby (1992) do suggest we have innate exchange mechanisms.

5.8 A note on the moral niche

By altering the moral structure of one's environment we affect not only present group members but also future generations of group members. Moral debate therefore changes the moral minds of future generations as well as our own by structuring their moral worlds and providing group solutions to many adaptive problems. This reduces the amount of moral deliberation and debate that would be otherwise required of individuals and groups, both within and over generations. If we are born to a world that has already solved many moral problems, we do not have to reinvent the moral wheel. We can make use of the moral information that is already present in our social environments. In this way solutions to many moral problems are distributed over both space and time.

Niche construction is the process by which organisms alter their own environment in such a way as to alter the fitness landscape of that environment (Odling-Smee, Laland et al. 2003; Sterelny 2003). Engineering habitats in this way may be physical (such as termites building their mounds), social (such as social organisation in the case of division of labour), or epistemic (such as the manner in which we organise and store information in books). Niche construction often has downstream effects; the niche construction of one generation has effects upon the next. The classic, and probably most cited example, is that of the beaver's dam. By creating a dam the beaver engineers not only its own environment but also the environment of its offspring, thus directly affecting the survival of future generations of beavers. Downstream niche construction is sometimes cumulative, where one generation alters the niche it inherits from the previous generation, and in turn the subsequent generation alters that niche and passes it on to future generations, thereby creating agent-environment feedback cycles occurring cumulatively across generations. In this way organisms are active participants in moulding their own evolutionary trajectories.

In the following chapter, I will argue that this evolutionary process characterises our moral worlds. We have the moral cognitive capacities that we do because we are born to the moral niche; in turn, future generations have the moral cognitive capacities they do because they are born to the moral niche that we provide for them. Through the use of embodiment effects and a normatively structured world we pass on moral concepts, norms, social structures and behavioural routines to subsequent generations, providing

them with the resources with which to become embodied and extended moral agents and to help in the creation and maintenance of moral space. By engineering our own moral habitats through externalised moral cognition, we create group structured cognitive scaffolds which subsequently alter the information available and use of that information by future generations. The creation of the moral niche therefore alters the fitness landscape of generation N and $N + 1$.

Importantly, this process is “visible” to natural selection (Sterelny 2003, 2004). In the extended phenotype cases that Dawkins (1982) presents such as the beaver’s dam and spider webs, traits do not stop at the organism boundary nor do their adaptive effects – organisms are not uniquely privileged targets of selection. As we have seen, our moral cognitive traits are not fully specified by the internal, psychological mechanisms that reside inside our heads. Those extended traits are often the bearers of adaptation: the target of natural selection includes dynamic, reciprocally influencing, embodied, cognitive feedback loops extending beyond not only the head, but also the body. Furthermore, the processes by which we, as a species, bootstrap ourselves from non-moral infants to mature moral adults are themselves visible to natural selection. As such, it is not just the construction of the moral niche that is visible to selection, but also the downstream cumulative effects of the moral niche. In the next chapter I will outline how this plays out in the evolution of moral cognition, through a gene-culture (dual inheritance) co-evolutionary process.

It should be noted that one need not accept a cognitively extended picture of moral cognition to accept the coevolutionary dual inheritance account presented in the following chapter. In the extended account of moral cognition I have presented here, both the development and execution of moral cognition are constituted by biological and cultural factors. Under the evolutionary account I will present in the following chapter, both developmental/learning resources and beyond the body cognitive resources are the product of dual inheritance evolutionary mechanisms. However, in a non-extended, non-nativist evolutionary account of moral cognition, only the developmental/learning resources are the product of a dual inheritance evolutionary system. In other words, even if one rejects the extension of moral cognition, a non-nativist account of moral cognition is still largely consistent with the dual inheritance account presented in the following chapter.

5.9 Conclusion

In the case where one uses a pen and paper to do long division, it is not the brain that is doing the long division, it is the entire *system*: brain, body, and external environment (Clark 1997; Giere and Moffatt 2003). The main thread running through this chapter is *how our moral cognitive system extends beyond not only the brain and associated neural systems but also the boundaries of the organism*. Structures external to the brain as well as those external to the organism itself enable cognitive tasks to be off-loaded. They also allow the performance of moral cognitive tasks that would not be possible without those bodily and environmental resources. In many cases it is the entire system that instantiates the moral cognitive process, not merely the internal, neural structures within the head. Moral cognition is often distributed over brain, body and world. What is also central to the above discussion is the ways in which we *actively affect our moral environments*. We debate, we punish, we gossip, we exact moral behaviours. We construct our normative worlds. Through the process of actively shaping and structuring our moral environments, we are also altering the fitness landscapes of those environments.

In summary, I have detailed four aspects of moral cognitive extension:

- 1) Embodiment effects are mechanisms that extend moral cognition through embodied coupling with other agents and our normative worlds.
- 2) Social groups act as distributed cognitive tools for tracking, remembering and controlling the moral behaviours of other agents, thereby extending the scope of our moral abilities.
- 3) Moral norms are public artefacts which we use as action-guiding information stores for our own actions and the control of others' actions.
- 4) Groups are essential mechanisms for creating external cognitive resources.

The extension of our moral cognition provides unique adaptive solutions. The fact that it is distributed amongst multiple agents and the environment should not detract from the

fact that it is a cognitive process. Should these processes be attributed to any non-extended individual agent, we would rightfully consider them a cognitive achievement. It is clear that much of our moral cognitive ability cannot be explained through the singular positing of internal representations of external states of affairs (“causal surrogates for distal features of the environment” (Chemero and Silberstein 2008)). Our best explanations pay close attention to the interactions between brain, body and world. Moral cognition is extended cognition.

5.10 Summary

In this chapter I provided evidence in support of the view that moral cognition extends beyond the boundaries of our bodies. Based on this fact, explanations of moral cognition must include the ways in which external cognitive resources operate as part of our moral cognitive system.

The main points to take from the above discussion are as follows:

- Much of moral cognition involves the tracking and controlling of other agents.
- Embodiment plays a special role in coupling agents with the normative world and other agents.
- External cognitive resources structure and shape moral cognition.
- Moral norms are public artefacts, storing moral information and reducing much of our cognitive load by doing some aspects of our moral problem solving.
- Much of moral cognition is distributed over brain, body and world, including the moral scaffolds provided by groups.
- Extended moral cognition creates an extended moral space (moral niche) which has fitness bearing effects.

As in the previous chapter, this chapter details the width of the moral cognitive system. It is extended into our environments. Those cognitive extensions are visible to natural selection. In the following chapter I explore the role of our moral environments in the evolution of moral cognition.

6

The Gene-Culture Coevolution of Moral Cognition

Hominids inherit more than genes from their parents. They inherit both information and developmental environments which allow that information to be used.

(Sterelny 2003:239)

They have shown that these intellectual achievements arise not only from the young mind's surprising capacities for inducing understanding from everyday observation, but also from the ways that caregivers scaffold understanding through the structure of daily routines or by how they talk with the child about recent events.

(Thompson, Laible et al. 2003:138)

Evolutionary ethics has typically neglected the role of culture (Mesoudi and Danielson 2008) often, one suspects, because it is viewed as playing a proximate role in the evolution of human behaviours. Yet such a view misses the novel contributions that culture makes to our evolutionary histories (Sterelny 2003; Boyd and Richerson 2004). We have already seen that our social environments play a central role in the development and deployment of moral cognitive abilities. One of the challenges of an evolutionary history of moral cognition is to appropriately account for the role of culture and fit it into an evolutionary point of view. In what follows I will present a dual inheritance model of

the evolution of moral cognition. Such an account emphasises the ways in which culture plays a unique evolutionary role by altering the fitness landscapes upon which natural selection acts. What we see in dual inheritance models is that not only does biology guide culture but, reciprocally, culture guides biology. In this way our cultural practices can radically alter our evolutionary trajectories. This is, I maintain, what has happened in the case of the evolution of moral cognition. I first discuss the nature of dual inheritance models before looking at how the evolution of moral cognition operates under such a model.

6.1 The making of moral phenotypes

The received view in much of evolutionary biology is that, while genes and environment combine to produce the phenotype, it is only the genotype that is inherited by the next generation; there “is no evolutionary consequential causal influence of N generation phenotypes on generation $N + 1$ ” (Sterelny 2007:179). The mechanisms of inheritance are, therefore, genetic. Spiders’ webs are expressions of the spiders’ web-making genes in web-enabling environments, both in the ontogenetic and synchronic sense. Webs are the phenotypic effects of genes coded for web making and require both the right ontogenic environments as well as the right environments in which to build their webs. In the case of the spider and its web, even though the environments are causally necessary for phenotype development and expression, it is the genes that code for web making that are inherited by the next generation, whereas the environmental constituents are not. Spider genes are inherited, spider environments are not.

As we saw in chapter 3, both the individualist moral grammarians and evolutionary psychologists believe that the *fundamental* causal determinants of our (moral) phenotypes and the underlying structure of our (moral) cultures are genetic. This fundamental structure is universal to all (normal) human beings and exists independent of one’s environment and experience, hence the search for universal features of ‘human nature’ that apparently signify the existence and shape of such structures (Barkow, Cosmides et al. 1992; Tooby and Cosmides 1989; Hauser 2006b). We saw in chapter 2 that one argument in support of internal, genetically specified moral structure is the claim that the informational resources required for normal moral development (of the moral phenotype) are impoverished. The required information must therefore be built into the

genome. Innate moral structures are necessary, but not sufficient, for the establishment of the moral phenotype. However, as I have attempted to show in earlier chapters, an innate moral faculty is not necessary for moral cognition.

In evolutionary terms, to talk of our moral cognition is to talk of moral phenotypes. Phenotypes are the product of the *interaction* between genes and environment. This is, of course, trivially true and no one denies that non-genetic factors play *some* role in phenotypic development – you cannot grow bones without a reliable source of calcium. But on a stronger reading of this claim, developmental environments play a more causally significant role in the expression of particular phenotypes, and this applies to both cognitive and behavioural phenotypes as much as it does to those of body morphology. Subsequently, there is much debate about the role, extent of influence, and effects of developmental environments in phenotype expression. Much of the debate, for example, between evolutionary psychologists and their opponents has been over the contributory role and extent to which developmental environments play in the expression of phenotypic traits (Barkow, Cosmides et al. 1992; Gangestad and Simpson 2007).

In contrast to nativist views, I proposed that moral cognition is characterised by the coordinated integration of cognitive processes across numerous and disparate domains. It often involves problem-solving routines utilising bodily and environmental resources through brain-body-world feedback loops. Coordinated patterns emerge from the integration of coupled elements that span brain, body and environment. Our moral phenotypes are extended and involve the utilisation and manipulation of these elements in a reciprocal feedback process. We affect the world and the world affects us, and it is through this interaction that we instantiate the moral processes that we do. Viewing moral cognition in this light, moral phenotypes are best thought of as *embodied and extended complex skills* that allow us to actively partake in, use (including cognitively integrate) and contribute to, the structure of our moral worlds. Because these skills often involve the coordinated use of brain-body-world interactions, we should consider those bodily and physically external aspects as part of the moral extended phenotype. This includes seeing the distributed cognitive processes presented in the previous chapter as part of our moral phenotype.¹ Moral phenotypes are embodied and extended complex

1. Although the inclusion of group phenomena as part of one's phenotype may sound rather startling, it is no more so than the view that termite mounds are the collective phenotypic expression of a colony of termites (Dawkins 1982).

skill sets, and I conjecture, they are skill sets that we can learn. They emerge over ontogeny from pre-existing (or at least simultaneously developing) non-moral cognitive processes and become increasingly cognitively sophisticated over the course of development.

These complex skills are bootstrapped from constituent processes via the appropriate structuring of our learning environments. Such scaffolds, as we have seen, allow both diachronic and synchronic expression of the moral phenotype. This is where our developmental environments play a crucial causal role in the fundamental structuring of our moral cognitive processes. Such a view does not privilege an internal fundamental structure with respect to the ontogeny of moral phenotypes, nor their evolution. What is important from an evolutionary point of view is that although selective mechanisms depend on the existence of heritable similarity between parent and offspring, they do not depend on, and are not sensitive to, the mechanism of inheritance. Here we can draw on some work from developmental systems theory, underwritten by the ‘parity thesis’, which states that “the roles played by the many causal factors in development do not fall neatly into two kinds, one exclusively played by DNA elements the other exclusively played by non-DNA elements” (Griffiths and Gray 2005:420). When we think of the causal roles of various elements that establish phenotypes, we see they do not map neatly into genetic and non-genetic elements. A relevant example of this, and one which we will encounter below, are the ways in which evolutionary hereditary systems exist beyond those of genetic based systems. This is in contrast to the standard genecentric view which takes genes as the sole inheritance mechanism, whereby genes map with inheritance and non-genetic factors map with non-heritable processes. According to the epigenetic view of inheritance, this mapping does not hold. Organisms are embedded in heritable environments that are themselves subject to evolutionary pressures. Therefore, it does not make sense to privilege genetic inheritance mechanisms as the only heredity systems in the evolution, development and expression of phenotypes.

We can see the importance of this view when we consider the role of the environment in the development and execution of moral cognition. What is clear from the discussions of extended moral cognition undertaken in the previous chapter is that we actively make our moral worlds and in doing so construct the normative and informational structures of those worlds. The structuring of our offspring provides the constituent resources

required for development and expression of the embodied and extended moral phenotype. Aspects of our moral phenotype are both genetically and culturally inherited and it is the structuring of those environments which allows the intergenerational transmission of moral phenotypes. This raises important issues with respect to the precise mechanisms of inheritance and the accuracy of transmission of those phenotypes which I will address below.

In earlier chapters we have seen how the evolutionary psychologist views the *fundamental* structuring of cognitive phenotypes and their aggregative cultural manifestation as coming from the inside in the form of innate (genetic), domain-specific cognitive modules. The channel of inheritance for this fundamental structure is genetic. However, in the moral case this fundamental structure can be learned. Because moral cognition is a learned adaptation, we need to be able to explain both genetic and cultural aspects from an evolutionary point of view. The relationship between genetic and cultural factors in the evolution of moral cognition is best captured by a dual inheritance (or gene-culture) model.

6.2 Dual inheritance models in evolution

As Boyd and Richerson point out in their book *Not by Genes Alone* (2004), standard approaches to evolutionary social sciences have typically treated culture the same as other environmental influences. In much the same way that calcium is required for the genetically encoded growth of bones, culture is required for the genetically encoded expression of various behaviours. From this perspective, natural selection acting upon genes is the *ultimate* cause of behaviour whereas other non-genetic elements merely implement the structures encoded into those genes, and in this way play a *proximate* role. We saw exactly this role afforded to culture in the cases of innate moral grammar and the evolutionary psychology discussed in previous chapters. But, as we will see, selection acting on culture often plays an ultimate role in the evolution of moral cognition.

It is clear that genes have shaped our culture and environment. One of the reasons supermarkets exist is because they provide for many of our basic biological needs. Yet

what Boyd and Richerson and others² make clear is that culture is not like other aspects of the environment in some fundamental ways. Most notably, much of culture involves mechanism of both learning and inheritance. Under certain conditions culture evolves cumulatively subject to the forces of natural selection. Culture affects fitness (see dairy farming example below), and in this way selection operating upon cultural practices can act as an *ultimate* cause of behaviour, rather than merely a proximate cause. Cultural changes can cause genetic changes. Put together, we have a *coevolutionary* dynamic whereby genes and culture exert reciprocal influences on each other that are visible to natural selection in both the genetic and cultural channels. Dual inheritance models along these lines have been gaining traction in the current literature on the evolution of psychological, behavioural and cultural phenomena. For example, Boyd and Richerson (2004) have given accounts of how many of our complex cultural phenomena, from kayak making to the emergence and sustenance of human ultra-society, are best explained via dual inheritance evolutionary models. Kim Sterelny (2003) has presented an account of how our folk psychological capabilities are best explained via a dual inheritance model. And Stephan Linquist (2007) has recently proposed that the evolution of the complex higher emotions is also suitable for such a model.

These models show how cultural adaptations can have selective effects on genetic lines of inheritance in such a way as to fundamentally alter evolutionary trajectories. The classic and often cited example is the advent of dairy farming and the spread of genes for lactose absorption (Boyd and Richerson 2004). After weaning, almost all mammalian animals lose the ability to absorb the lactose in milk when their bodies switch off the gene for the production of lactase, an enzyme necessary for the digestion of lactose in milk. Without this enzyme, lactose remains in the milk and is subsequently fermented by bacteria as it passes through the gut thereby producing gas and leading to stomach irritation, cramping, bloating and flatulence. Significantly, very high levels of lactose intolerance are observed in cultures that do not consume fresh milk. Cultures that traditionally practice dairy farming and consume fresh milk on the other hand have very high levels of lactose tolerance in their populations. The ability to absorb lactose beyond weaning evolved in those cultures that practiced dairying. This is a clear example of genetic adaptations evolving in response to changes in cultural environment.

2. See for example Bowles and Gintis (2003), Odling-Smee, Laland et al. (2003), and Sterelny (2003, 2006).

There are some important points to note from this example.

- 1) *Cultures alter genes.* The cultural practice of dairy farming alters the fitness landscape. Thus there is selection for genes that code for the enzymes that breakdown lactose, increasing those genes' fitness relative to the rest of the population and causing the spread of lactose absorption genes.
- 2) *Genes alter culture.* When the gene favouring lactose absorption spreads in the population that in turn gives rise to selection pressures for increased dairying practices.
- 3) *The two work together in a coevolutionary process.* Evolution in one dimension (genes) influences the evolution of the other dimension (culture) and vice versa. Boyd and Richerson colourfully describe this as a process by which the “evolving pools of cultural and genetic information carried by human populations are partners in a ... swirling waltz.... Each partner of the coevolutionary dance influences the evolutionary dynamics of the other” (2004:192-193).
- 4) *Cultural inheritance is cumulative and subject to selection.* It is not the case that the cultural traditions of dairy farming could be recreated anew in each population or could appear de novo. The traditions are complex and arise through the cumulative modification over generations. Successful alterations are selected and kept, whereas unsuccessful alterations are discarded.
- 5) *Genes don't cause the resultant dairy farming behaviour.* The gene itself does not *cause* dairy farming behaviour; genes for production of lactase that enable the absorption of lactose are not genes for dairy farming (Laland and Brown 2006). But, they do have consequences that induce evolutionary change at the cultural level.³

3. It can be said that genes for lactase increase the probability that the bearer will dairy farm. Although Richard Dawkins (1989) would perhaps call these “genes for farming”, the distinction is merely terminological rather than substantive. Genes for lactase do not cause dairy farming.

Having the gene for lactose absorption increases the relative fitness of agents with that gene within milk consuming populations. Groups that undertake dairy farming thereby increase the spread of that gene through that population. Culture influences genes and genes influence culture. The spreading of the lactose absorption gene “in turn may have changed the environment-shaping cultural practices, perhaps favoring more whole-milk consumption, or more serendipitously, giving rise to the evolution of ice cream” (Boyd and Richerson 2004:192). The genetic makeup of the population is altered in response to the culture of that population, and the culture of that population alters in response to its genetic makeup.

Dual inheritance theories such as these model both genetic channels as well as cultural channels of inheritance. They do this by tracking changes in both the genetic information present in gene pool and the cultural information present in cultural pool, as well as the interactions between the two (Boyd and Richerson 2004; Mesoudi and Laland 2007). One important aspect of this model is that the cultural channel can be cumulative, whereby modifications can be retained over time allowing for gradual modification and selection to occur in a similar manner to that occurring in the genetic channel. This type of model can explain how we come to have complex cultural, habitat-specific adaptations, such as dairy farming, that can arise more rapidly than could otherwise occur in the case of natural selection operating at the genetic level alone (Boyd and Richerson 2004). Cultural inheritance allows rapid adaptive response because it can cumulatively preserve advantageous changes and reject non-advantageous ones over very short timescales relative to genetic evolution. In these types of models, social learning mechanisms, such as copying the successful, operate as low cost alternatives to individual trial and error learning. They allow the rapid transmission and retention of complex adaptive behaviours within one or a few generations rather than over multitudes of generations. Hence, one of the perceived advantages of dual inheritance models is that it allows us to model the evolution of complex skills without having to commit to the view that those skills are encoded in the genome (such is the claim of evolutionary psychology). Cumulative cultural change involving domain-general learning mechanisms is adaptive “because it can do things that genes cannot do for themselves” (Boyd and Richerson 2004:145).

Viewing culture as incorporating an evolutionary efficacious inheritance mechanism

raises specific issues about how these complex skills are passed on: we need to explain the reliability and fidelity of their transmission. One of the ways these issues of inheritance can be resolved is through niche construction. Especially in the case of hominids, downstream niche construction provides a robust, high fidelity, and reliable channel for the inheritance of cultural information.

6.3 Niche construction

As Kim Sterelny informs us, “hominids are ecological engineers with a vengeance...” (Sterelny 2003:149). We physically alter and structure the ecological environments which we inhabit. Importantly, by altering and structuring our own environments we are also altering and structuring those of our offspring. This environmental engineering will often have fitness effects and, accordingly, will often have fitness effects on subsequent generations. Building a warm and dry shelter that successfully buffers the occupants from the negative effects of the environment increases the relative fitness of those occupants. Should the shelter be occupied by the next generation then they will also benefit from that altered environment. We alter our environments and those of our offspring in ways that have fitness effects on both generations N and $N + 1$.

Ecological engineering of our physical, social, and epistemic environments is niche construction (Odling-Smee, Laland et al. 2003; Sterelny 2003, 2008). Animals change their environments in many different ways. Often these changes will be in a manner which regulates the fluctuations and variability of environmental conditions that bear upon the animal. In this way the animal alters its environment in response to external pressures, another example of an organism-environment feedback process. Those altered environments in turn have selective effects which provide feedback to an organism's niche constructing behaviour. This behaviour then feeds back to bear upon the construction of the niche, and so on (Odling-Smee, Laland et al. 2003; Laland and Brown 2006). When this feedback process runs across generations, the construction of the niche becomes the construction of the niche of downstream generations. When this process allows the accumulation of modifications over generations, it becomes an evolutionarily significant process.⁴

4. This is not to claim that all instances of niche construction are fitness enhancing. But those that are fitness enhancing will be selected for whereas those that are not fitness enhancing typically will not.

6.3.1 Downstream niche construction

As Odling-Smee, Laland and Feldman (Odling-Smee, Laland et al. 2003) point out, the picture of organism-environment feedback processes operating over generations is in contrast to many conceptions of natural selection that see lineages as adapting to the environments that they inhabit, whereby selection shapes lineages to fit the niche. Rather, downstream niche construction sees lineages as altering their environments in response to external pressures. The arrow of change therefore flows both ways; lineages adapt their environments as opposed to only adapting to their environments. If this is the case, then we should expect to see, in addition to adaptations that are responses to the environment, adaptations (physical, behavioural and social) that bring about change in the environment. This type of niche construction is seen in such cases as termite mounds and beavers' dams where generation N construct and alter the selective environments of generation $N + 1$.

This model of evolution is very important with respect to the evolution of cognition because not only are we engineers of our physical worlds, but also our epistemic worlds and those of our offspring. Culture is a rich source of epistemic information required for cognitive development, so cultural niche construction is relevant for a dual inheritance model of the evolution of some of our cognitive phenotypes. The cumulative engineering of our cultural worlds works as a system of inheritance in its own right that operates as one of the channels in a dual inheritance evolutionary mechanism.⁵

There are, however, some important constraints as to when niche construction can operate as an intergenerational inheritance mechanism. Complex adaptations appear over time through the selection of accumulated modifications over successive generations. Genes do this very well (Dawkins 1989). Cultural transmission on the other hand is subject to a variety of error-inducing effects such as noise and bias which make sustained intergenerational transmission over time difficult. For this to happen successfully there

5. There is some debate as how to classify niche construction. Rather than an aspect of cultural inheritance, Odling-Smee, Laland and Feldman (2003) view it as a third-channel of inheritance, alongside genetic and cultural channels. This is because niche construction need not work by learning in the $N + 1$ generation. It is unclear whether this is a substantive distinction or merely a terminological one. Either way, this distinction need not concern us here.

needs to be some means of sustaining reliable, high-fidelity transmission between generations. For downstream niche construction to be of any evolutionary significance in the generation of complex cultural adaptations, it too must be a cumulative process and therefore must be able to explain the retention of accumulated modifications over successive generations.

In the moral domain, numerous mechanisms maintain reliability and fidelity and reduce intergeneration information degradation. Downstream niche construction, biased learning and participatory learning all play roles in the transmission of moral phenotypes from one generation to the next. We see these in the structuring of children's moral environments, psychological biases and children's interactions with their moral worlds. Each process plays an important role in the evolution of moral cognition. Here we can start to see connections to the picture of moral cognition as presented in earlier chapters and the evolutionary importance of embodied and extended moral cognitive processes. This picture of moral cognition was one that (often) "looked outward" and involved not only engagement with the world, but the active, embodied and cognitively extended manipulation of the world (including the behaviours of other agents). These process of manipulation structure the moral learning environments of offspring, restricting and directing the moral learning resources available to them.

6.4 A dual inheritance model of the evolution of moral cognition

Products of evolution do not arrive on the scene *de novo* and morality is no different; the evolution of morality is an incremental process. Moral norms, for example, evolve in a series of stages (Harms and Skyrms 2008). This process involves the accurate transmission and retention of modifications in the moral phenotype over time. Under the right conditions, successful changes in moral phenotypes can be transmitted to, and preserved by, future generations.

Moral phenotypes are evolved complex skills that arise via cumulative modifications over generations with components transmitted in both the genetic and cultural channels. Moral behaviours and cultural practices alter fitness landscapes, increasing selection pressures for genes that code for various (non-moral) psychological capacities that enable

the development of moral phenotypes through enculturation. When genes supporting these dispositions spread in the population they in turn increase selection pressures for cultural practices and resources that scaffold the expression of those phenotypes. These include caregiver practices and collective cultural behaviours that support the development and execution of moral cognitions. This feedback process is cumulative and subject to selection. The environment plays an essential role in the development and expression of moral cognitive abilities (moral phenotype) over both phylogeny and ontogeny.

One of the problems with dual inheritance models is that cultural learning environments are noisy, error prone and developmentally variable (Boyd and Richerson 2004; Sterelny 2006). If moral phenotypes are partly constituted by heritable environmental resources, then the transmission of those resources between generations must be accurate, stable and of high-fidelity. Individuals need to have the right moral phenotypes at the right time, and it cannot be solely up to the individual learning to make sure this happens. Because individual learning alone is prone to error in noisy environments, thereby degrading the transmission from N to $N + 1$, there is problem of accurate transmission of moral information between generations.

Within the cultural channel there are different types of transmission within and between generations. The first type is horizontal transmission in which information is transferred between agents within a generation. The second type, vertical transmission, occurs between parents and offspring. The third type is that of oblique transition which occurs between non-relatives over generations (an expert may teach non-related children some particular skill or adaptive information, for example).⁶ All three types of transmission occur in the moral case and there are various mechanisms that operate within the cultural channel that minimise transmission problems, allowing the successful transmission of moral phenotypes from one generation to the next.

In previous chapters we have seen many of these transmission processes in action. Mechanisms of reputation, punishment and inter-group competition, for example, can

6. Sterelny (2006) divides oblique transmission into two types: 'Discrete oblique transmission' occurs when only a few individuals transfer information to many in the next generation. 'Diffuse oblique transmission' occurs when individuals in the subsequent generation are influenced by many in the previous generation. These distinctions need not concern us for the present purposes.

spread and maintain norms of cooperative action within groups. The moral behaviours of group members towards other members can encourage the spread of moral information horizontally between agents of the same generation, as well as over generations. We have also seen vertical transmission occurring where children adopt the moral phenotypes of their parents. Parents teach and structure the child's learning environments such that children can learn how to be moral. Parents also treat their children morally and in this way teach by doing; by acting in moral ways they structure the learning environments of their children. In the case of oblique transmission, moral groups help to structure developmental environments and provide cognitive resources of the next generation. We have seen, for example, how cultural group selection can favour the spread and maintenance of group cooperative norms which also apply to the moral behaviours of new generations. We have also seen downstream moral niche construction in the way that groups create a socially shared moral space which is occupied and utilised by subsequent generations as developmental and cognitive resources.

I will now look more closely at how moral niche construction, participatory learning and learning biases minimise between-generation information degradation. The cumulative engineering of our moral worlds is a heritable process whereby our alterations of the moral world and the practices which bring about those alterations are passed on to future generations. Our (embodied) psychology also biases the information to which we attend, strengthening the flow of specific types of information between generations. We also learn to become moral agents through the embodied, active and guided exploration of our moral worlds. It is through these mechanisms, I argue, that group constructed moral information (developmental environments and structures that realise moral cognition, such as moral norms, concepts, beliefs, moral practices, parenting practices) is passed on to future generations.⁷ I will examine how these processes fit together in a feedback process operating cumulatively over generations which is subject to selection.

7. There are various types of moral information that is transmitted between generations. These include, for example, specific information about how to act in the moral world such as moral norms and prohibitions. It also includes the moral world itself and associated developmental environments. There is much overlap between the various types. For ease of presentation, I will generically refer to the various types of transmitted information as "information".

6.4.1 Moral niche construction, biased moral learning and participatory moral learning

One of the significant points of the preceding chapters was that we are active and embodied participants in our moral worlds and that the patterns of behaviours that constitute our morally normative worlds exist *because* we are active and embodied participants. Humans create a socially constructed moral environment (i.e., moral niche) and use resources of that environment in ways that are central to our adaptive successes. Importantly, this moral environment supports both the development and execution of moral cognitive processes, operating as a mechanism by which to transfer moral phenotypes from one generation to the next. Children receive both explicit and implicit instruction from their groups for moral development and how to use the moral world as a moral cognitive resource. Generation N 's moral activities engineer the environments of their offspring in generation $N + 1$, thereby directing and structuring the transmission of moral phenotype between generations – downstream moral niche construction in action.

As we have seen in chapter 2, there is empirical evidence detailing the ways in which parents structure the child's learning environments and the effects this has on moral development. These were examples of vertical transmission between parent and offspring. Furthermore, non-related individuals also contribute to the structuring of the child's learning environment and directing their development when caregiving duties may be shared amongst non-related adults – an example of oblique transmission. Groups also transfer moral information horizontally within generation N , who then transfer that information vertically to $N + 1$. A clear example of this horizontal-to-vertical transmission can be seen in the ways groups provide parenting advice and norms of parenting practice. If parents adopt the moral phenotypes and parenting practices of their group, then they will transfer them to their offspring because those practices are the ones the learning child is most likely to experience during development. This is an example of frequency dependent bias, whereby what we are likely to learn is biased by the commonness of the information that we experience (Boyd and Richerson 2004). We can see this type of information transfer in the group production of moral narratives in stories and mythologies which are relayed to the next generation. Those moral narratives that are not sanctioned by the group are unlikely to be spread amongst members, and hence the next generation.

Furthermore, studies have shown that much of children's moral experience is with peers and siblings (Smetana 1989). Children of generation $N + 1$ can obtain moral information from generation N via horizontal transmission from their peers in generation $N + 1$; parents teach children, who then interact with other children in their group, thereby spreading that information horizontally within $N + 1$. When we also consider the many processes that bring about group conformity of moral behaviours (as discussed in chapter 1), we can see there are numerous ways in which generation N structures the moral environments of generation $N + 1$. Group structures help to mitigate errors in the transmission of moral phenotypes between generations, increasing both reliability and fidelity.

Importantly, it is not only the engineering of our moral worlds that maintain the reliability and high-fidelity of these channels. Children also contribute substantially to the transmission of moral information between generations because the child's psychology biases her learning experiences. If there is one thing that the poverty of the moral stimulus proponents have got right, it is that disembodied non-participatory observers will have inadequate resources to become moral agents (see chapter 2). But, as we have seen, we are not disembodied, non-participatory observers. Not only are our moral learning environments directed and structured, but our psychology biases our experiences of our learning environments, focussing attention towards morally relevant information and making that information more salient and memorable. Moral learning is biased by, and structured around, the embodied reactions of ourselves and others, directing the transmission of relevant information. In this way, our social environments and our embodied psychologies combine in the production and development of the moral phenotype.

The structuring and biasing of our learning environments highlights an important relationship between morality and the learning of morality, namely that we *learn by doing*. Much of what we learn in morality is through participating in morality. We learn that it is wrong to hit others when we hit others or others hit us. We learn fairness and unfairness through the experiencing of fair and unfair situations. Here, generation N acting morally towards generation $N + 1$ serves dual functions, one being that of controlling moral behaviour, but also teaching moral behaviour by highlighting the causes and consequences of moral violations. We have also seen in chapter 4, the idea that

decoupled moral learning involves *learning by virtual-doing* by way of embodied simulations in which we enact our sensory-motor structures as if we were there. These simulations – such as in the case of children’s fairytales – are directed by pedagogical practices which exploit our embodied psychologies. The developing moral agent learns by biased and embodied participation with his highly structured moral environment and the scripting of moral episodes by caregivers.

Here I have outlined some of the mechanisms which allow the reliable, high-fidelity transmission of moral phenotypes from one generation to the next. Moral development is not a case of trial and error individual learning, but a process of social learning. As Boyd and Richerson (2004) are at pains to make clear, social learning dramatically cuts the cost of individual learning by directing the learner and making the learning target more salient (see also Sterelny 2003). Incorrectly interpreting the features of one’s environment can be costly. Learning moral disgust norms purely by trial and error alone can have disastrous effects upon one’s reproductive success, being able to ‘pick up’ the moral phenotypes of one’s social environment is through directed learning is adaptive.

Having outlined the transmission mechanisms (niche construction, biased learning and participatory learning) in the gene-culture coevolution of moral cognition involving, I will now look at some of the ways these mechanisms are employed.

6.4.2 Moral culture alters psychology

In chapter 2, I demonstrated that our moral environments are not impoverished and that we have the environmental and individual resources to learn to identify the moral domain from the conventional domain. Emotional capacities, theory of mind, and consequential and counterfactual reasoning, combined with highly structured learning environments, enable the learning child to identify moral from non-moral norms. I also tentatively concluded that in such developmental environments a child can develop basic moral concepts based on the emotional responses to harm and welfare. The ability to be moral can be learned.

If moral phenotypes are learned phenotypes then there need to be mechanisms that enable intergenerational transmission of moral information and environments to enable

the use of that information. There are numerous individual mechanisms that facilitate this process. One class of these involve the ways in which caregiving practices and developmental environments scaffold the developing moral phenotype of the child to enable the reliable, high-fidelity transmission from generation N to $N + 1$.

Importantly, moral development takes time. We are not in possession of many of our more complex moral beliefs or concepts until our teenage years (Smetana 2006). It takes time for individuals to develop moral beliefs and norms that are different from their parents and moral groups, if this occurs at all. It is not often you will find a pre-teen moral vegetarian whose parents are not also vegetarian. There “are surprisingly low correlations between parent’s and children’s preferences (e.g., for food, music) and much more substantial parent-child correlations for values (e.g., political preferences, attitudes to abortion)” (Rozin 1999:218). If these values include moral values, this would suggest that our moral phenotypes (at least in modern western populations) are mainly transmitted vertically from parent to child. This is probably due to long periods of juvenile dependence spent in familial relationships as well as the nature of these relationships – we spend a lot of our time with parents and they are typically the punishers of moral transgressions and the teachers of moral concepts and norms. However, it is doubtful that vertical transmission would be the sole channel of transmission in all populations. In populations where caregiving is communal rather than familial for example, one would expect to see more oblique transmission from groups of caregivers to individual children.

We see evidence of the transmission of both biological and cultural aspects of moral phenotypes in the development of guilt. Young children who have high levels of anxiety are more likely to develop high levels of guilt, if they have been brought up with inductive parenting styles as opposed to authoritarian environments. The emerging guilt levels of children with low anxiety are not affected by which parenting environment they are exposed to (Kochanska 1991; see also discussion in Linquist 2007 with respect to a dual inheritance account of the higher emotions). This is a clear example of how the development and expression of a moral phenotype involves the interplay between both genetic and culturally inherited factors, and its development is structured around embodied dispositions, in this case anxiety.

There are many ways in which generation N engineers and scaffold the moral learning environments of generation $N + 1$. We encountered many of these in previous chapters so I will only briefly address them here:

Moral displays as exemplars – The moral behaviours of parents and caregivers provide exemplars of moral displays and scripting of moral experience for the learning child. Children experience the causal sequences of, and appropriate responses to, moral situations. (See sections 2.4.2 and 2.5.2.)

Parental conflict management and disciplinary practices – Young children are often involved in emotionally charged moral conflicts with siblings, peers and parents. Constructive and reasoned explanation, as opposed to authoritarian disciplinary practices, that appeal to the emotional effects of action during toddler conflict resolution is shown to produce higher levels of moral understanding. In learning environments such as these, “emotion-laden discourse during conflicts ... fosters the development of moral emotions such as guilt” (Dunn 2003:340). (See sections 2.4.2 and 2.5.2.)

Moral and emotional salience of environments – Caregivers’ own moral phenotypes structure their care practices thereby structuring the intergenerational transmission of moral phenotypes. Parents will point out moral features of their environments *because* they are morally and emotionally salient to them, thereby teaching children their own moral phenotypes by highlighting specific aspects of the environment and appropriate responses. (See sections 2.4.2, 2.5 and 4.3.1)

Explicit teaching and moral narratives – Members of our community direct and explain the narrative of moral performance by ourselves and others. Adults communicate causality and responsibilities of moral events to which children have emotional reactions, helping them develop the appropriate moral responses (Thompson, Meyer et al. 2006). Children’s developmental environments are immersed with fairytales and other stories that are imbued with moral themes, actions and concepts that provide exemplars of moral behaviours and the scripting of moral episodes. (See sections 2.4.2. and 2.5.)

Moral language and moral discourse – Moral language in the developing child’s social environment enable the child to label their moral experiences. Words like ‘naughty’,

‘wrong’ and ‘good’ enable the child to associate feelings, intentions, and causes with the different contexts within which they arise. Caregivers talk to their children in evaluative terms such as “when you helped mommy that was good wasn’t it”, “don’t hurt your brother, that is bad”. Studies have shown that mothers who frequently refer to feelings and moral evaluations in their discussions of moral events raised children who were more likely to express guilt for their actions after doing wrong (Dunn 2003). Moral language and the linguistic articulation of moral maxims provide a medium within which moral agents can engage in moral debate, moral perspective taking, and ultimately collaborative moral problem solving (Clark 1996). Teaching children moral language provides the tools with which to become active participants in their moral worlds. (See section 2.4.2. and 5.6.)

These are just some of the ways in which the development of our moral phenotypes is supported by the moral niche. By relating embodied responses to causes, consequences, behaviours, verbal classifications and social contexts, patterns of behaviours in groups and socialisation practices enable the reliable, accurate, high-fidelity transmission of embodied and extended moral phenotypes between generations.

In addition to moral niche construction, there are many ways in which children contribute to the moral learning situation by biasing the flows of information between generations. Children are active and embodied participants in the learning situation. They have a raft of non-moral psychological and physiological resources that provide the platform for the process of moral elaboration but also bias between generation information transmission, thereby reducing noise and error. Capacities such as emotions, shared intentions, and imitation all affect the information flows between one generation and the next. We saw many of these resources in chapter 2 and have been summarised in table 6.1, along with the contributions they make to the transmission of moral phenotypes.

Table 6.1 Children’s psychological resources that contribute to high fidelity transmission of moral phenotypes

Resource:	Description:	Contribution to transmission of moral phenotypes:
Emotions and Emotional Contagion	Emotional displays are a primary means of communication. From birth children show basic emotional responding and the ability to respond to others’ emotional states (Saarni, Mumme et al. 1998). Often the recognition of others’ emotional states will generate the adoption of that state (emotional contagion) (Hastings, Zahn-Waxler et al. 2006; Eisenberg, Spinrad et al. 2006; Dunn 2003).	The learning child pays particular attention to emotionally (and thereby often morally) salient events which facilitate the joint attention in moral learning.
Shared Intentions	Through the ability to distinguish the directionality of another’s embodied signalling (for example via a mother pointing at a dangerous object), children attain the understanding that emotional episodes are directed at events or objects (Saarni, Mumme et al. 1998).	Children acquire the capacity to interpret emotional situations and assign meaning to particular emotional and moral episodes. This is important in the transmission of moral information to the learning child.
Social referencing and Embodied Meaning	By the end of the first year children are able to pick up on the embodied displays of others to evaluate their environment and regulate their responses appropriately. (Dunn 2003; Saarni, Mumme et al. 1998). Children as young as nine months will use the reactions of others as a guide to behaviour in situations of uncertainty (Saarni, Mumme et al. 1998).	Children start to understand and appreciate the meaning of others’ embodied reactions and use those reactions as resources for appraisal in moral learning.
Development of the Self	Much of children’s moral development is congruent with the development of self-representation and seeing oneself as participants in a social world. This allows the development of “self-referential emotions such as guilt, pride, shame, and embarrassment” (Thompson, Meyer et al. 2006).	Self-representation enables sensitivity to the evaluations of others and the feelings (pride and guilt, for example) that arise from those evaluations. This motivates agents act in accordance with others moral evaluations.
Intersubjective Understanding	At around 2 to 3 years of age children start to gain an awareness that people have divergent mental and emotional states that may differ from their own (Saarni, Mumme et al. 1998).	Intersubjective understanding is important for learning the consequences and motivation of peoples’ actions.
Imitation	Imitation plays a major role in the learning of cultural skills and is central to the evolution of culture as it allows the transmission of cumulative improvements over time (Tomasello and Whiten 1999; Sterelny 2003).	Imitation of moral behaviours and practices supports moral learning and the transmission of moral phenotypes. It also enables changes to the moral phenotype to be retained in the next generation.
Active Exploration	Young children will actively explore and test the normative structures of their moral environments in a form of trial and error learning. We see this in the period of development infamously known as the Terrible Twos.	Exploring one’s moral environment allows the child to experience the moral environment and the consequences of action, gaining moral information about that environment used in moral learning.

Children are not passive moral learners. They are heavily biased towards, and actively explore, particular aspects of a highly structured moral environment which in turn exploits these biases to transfer moral information to the child. This mix of social transmission and individual exploration form what Sterelny calls hybrid-learning (Sterelny 2006). It is through these mechanisms, I maintain, that the learning moral child builds up increasingly complex and elaborated moral knowledge structures.⁸ (See Figure 6.1).

Figure 6.1 Culture and the development of moral phenotypes

8. Note also that this is a multi-stage process where development will be constrained by levels of cognitive sophistication; you can't have concepts of equity without first having concepts of fairness.

In summary, the reliable, high-fidelity transmission of morally relevant information from the moral niche to the developing child enables the development of the moral phenotype. This involves:

- 1) Structured and directed moral learning environments which exploit the developing child's non-moral psychology⁹
- 2) Non-moral psychological biases that direct the learning child towards morally relevant aspects of the moral niche.

6.4.3 Moral Psychology Alters Culture

We have seen above some of the proximate mechanisms that structure and bias the flow of information from moral environments to the developing moral agent – the focus was on those mechanisms that facilitate the intergenerational transmission of moral phenotype from one generation to the next. This was one half of the feedback loop between the agent and their moral niche. The other is that in which the causal arrow flows the other way, where psychology influences and structures the moral niche.

We have already seen in chapter 1 how selection as an ultimate cause can favour the evolution of moral behaviours. Many moral capacities, norms and practices are selectively favoured because they enhance cooperation. Individual selection can operate in favour of moral behaviours in small groups. Group-level cultural evolution was necessary for the evolution of morality in large-scale groups. In each case, our moral worlds exist because of the cooperative benefits they confer. Proximate mechanisms also help determine the structure of our moral worlds, because the makeup and nature of our moral worlds is not independent of our psychology. Because biases play a central role in the transmission of moral information they also affect the structure of our moral worlds. This is a crucial point: our beliefs, desires, choices, preferences and embodiment effects have a direct effect on the transmission of specific cultural variants, which, as we have seen above, are partly constitutive of the moral phenotype.

9. I use the term “non-moral” to denote psychology that is not specific to morality.

Psychology helps determine the salience, memorability and motivations to exhibit moral information, thereby affecting the transmission of moral information between agents. This bears upon the moral phenotypes that develop and, in turn, the patterns of behaviours that give rise to the construction of our moral worlds. Take, for example, moral norms. Those that have strong emotional content, such as protecting children from harm for example, are more salient, memorable and elicit strong motivations; they resonate with our emotional sensibilities. It is more likely they will be transmitted between agents within a population than norms that promote harming children.

Shaun Nichols (2002a) presents a persuasive account of how these effects play a role in the cultural evolution of norms. Through tracking historical patterns of social norms, he highlights just how our affective capacities bias the transmission of cultural information within cultures. According to Nichols, affect backed norms (such as those to do with harm and disgust) have a greater chance of transmission than non-affect backed norms and are therefore more likely to spread within and across generations. Because of their affective nature they are regarded as more serious, more important and easier to recall. This biasing effect is a special case of what Boyd and Richerson (2004) call *content bias*, whereby individuals are motivated to adopt practices which appeal to them or those that are deemed to be most successful or advantageous. Content biases depend on the content of the information transmitted. Our affective responses operate as 'cultural attractors' which bias the evolution of moral norms. As Nichols explains, normal people have aversive responses to suffering in others. Norms that prohibit harming others are therefore more likely to spread through populations over time. Norms that reduce negative affect (or increase positive affect) will be more appealing to the psychologies of normal agents than norms that promote negative affect or are affectively neutral. There are good adaptive reasons why we have affective responses to suffering, and these responses can be harnessed by cultural evolution to promote cooperative behaviour. We construct our moral niche around the embodied responses of human agents.

Here we can see how the psychology of agents influences the construction of the moral niche. Moral cognitive capacities spread amongst populations via cultural group selection because they promote group cooperative behaviours. But our affective responses also create basins towards which cultural evolution will gravitate. The psychologies of agents within a population will therefore have direct effects upon the cultural structure of that

population. This will include the types of moral norms, behaviours and parenting practices which structure the child's moral learning environment. (See Figure 6.2.) In this way, psychological biases help to 'fix' some of the content of moral cultural evolution. Some products of cultural evolution will align more closely with our psychologies than others. When psychological biases are aligned with moral information, basins of attraction stabilise the propensity for that information to remain with the population. Moral solutions to cooperative problems that are aligned with our psychology are more stable than those that are not.

Figure 6.2 Moral phenotypes and their influence on the moral niche

We can see these types of effects in the dairy farming example encountered earlier. Human preferences for the taste of milk provide basins of attraction which norms of dairy farming will gravitate towards. If humans found the taste of milk disgusting it is unlikely that we would have seen the cultural evolution of dairy farming and hence the spread of genes for lactase within dairy farming populations. Taste preferences help stabilise the cultural practices that evolve around them, making them more likely to be retained in a population as well as be passed on between generations. Similar effects are, I maintain, at play in the case of moral evolution.

6.4.4 Putting the Feedback Loop Together Again

So far, we have seen how the construction of the moral phenotype and transmission between generations involves multiple mechanisms. The individual develops into a moral agent because of his (not exclusively moral) psychology and moral niche to which he is born. A combination of scaffolded learning environments interacting with biased psychologies facilitates the reliable, high-fidelity transmission of the moral phenotype (complex skill) from one generation to the next. Cultural evolution utilises, and is influenced by, many of the biases inherent in our embodied psychologies to establish groups of cooperative agents. The moral niche itself exists because of the interactions of multiple moral agents. Putting these aspects together, we have a stable feedback loop occurring over generations that is subject to selection at the group level as well as the individual level. Selection at the level of the individual favours those capacities that allow the successful cultural elaboration of an embodied psychology that underwrites the moral phenotype. At the group level, selection occurs for the construction of a structured learning environment (moral niche) that gives rise to the group beneficial moral phenotypes. Hence our moral phenotype is a product of biological as well as cultural selection operating via a dual inheritance evolutionary model. Importantly the two processes impact upon each other. Once we have feedback loops occurring over generations involving high-fidelity transmission and incremental change we have evolutionary efficacious downstream niche construction.

The cultural generation of primitive moral practices that bring about cooperative benefits alter fitness landscapes, leading to increased selection for the non-moral psychological capacities which enable the cultural elaboration of the moral phenotype. This includes capacities such as emotions, theory of mind and imitation. Their spread within the population gives rise to further selection pressures for moral cultural practices (moral niche construction) which leverage greater cooperative benefits. This in turn creates greater selection pressures for the spread of psychological capacities that favour this elaboration, and so on. The two dimensions operate in a coevolutionary process: culturally-mediated changes in phenotype cause changes in selection on genetically-mediated traits, and vice versa. As long as those cumulative improvements are maintained over generations we will see the ‘ratcheting up’ of complex moral skills and

the elaborate moral and social systems that we see today, and the cooperative benefits they bring. This feedback process crossing over brain, body and world is a dual inheritance model of the evolution of moral cognition. In Box 6.1 I have given an account of how this process operates in a gene-culture coevolutionary model of moral disgust.

Box 6.1 Case Study: The Gene-Culture Coevolution of Moral Disgust

Moral disgust is an emotion-based response to a class of morally forbidden behaviours that typically promotes withdrawal of contact and association with perpetrators of those behaviours. Rozin et al. suggest that the role of this moral response is a “powerful form of negative socialisation” (2000:650) and in this way acts as a mechanism that promotes adherence to the social order.

Rozin and colleagues present evidence that moral disgust is culturally elaborated from more basic food related emotional responses and proceeds incrementally through a sequence of four stages involving increasing levels of abstraction. The development of moral disgust starts with the *Core Disgust* originating from more biologically primitive distaste responses to foul tastes. Core disgust is a uniquely human emotional response to protect the body from disease and poison, especially potential contaminants such as food, animals and bodily products. It is not witnessed until after infancy, and initially not considered to involve concepts of ‘contamination’ such as that seen in more mature subjects. Core disgust then proceeds to *Animal Nature* disgust which involves negative responses to violations of the animal-human boundary through to contact and offensive sights. This involves responses to such things as death, sexual acts and inadequate hygiene. Rozin et al. propose that these involve the rejection of things that remind us of our animal nature. The stages of disgust then move to *Interpersonal Disgust*, involving rejection of contact with, and hence potential contamination from, other people and the application of more abstract concepts of contamination. The final stage is that of *Moral Disgust*, which involves responses to abstract categories such as racism, murder and the degrading treatment of others, often pertaining to violations of the social order. Rozin et al. suggest that moral disgust “may represent a more abstract set of concerns about the human-animal distinction, focusing not so much on the human body as on the human body-politic – that is, the human as a member of a cooperating social entity” (Rozin, Haidt et al. 2000:644).

The development through the various stages of disgust involves increasingly sophisticated and abstract ideas of contamination “requiring a separation of appearance and reality” and notions of invisible entities “that are vehicles of contamination” (Rozin, Haidt et al. 2000). Full development of the most abstract of these concepts can only be achieved with the attainment of a level of cognitive sophistication which does not arise until the age of about seven years (in North American children (Rozin, Haidt et al. 2000)). In this way, the development of moral disgust is *cognitively constrained* because it is dependent on the development and integration of various cognitive competencies. Development through these levels is also *incremental*, progressing from simple to complex levels of abstraction.

There is evidence that this expansion process has a cultural basis. Firstly, studies of feral humans found that they do not show disgust responses (Malson, reported in Rozin, Haidt et al. 2000). Secondly, core disgust and animal nature disgust show marked similarities across different cultures, whereas the more abstract higher levels of expansion in the socio-moral categories show greater between-culture differences (Haidt, Rozin et al. 1997; Rozin, Haidt et al. 2000). Thirdly, we often see quite rapid change in moral disgust prohibitions, as seen by, for example, recent changes in moral attitudes towards sex, sexuality and obesity.

This expansion of disgust is developmentally scaffolded by moral environments. The transmission of contamination concepts, norms of behaviour, and associated eliciting cues are socially transmitted. For example, infants do not display aversive reactions to faeces, but Rozin et al. suggest that children pick up these responses from parental behaviour through witnessing acts such as aversive responses to faeces, i.e., social referencing. The screwing up of the face and rejection of offending object are clear embodied signals to the child which involve expressions of inhibiting bodily entry (screwing up of mouth and nose) and expulsion (gape indicating food rejection) (Rozin, Haidt et al. 2000). This is downstream niche construction in action. Parental displays structure and direct the child’s learning environment through embodiment displays, normative action and communication.

Extended periods of infant and juvenile dependency are likely to play a role in the fidelity of the transfer of the moral disgust phenotype. If the learning environment is primarily with the parent then they will most likely adopt the moral phenotype of that parent. By the time children have reached moral maturity they are less likely to interfere with or drastically alter the scaffolding process because they have already adopted the

moral dispositions that underwrite that process (see Linquist (2007) with respect to intergenerational transfer of emotional phenotypes).

Children's basic emotional dispositions for core disgust bias their learning by directing them to emotionally salient aspects of their environment and thereby reinforcing mechanisms of intergenerational transfer. Parenting practices also exploit the psychological dispositions in a child that is already primed for disgust-based learning. In this way disgust based emotions operate as basin of attraction which learning processes will gravitate towards. Children are much more like their parents with respect to values than to preferences (Rozin 1999) and the robustness of this vertical channel of cultural transmissions through biasing effects goes some way to explaining why.

The psychological mechanisms underlying moral disgust norms also operate as cultural attractors for the (horizontal) cultural transmission of those moral norms. Research by Shaun Nichols has shown that norms "prohibiting actions that elicit negative emotions are more likely to survive than affectively neutral norms" (Nichols 2002a:251), and his research confirms this with respect to norms prohibiting actions that elicit core disgust.

On first appearances, this picture seems applicable for our more biologically grounded "core disgust" reactions, but what about more abstract moral disgust? Once the association between emotional based disgust reactions and moral actions is made in development, then those actions will reliably elicit emotionally grounded reactions, thereby biasing and increasing the likelihood of the transmission of disgust norms and behaviours. As mentioned above, those associations are less likely to be broken if they are acquired during development, becoming individually entrenched. Here we can see biased stabilisation effects operating at the level of culture as well as in the feedback loops transferring the moral phenotype from one generation to the next.

Looking at the development and cultural transmission of disgust in this way we can see how the cultural elaboration of disgust fits a dual inheritance evolutionary model. At the genetic level, selection favours distaste based basic emotional mechanisms and other cognitive capacities that allow cultural elaborations of disgust. At the cultural level, selection operates on both group norms within a population as well as associated pedagogical practices (i.e., the moral niche) that give rise to the cognitively enriched disgust based responses.

As such, core disgust is co-opted by culture to extend the avoidance mechanism beyond the initial elicitors of core disgust. There are clear and direct individual adaptive

advantages for many of the stages involving poisons, contamination, and disease and at the more abstract level the “need to reject many things, including sexual and social “deviants”” (Haidt, Rozin et al. 1997).¹⁰ The direct adaptive advantages are not always so obvious. Think, for example, of how people in Western societies typically find the eating of cats morally disgusting whereas eating pigs is not. In cases like these, adaptive value often lies in as a mechanism of group conformity and control as well as advertising difference to other groups. Through the alienation and stigmatisation of deviant individuals, moral disgust operates as a mechanism that protects the social order and maintains large-scale social cooperation. Mechanisms that originate in aid of self-preservation are culturally elaborated to become mechanisms of group coordination and cooperation. In the case of group coordination, the content of the norms themselves may not be adaptive¹¹ but group conformity is. The skill of being able to make judgements of moral disgust is learned, in addition to the content of the norms. In other words, it need not be the content of the disgust norms that is necessarily adaptive, but the complex skill of being able to make disgust based judgements that conform to those of your group that is adaptive.

10. Haidt and Rozin et al. point out that we should be careful in proposing biologically adaptive reasons for the evolution of disgust (such as poisons and disease etc) for the fact that children do not display disgust reactions early on, will often put all sorts of things in their mouths and do not develop ideas of contamination until at least 5 to 7 yrs (Haidt, Rozin et al. 1997). They suggest that this could indicate that the social function of disgust could be the more important adaptive role of disgust. I agree that we should be careful. One reason for late onset could be because long periods of infant and juvenile dependence will mean children have limited exposure to dangerous and infectious objects as long as they are adequately supervised. However, the age of onset should not matter as to whether there are biological advantages to having such knowledge. Biological advantage can also have late onset. It may take a few years for young children to learn that cars are dangerous, but that does not mean it is not a biologically adaptive advantage for them to do so.

11. As Boyd and Richerson make clear, sometimes culture is adaptive, maladaptive, or neutral (Boyd and Richerson 2004).

Before concluding this chapter there is one issue I would like to address. I started this thesis by outlining many of the mechanisms by which moral cognition evolved. One feature was how within-group similarities and between-group differences can be sustained through cultural group selection. In subsequent chapters, we have also seen many of the mechanisms that account for the transfer of within-group similarities between moral agents within and across generations. However, I have said very little about between-group similarities in moral phenotypes.

6.5 In Group / Out Group Similarities and Differences

A culture and learning dependent view of morality leads to expectations of moral difference between communities. The challenge, therefore, is to show how this account is compatible with cross-cultural similarity. In chapter 1 I outlined how cultural group selection stabilises uniform moral behaviours that favour cooperation; within-group uniformity can be an adaptive advantage. We saw that cultural group selection for mechanisms of cooperation only work if there is conformity within groups and variation between groups as is required for cooperative enterprises to evolve in large groups. It is uncontroversial to claim that there are many cultural similarities within large groups and that there are many moral differences between groups. We need only think of moral

norms pertaining to marriage, sexual practices, rights and food for examples of these. However, if morality is not innate, why do we also see *between-group similarities* in divergent populations? If, as Boyd and Richerson maintain, norms of enforcement and punishment can stabilise almost *any* norm, why are there commonalities between moral cultures?

It would appear that all cultures have moral systems as solutions to cooperative problems (to some degree or other), including notions such as fairness and justice. Universality in general is touted as *prima facie* evidence for innateness (Barkow, Cosmides et al. 1992) and this is also true in the moral case (Hauser 2006b). Consequently, there is debate as to whether there are moral universals. As we have seen in earlier chapters, Marc Hauser, John Mikhail and Susan Dwyer believe that we have a universal moral grammar. Jesse Prinz (2008a), on the other hand, maintains that there are numerous counterexamples to oft cited moral universals such as incest prohibitions and universal norms against killing. These counter-examples, Prinz argues, undermine claims of universality and give us some reason to doubt moral innateness. However, as both Hauser and Prinz recognise, evidence of universality is not sufficient to establish or refute nativism, it needs to be presented with acquisition evidence. Moral universals and their role as evidential support for nativism are open questions but not ones that I will focus on here. Instead, I will look at the lesser claim, that there are some moral similarities between some human populations that are more than merely random chance, or obvious selection pressures (such as the fact that there are no cross-cultural norms promoting celibacy). Here I will briefly look at some of the theoretical tools we have for accounting for cross-cultural similarities and the relevance to moral cognition.

We have already seen that learning biases make certain moral information likely to spread. Work by Shaun Nichols (2002b) provides examples of how these biases affect cultural transmission. Further, with similar selection pressures it is likely that different and disparate groups with the same biological drivers and cooperation and coordination problems will often converge on similar solutions.¹² The example of trade and finance from previous chapters can be applied here. Trade and financial systems are cross-cultural phenomena, if not universal. All normal people have at least rudimentary notions of ownership, property, trade, exchange and monetary value. A non-nativist account of

12. See also Dennett's discussion of "a Good Trick" whereby general intelligence can lead "again and again to same bright idea" (1995:487).

these systems and associated concepts and behaviours involves the cultural elaboration of dispositions to desire and value. Human groups face similar resource distribution problems, many of which are solved by the construction of economic systems. The evolution of these systems is biased by our non-economic biologies. An evolutionary account based in biology and culture would see the emergence of trading system such as bartering, leading to more complex systems of money and finance across disparate groups.

In the moral case, we are likely to see similar mechanisms. Feedback cycles that involve the same innate psychological biases and similar cooperation and coordination problems will converge on similar solutions. Physiological and psychological biases are likely to give rise to basins of attraction which will, for example, favour the evolution of particular norms and not others. In this way culturally unrelated groups will converge on many of the same locations in moral space. Some insightful work on the ways convergent solutions arise can be seen in Brian Skyrms' *The Evolution of the Social Contract* (1996). Skyrms applies evolutionary game theory to problems of cooperation and much of this work can be read as showing how evolutionary dynamics will often converge towards particular strategic solutions to cooperative problems, even from different initial conditions. This is not to say that Skyrms claims that there is one true solution to cooperative or coordination problems. On the contrary, he claims, for example, a variety of different norms can arise: "It is apparent ... that the typical case is one in which there is not a unique preordained result, but rather a profusion of possible equilibrium outcomes" (1996:109; see also, Harms and Skyrms 2008). But Skyrms does show that in dynamic interactions, there are stable states which have an increased likelihood of evolving – these are basins of attraction for the evolution of cooperative strategies.

Skyrms considers the problem of fairly dividing a cake between two people. Assuming that cake possession is correlated to fitness in a reproducing population, for randomly paired individuals partaking in symmetric¹³ division of the cake there is one evolutionary stable strategy available, a 50%-50% equal division. An evolutionary stable strategy is "a strategy such that, if most of the members of a population adopt it, there is no "mutant"

13. A game is said to be symmetric when the outcomes are the same for each player when under the same circumstances. i.e., the outcome for player A adopting strategy S, is the same outcome for player B adopting strategy S. The outcomes depend on the strategy being played, not who plays the strategy. Symmetric equilibrium is when all players use the same strategy. e.g., all players asking for 30% of a cake.

strategy that would give higher reproductive fitness” (Maynard Smith and Price 1973:15). These strategies are more likely to persist in populations because they produce greater, or equal, fitness outcomes compared to alternative strategies. In cake dividing games such as this there is one clearly preferential strategy (see Box 6.2 for more detailed explanation). Yet, even in cases where there is no *one* clear preferential strategy, we still find instances of convergent solutions. One example is how, in some circumstances, even very small biases can see the universal establishment of particular norms (Skrms 1996).

In some coordination situations there will be multiple evolutionary stable strategies available. Sometimes the payoffs for these strategies will be equivalent so there is no principled way to choose between the two. In these situations, we are said to be in a position of *symmetry*. Skrms argues that breaking of this symmetry can occur due to small differences between the participating agent’s resources. He shows that the slightest perturbation which alters the cost / benefit of any one player (i.e., induces a bias) will alter the reproductive fitness of that player so that the strategies in a reproducing population will no longer considered to be symmetrical. Perturbation events create a basin of attraction towards which the population will gravitate. Over time, the strategy to which the bias leans will dominate the population and will therefore become the norm.

Box 6.2 Cake dividing games and the evolution of fair division

We can see why a 50%-50% fair division strategy is likely to arise in a population of cake dividers when we look at the other strategies available to cake dividing populations. As Skrms (1996) explains, should a population adopt a strategy of requesting 60% division, then each individual gets nothing because they will only encounter others also demanding 60%. But, if any one individual in the same population were to ask for anything less than 40%, they would have the rest of the population to cut cake with (i.e., all those asking for 60%). They would do better than the average of the population and hence have better than average reproductive fitness.

We can also imagine the same outcomes in a population in which all people ask for 30% of the cake. If any alternative strategy that arises in the population asking for more than 30%, but no greater than 70%, that strategy will do better than the population average.

In each of these examples, what we see is that those with the higher payoffs (i.e., higher fitness) than the average will increase in numbers in the population. Population-wide strategies of 60% or 30% are evolutionary unstable strategies because there are

mutant strategies which have higher fitness. However, a crucial point is that as those mutant strategies spread in the population they become less profitable because they are more likely to meet their own kind. A strategy of asking for 70% in a population of others asking for 30% will initially have higher than average fitness and therefore spread in the population. But, the advantage of this strategy will decrease as their kind becomes more common and therefore increase the probability of encountering 70%ers with whom they cannot divide cake.

This leaves only one strategy which is evolutionary stable; that is, when most of the population adopt that strategy no alternative strategy can better it. That strategy is 50%-50% equal division.¹⁴ In cases where an alternative strategy arises requesting a division less than 50%, they will always receive less than the population average. In cases in which an alternative strategy arises which requests a division of more than 50%, there can be no cake division because any attempted division would be greater than 100%. What we see is that 50%-50% equal division is both an evolutionary stable strategy (i.e., unlikely to be bettered by any other strategy) as well as an attracting equilibrium, whereby reproducing populations of cake dividers will evolve towards a population of 50% cake dividers. Any move away from that strategy would induce selective pressures back towards that equilibrium. Equal division is a basin of attraction in the cake dividing game. Strategies above and below 50% are unlikely to invade the population. Therefore, demanding 50% of the cake is the only evolutionary stable equilibrium strategy and more likely to evolve in populations of reproducing symmetric cake dividers. Skyrms also shows that in bargaining cake division games that involve mixed strategies (i.e., populations have a mixture of different bargaining strategies) 50%-50% equal division is still likely to evolve from larger numbers of initial starting states than any unfair division strategy, because equal division strategies provide large basins of attraction towards which populations will evolve. (See chapter 1 in Skyrms 1996).

A useful application of this line of thinking is Neil Levy's (2004) treatment of gender role

14. Note that a strategy where each member of the population demands 100% of the cake is, like the 50% strategy, an equilibrium strategy (Skyrms 1996). This is because everyone in the population will get nothing, including any mutant strategies that arise, because they are likely to only meet agents undertaking the 100% strategy. Therefore the population is said to be at equilibrium. But, the 100% strategy is not a stable strategy. This is because if only a small number of mutants arise who ask for 50% or less, then they may meet and share the cake with each other, thereby attaining a higher than population average payoff. If this were to occur, the mutant strategy would spread within the population. This is not the case with equal (50%-50%) division, where no other mutant strategy can obtain higher than average payoff.

norms. Child-rearing is a resource allocation problem that has many possible solutions, but there are two evolutionary stable strategies that are in a position of symmetry where each has the same payoff: either the man can care for the child, or the woman can care for the child. In reality, the one solution that we typically see is that the task of child rearing is allocated to women. The reason, Levy suggests, is due to *physical* differences between men and women.¹⁵ Because mothers initially bear the costs of child rearing when the child is born (such those of childbirth and breast-feeding), a naturally occurring asymmetry arises with respect to the labour allocation of child rearing between the sexes. This symmetry-breaking event drives the population towards adopting one coordinating strategy over the others. The asymmetry need only be small for symmetry-breaking to occur. If these symmetry-breaking features are common to independently evolving populations involving the same coordination problems, it is highly likely the same coordination strategy will arise in each of the populations.

This work is relevant to the evolution of morality and cross-cultural moral similarities. Firstly, independent populations can converge (indeed, are likely to converge) upon many of the same solutions to various problems such as resource distribution. We have just seen, for example, how evolution will often lead to selection for equal division from many different starting conditions and strategies. Basins of attraction can draw evolving populations towards certain evolutionarily stable strategies. Fairness is a clear example of this. Secondly, symmetry breakers (which often seem rather innocuous and independent of the phenomena being explained) can drive the adoption of cooperation and coordination strategies in independently evolving populations. Non-moral biases can favour the evolution of particular moral norms. Thirdly, much of the point of Skyrms' work is to show that natural selection operates on cultural dynamics, or as Skyrms himself explains, "the Darwinian story can be transposed into the context of *cultural evolution*, in which imitation and learning may play an important role in the dynamics" (1996:11, italics in original). We can invoke cultural evolutionary dynamics to explain many features of populations without the need for innate, domain-specific psychologies that evolved in the 'Environment of Evolutionary Adaptedness of the Pleistocene'.

Having seen how moral norms of fairness can evolve, we can hypothesise similar

15. As Levy points out, evolutionary psychologists such as Steven Pinker will often claim that it is innate psychological differences between the sexes that evolved in the Environment of Evolutionary Adaptedness which best explain gender roles (see Pinker 2002).

histories for norms of justice and associated concepts. Concepts such as good, bad, right and wrong, and moral prohibitions against cheating and theft, for example, are likely to evolve in divergent populations because they are evolutionary stable strategies to cooperation and coordination problems.¹⁶ We can also imagine Levy's norms of gender roles applying to the morals of sex and sexuality often witnessed across cultures, such as non-equivalent moral judgements and norms applying to men's and women's sexual promiscuities and associated reputations. In the moral case more generally, it is likely that biases, even small ones, can result in convergent evolution in divergent populations where those biases are enough to establish basins of attraction for norms of behaviour. A universal set of basic non-moral emotions, for example, can trigger basins of attraction for the evolution of norms as well as pedagogical practices that spread those norms. Emotions can therefore operate as symmetry breakers which favour some particular moral strategies over others. Human trajectories through moral space are going to be constrained by the effects of our underlying embodied psychologies, such that the trajectories that groups independently travel will be very similar indeed. Many of our moral norms and cultural and moral pedagogical practices will be convergent, culturally evolved solutions to cooperation problems. We need not, therefore, appeal to an innate universal moral psychology to explain the existence of cross-cultural similarities in moral psychology or moral systems.

6.6 Conclusion

I have contended that moral phenotypes are learned phenotypes. They are complex skills that we learn over the course of development which emerge from the patterned integration of aspects of brain, body and world – they involve both biological and cultural aspects. These skills cannot be learned individually because individual learning environments are typically noisy and prone to error. The learning and expression of moral phenotypes requires socially structured interactions with our moral worlds – both the development and expression are scaffolded and structured by our moral environments. This places particular importance on the role of the environment in establishing moral phenotypes.

16. This is not to claim that *all* morality evolves. There are surely other mechanisms at work, but the claim is that much of morality evolves this way.

These moral environments are themselves the product of adaptive cumulative change over generations. In addition to the biological components of moral phenotypes being inherited, cultural structures that enable moral phenotypes are also inherited. Together these operate as part of a dual inheritance system – aspects of the phenotypes are inherited in both the genetic and cultural channels. Because cumulative evolution of moral phenotypes requires the reliable, high-fidelity intergenerational transmission of moral information, there are some issues around how these phenotypes are inherited between generations. I have detailed a number of mechanisms that minimise transmission errors in the cultural channel. These allow the transmission of group moral phenotypes from one generation to the next. We have also seen that this model is coevolutionary. Evolution in one dimension (genes) influences evolution in the other (culture) and vice versa. Psychology begets culture and culture begets psychology in a gene-culture feedback loop (the psychological/genetic component which does not, as I have stressed, constitute an innate moral sense). In this way, culture is an evolutionary efficacious mechanism – selection acting on and through culture guides our evolutionary trajectories.

In sum, parents direct the social learning of the child. Children also contribute to the learning experience by actively engaging with their moral worlds. Both mechanisms improve fidelity of transmission in potentially noisy and error prone environments, allowing the cumulative retention of variation necessary for natural selection. The transmission of moral cultural practices and environments over generations allows the cumulative evolution of moral cognition.

Some of what I have outlined may be speculative, but it is none-the-less empirically tractable.¹⁷ Indeed, many of the empirically informed gene-culture coevolutionary models, such as those by Boyd and Richerson, are suggestive of such an account of the evolution of morality. Recent empirical work on coevolutionary models in the field of morality by Mesoudi and Laland (2007) and Mesoudi and Danielson (Mesoudi and Danielson 2008) lend some empirical weight to dual inheritance models in moral evolution.

17. Evidence for dual inheritance models themselves is not necessarily evidence for non-nativism. An innate moral sense can be consistent with elements of morality being inherited in the cultural channel. A proper evolutionary account will therefore involve an account of moral acquisition. Hence, the dual inheritance model I have presented in conjunction with the acquisition account of previous chapters specifically details a non-nativist account of the evolution of moral cognition.

6.7 Summary

In this chapter I have argued that moral cognition is a learned complex skill that evolves via a gene-culture coevolutionary process through dual channels of inheritance. Some of the main points are:

- Moral phenotypes are a product of genes and non-trivial structural contributions from the environment.
- The evolution of moral cognition is suited to a dual inheritance model of evolution.
- Dual inheritance models require reliable, high-fidelity transmission in both genetic and cultural channels of inheritance.
- Reliable, high-fidelity transmission in the cultural channel is achieved through mechanisms such as downstream niche construction, biased learning and children's participation and exploration of their moral worlds.
- Agent-environment interactions create feedback process that runs across generations. When this process allows the accumulation of modifications over generations, it becomes an evolutionarily significant process.
- Convergent evolutionary processes help explain non-trivial cross-cultural moral similarities between groups.

Conclusion

With the recent popularity of evolutionary psychology, nativism has often been viewed as the default position in evolutionary accounts of human psychology; our minds are collections of genetically specified psychological adaptations that evolved in response to evolutionary pressures deep in our lineage's history. In the case of morality, we are moral agents with the moral psychologies and culture that we have because of selection for our innate moral sense in response to adaptive challenges of the Pleistocene. The aim of this thesis has been to question this picture of morality and to develop alternative explanations of the moral mind and its evolution. To conclude, I will briefly summarise the core results and outline some future research directions. I will then look at some interesting implications, including the constraints on individual moralities, the realisability of moral theories, and the costs and constraints of moral obligation.

The alternative account of the evolution of moral cognition presented in this thesis is based on a number of key aspects. I detailed how moral cognition is selected to leverage the benefits of cooperation, and how there are a plurality of mechanisms through which it evolved. However, I argued that moral cognition is not innate; it is a learned adaptation. In support of this, I described some of the ways in which our culturally based developmental environments play a specific role in the transferring of moral capacities from one generation to the next, without recourse to innate moral information. I also argued against an assumption inherent in much of philosophy of mind and the cognitive sciences: our cognitive processes are bound by the brain and associated neural structures. Instead, much of moral cognition incorporates both body and environmental structures.

Again, culture plays a central role, supplying resources necessary for both the development and execution of moral cognition. One particular aspect I explored was the way in which our embodied social interactions create the very moral space in which we and importantly future generations operate. We construct our moral worlds and our moral worlds construct our moral minds. The moral world we create plays a vital evolutionary role in the transmission of moral phenotypes from one generation to the next. Pulling these threads together, I developed a dual inheritance model of the evolution of moral cognition. I explained how many of our moral cultural practices such as niche construction, psychological biases and the ways in which the learning child interacts with their world enable the reliable high-fidelity transmission of moral information between generations required for the cumulative evolution of moral phenotypes.

This view of the evolution of moral cognition leads to many new and interesting research avenues, both empirical and theoretical. Future work will involve developing a more detailed understanding of the exposure-competence relationship in moral development, including further analysis of the influence of parenting practises across and within cultures and the role of psychological biases. This will enable us to gain more insight into moral learning and help to resolve the nativism/non-nativism debate. As we have seen, an embodied cognitive science offers unique perspectives on this research and will contribute significantly to unravelling the underlying structure of morality. This will require further exploration of the roles that representations and non-representational components play in moral cognition and the many ways we integrate the body and world into our moral problem-solving routines. Future research in extended moral cognition will involve individuating different moral phenomena and identifying the various roles of brain, body and world. Progress in understanding the evolution of moral cognition requires developing a more detailed, empirically informed, dual inheritance model. We need empirical research that tracks the information flows through multiple inheritance channels and which locates the specific roles of learning and niche construction in this process. The evolution of moral disgust is a model of such research, and the aim is to extend this model to other aspects of moral cognising. Moral cognition, in turn, is a model for exploring philosophical questions on the nature of explanation in cognitive science and the boundary disputes in philosophy of mind and cognitive science. Further exploration of the relationship between the extended mind and evolution will provide

another avenue of future research. The extended mind involves cognitive manipulations of the external world which, as we have seen, have important downstream effects on future generations which are visible to selection.

The work in this thesis also offers new ways of looking at the moral domain, with some interesting implications. A central thread running throughout has been to question many of the traditional boundaries of explanation. I did this in two ways. I argued firstly that explanations of moral cognition are not confined to the brain and, secondly, that the evolutionary project involved redrawing the boundaries within which selection operates: selection does not act solely on genes, it also acts upon culture. In each of these explanatory frameworks, we saw the significance of the dynamic and reciprocal interactions between biology and culture both within an agent's lifetime and over generations. These interactions place central importance on both developmental environments and the environments in which the mature moral agent is embedded.

This raises the interesting question as to what philosophical moral thought experiments are uncovering. If our decision making is reflected in our responses to moral thought experiments, then those experiments are uncovering many of the non-moral biological, culturally-specific and evolutionary influences that determine individual moral judgements, rather than a deep universal truth about the innate structure of our moral psychology. We have already seen how our moral cognitions involve a confluence of factors, many of which can arise independent of any moral considerations. Understanding these, the situated nature of moral judgement and the framing effects of the thought experiments will lead to a better understanding of the determinants of moral thought and action.

One consequence of the perspective I have developed is an emphasis on the plasticity of human moral cognition; it is rich in possibility. We have good reason to doubt that the fundamental determinants of our moral cognitions and moral culture are to be found deep in our evolutionary history; our moral brains are not hardwired in our ancestral past. Instead, many of our moral capacities evolve as strategic responses to cooperation and coordination problems that can arise over comparatively short time scales. Our moral worlds and our moral minds can and do change in fundamental ways; the trajectory of the development and evolution of moral cognition is open. Yet one of the

implications of the account I have provided is that moral development and evolution involve both flexibility and constraint that affect individual moralities and group moral systems. The source of these constraints and biases differ from that of the nativist account in fundamental ways, meaning that different drivers influence the outcomes of development and evolution.

We can see these two elements at play in individual moral decision making. Under a general learning account we expect a large degree of developmental plasticity and a correspondingly large range of individual moralities. Yet there are a number of reasons why individual moralities and moral systems will be somewhat constrained. Firstly, the long period of juvenile dependence in which moral development occurs limits the variety of moral information to which we are exposed, developmentally entrenching that information and the moral capacities to which they give rise. Moral change will often be slow. Secondly, biases influence the transmission of moral information between generations and hence bias the moral psychologies we develop. Thirdly, strong emotional and other biases will constrain the types of moral beliefs and behaviours we find salient. For example, moral beliefs and judgements that involve emotions such as empathy and sympathy are less likely to be subject to strong individual decision-making forces. This is because psychological and cultural biases dampen the individual decision making involved in the processing and transmission of information. As Boyd and Richerson (2004) point out, we tend to adopt the same beliefs as family and friends when decision-making forces are weak. Finally, the capacity for moral cognition is dependent on the worlds within which they are embedded. The structure of those worlds will, therefore, in part determine the informational resources available to us, and the cognitions that we can and do make. This last point highlights one of the roles that the evolution of moral worlds plays in constraining individual moralities. Because our moral psychology in part reflects our moral world, the selective effects on those worlds impact on our moral psychologies. The evolutionary designs upon those worlds and the fitness benefits they confer help determine moral thought and action. Understanding the flexibility and constraints on moral thought and action will therefore involve, as I have done in this thesis, detailing the interaction between individual and world through an evolutionary lens.

This picture of moral cognition and its evolution is relevant to debates about normative

theories and the moral actions we choose. For example, the emphasis on beyond-the-brain structures impacts upon our understanding of the psychological realisability of moral theories, the true costs of our moral obligations, and the constraints in satisfying those obligations.

At a fundamental level, we want to know if our moral theories that prescribe thought and action are psychologically possible for normal agents – are our moral theories psychologically realisable? (Flanagan 1991).¹ The realisability of moral theories will depend on those elements constitutive of moral cognition. For example, certain moral theories will simply not be realisable in a population of psychopaths because they lack elements, such as particular emotional capacities, to think morally in the right ways. Based on the work in this thesis, we can see there are numerous additional realisability constraints placed upon moral theories beyond the mere brainbound psychology of traditional cognitive science. If moral cognising and the potential resources that support moral decision making are dependent on body and world, then they will play a central role in establishing the realisability of moral theories in human populations. This will include group and information structures and the many biases and constraints they afford. What we can do depends on what we can think, and that in turn crucially depends on our physical and social environment. With additional consideration to developmental and evolutionary effects, we have the tools to consider what types of moral systems are possible.

This view of moral cognition and its evolution also impacts upon the way in which we identify the costs of, and constraints upon, moral compliance. This is important because the costs involved in obtaining a moral end are relevant to our calculations as to how, and even if we ought, to pursue that end. Whether we intervene to save the lives of others for example, will depend, in part, on how many other lives will be at risk. These issues are important for such things as moral satisficing: in pursuit of maximising moral outcomes, we may find the sub-optimal implementation of a particular moral norm could in fact be worse than adopting an alternative, less demanding norm. Consider our moral obligations to future generations in the face of global climate change, for example. We

1. Owen Flanagan specifies the Principle of Minimal Psychological Realism as follows: “Make sure when constructing a moral theory or projecting a moral ideal that the character, decision processing, and behaviour prescribed are possible, or are perceived to be possible, for creatures like us” (Flanagan 1991:32). Note that this is a slightly different claim from “ought implies can” which applies to individual moral obligations rather than moral theories.

can either mitigate climate change, saving many millions of lives and preventing the suffering of future generations. Alternatively, we can focus our resources towards adapting the best we can to the consequences of climate change, but in doing so not save nearly as many lives or prevent as much suffering. Because of limited time and resources, we cannot successfully both mitigate and adapt. However, failure in pursuit of mitigation could have disastrous consequences. The subsequent loss of millions of lives would be a far greater moral cost than adapting. Therefore, faced with very difficult task of mitigation and the very high costs of its failure, adapting to climate change may in fact be morally 'good enough'. Understanding the true cost is important when attempting to achieve the best outcome. It is therefore important to understand those factors which constrain moral maximisation and the costs involved (Goodin 2009).

Costs and constraints are also important for establishing the demands of our moral obligations. Morality places demands on individuals and it is (often) deemed morally wrong to fail to meet those demands. However, moral demands vary, as do the costs and constraints on meeting them. It is claimed that some moral theories make demands that ask too much of moral agents (Cullity 2004). A moral theory that asks the affluent to give the vast majority of their money to save the world's poor, for example, is thought by many to demand too much. There are a number of reasons why moral obligations may be judged as too demanding. They may be too costly because they ask us to sacrifice too much of our way of life; they conflict with other obligations, such as those to friends and family; they cause physical or psychological harm; or, they are simply too inconvenient. The problem of excessive demands leads us to question how much a moral theory can legitimately demand from us, and highlights the importance of understanding the true demands of moral theories.²

The case I have put forward is especially relevant to a related problem: identifying the demands. Our ability to meet moral obligations may be constrained by difficulties in obtaining the information required to take the right course of action, contingent on the

2. It is often claimed that if a moral theory imposes extreme demands it may give us reason to question aspects of morality. Garret Cullity (2004) identifies three different views in response to the problem of (extreme) demandingness: (1) that the particular moral outlook be rejected because of the absurd self-denying demands that it imposes; (2) extreme demandingness is a problem for morality itself because the unreasonableness of the demand gives one a reason to reject morality; and (3) the problem of demandingness is not a problem for morality, but for the life of one who tries to live a moral life – morality may be more demanding than what one would like it to be.

costs they incur and the resources we have available to satisfy them. The level of moral demand often depends heavily on the structures of our social worlds, and especially the actions of others. If we all make a sufficient number of small charitable donations, thereby satisfying the demands of charity, there is reduced moral demand on individuals to make large donations. The ways in which the world is organised is therefore intimately related to solving problems of demandingness (Goodin 2009).

In highlighting the role of social structures in moral cognising, the work in this thesis contributes to understanding the true costs of, and constraints upon, fulfilling our moral obligations and the demands they make. Everyone realises the social world of other agents influences the results of any individual's choice. But this understates its role. Because our extended moral cognitive abilities rely on external social structures, those abilities will often be constrained by the organisation of those structures. We need, therefore, to have the appropriate social structures and organisation in place to scaffold the development and execution of the individual moral capacities required to meet moral obligations. We need also to understand the transition costs of making changes to our moral worlds because they are relevant to the actions we ought to take. The demands of moral theories are dependent on the structure and dynamics of our moral worlds and our engagement with them. As we have seen, the organisation of our moral worlds has many drivers, constraints and biases, including evolutionary effects. Understanding that moral cognition and its evolution is reliant on brain, body and world provides new tools for establishing the true costs and constraints of morality.

One way of visualising these ideas is by detailing the Andy Clark-style moral space we encountered in earlier chapters.³ Moral space is a socially constructed moral ecology within which we live. Its shape is in part determined by the diverse and interacting needs and desires of its occupants. The creation of tools such as moral concepts, norms and maxims help us navigate moral space. The structure and dynamic nature of moral space is important for both individual and group moral problem solving, its configuration

3. See also recent work by Marc Hauser (2009). He has recently explored the idea of moral space (or as he refers, a 'moralspace'). Based on the idea of morphospace, which plots all possible morphologies of a particular organism, Hauser has proposed applying similar modelling principles to multiple domains of human psychology and culture such as linguistics and music, including a moralspace that details all possible moral systems. One significant area where Hauser differs from the notion of moral space used here is in the role he attributes to innate moral grammar as the central generative element and constraining factor in all possible moral systems. This is in contrast to the dynamic nature of a socially constructed moral space. (See also Owen Flanagan's (1991) discussion of 'psychological possibility space'.)

having consequences on constraints and costs. The possible space we occupy is, as Clark tells us, dependent on the histories we have. We have seen that this will also include evolutionary histories.

If we conceive of the moral space we presently inhabit, we can also conceive of the past moral space we have occupied and future possible spaces we *can* inhabit. By modelling the changes in moral space over time and plotting trajectories through that space we will see the various processes that impact upon the likelihood of various future trajectories. We can identify the costs of, and constraints upon, changing moral space (i.e., changing our moral worlds to meet our moral obligations) and the effects these changes have on such things as moral demands. Positive and negative constraints will be represented as troughs and hills across the terrain of possible moral space. Some regions will form strong basins of attraction towards which nearby trajectories will be drawn. Some possible moral space will be uninhabitable or, due to the difficult terrain that leads to them, inaccessible.

We can distinguish two types of constraints that bear on moral trajectories. The first type of constraint is *moral transition costs* which estimate the moral costs of adopting future moral obligations. These will include such calculations as how many lives are lost if we attempt to obtain a moral end, and the consideration of trade-offs that occur in adopting principles that are fair to some and unfair to others, for example. The ways in which the world is organised can impact on the moral costs of meeting moral obligations and the ease of transitions. Systems of greater resource distribution, for example, will reduce the price of individual moral commitments to the poor, improving our abilities to meet moral demands and reduce moral costs. Although moral transition costs may not *necessarily* make a trajectory less likely, they do figure heavily in our moral calculations.

The second type of constraint is *causal constraints*. Some moral trajectories are impossible, or very unlikely, because of the histories we have, our current location in possible moral space, and the structure of the world and human psychology. There will be regions of possible moral space that are more psychologically amiable than others as well as better structured to meet certain moral obligations and implement change.

The work in this thesis has clear ramifications for identifying causal constraints. The impacts of evolutionary effects, such as generative entrenchment for example, will become apparent. Here the evolution (biological and cultural) and development of complex phenomena, such as moral phenotypes, occurs against the background of those structures and processes that instantiate them - their evolution and expression depend upon upstream structures and processes, meaning they are resistant to developmental and evolutionary change (Wimsatt 2007). Aspects of moral cognition and our moral worlds are resistant to change because they depend on the structures and processes from which they were built – what Wimsatt calls “accumulated dependencies”. This has obvious implications for individual moral psychologies. Although morality is learned, certain concepts, norms, behaviours and moral resources are generatively entrenched features of evolved systems, so it is unlikely we will lose them. The direction of possible future moral cognition and moral worlds will be constrained by the requirement to maintain those dependencies. We need only think of the many capacities, concepts and norms upon which moral thinking relies. These capacities are entrenched because we cannot lose them without losing the moral structures which rely on them. Generative entrenchment also has implications for group moral change. Concepts of right and wrong, for example, are fundamental to the operation and adaptive success of many of our group-wide norms and social interactions. They are entrenched features of both individual moralities and group moral space, necessary for the existence of the worlds we presently inhabit and constraining the future possible moral worlds we can inhabit. Trajectories towards future moral worlds will involve both moral transition costs and causal constraints.

When we consider this in the context of future possible moral space, we may end up in regions from which retreat, or the navigation to new regions, will be difficult and sometimes impossible because they involve alterations to generatively entrenched features. Our future moral trajectories are thus constrained. Fairness is one example of a generatively entrenched component of individual moralities and complex group moral systems. The evolution of the concept of fairness provided adaptive solutions to many cooperative activities such as food sharing and hunting through to modern day trade and exchange. It transformed the ways in which we structure the world, altering the moral space which we occupy. Much of this newly reorganised moral world and its adaptive benefits depend on the concept of fairness, such as norms that guide our behaviours,

institutions, legal systems, as well as new concepts that are further elaborations upon fairness. The social construction of complex adaptive structures stabilise the emergence of fairness and the developmental processes that instantiate it. The removal of concepts such as fairness from our moral world, and the means by which it develops, are unlikely to occur because such alterations would have disastrous effects for many of those (adaptive) downstream elements that depend on it. The future possible moral worlds we can inhabit are constrained by the elements of morality upon which much depends, meaning our trajectories through one moral world takes us further away from some possible moral worlds and closer to others. Generative entrenchment is an example of how costs and constraints (including evolvability constraints) can impact upon the organisation of future moral worlds, and in turn the fulfilment of moral obligations.

When we apply these ideas to traditional ethical dilemmas we may find that possible solutions similarly lead to locations in possible moral space that are uninhabitable, or distant regions which require traversing difficult, if not impossible, terrains. Consider our obligations to the world's poor, for example. Peter Singer has claimed that "if it is in our power to prevent something very bad from happening, without thereby sacrificing anything else morally significant, we ought, morally, to do it" (Singer 1972:235). The implication, according to Singer, is that the affluent world ought to make considerable sacrifices of their wealth to the third world, as long as it doesn't bring about equal deprivation. One of the core components of this view, and that which that underwrites his utilitarianism, is Singer's commitment to impartiality; we must look beyond the interests of our own society, and give all people equal consideration of interests. By modelling these possible worlds and the paths that lead to them we can get an understanding of the costs involved, and the constraints upon satisfying moral demands. Occupying regions of possible moral space that satisfy certain moral requirements may not be feasible. Even if those regions are themselves habitable, the transition costs in reaching them may mean it is excessively demanding to satisfy those moral obligations. Furthermore, partiality is an entrenched feature of many of our (adaptive) complex moral systems. As we have seen, partial behaviours, such as those towards one's group, allow the generation of cooperative benefits. Major shifts in partiality could have disastrous downstream consequences. The generative entrenchment of partiality constrains and imposes costs on future moral directions. Evaluating the possibility, cost and constraints of our moral theories will involve looking beyond mere brainbound psychology,

broadening our explanations to include beyond-the-brain resources and evolutionary considerations.

In this thesis I have explained how a non-nativist moral cognition can evolve and how we ought to broaden our explanatory scope beyond innate, genetically specified, brainbound psychological mechanisms to fully understand moral cognition. Although this move away from nativism in phylogenetic, ontogenic and synchronic explanations of moral cognition goes against the grain of much modern theorising in the biological and cognitive sciences, the work I have presented offers the beginnings of a promising and empirically tractable research programme. This will ultimately provide a new and deeper understanding of moral cognition and its evolution.

References

- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2), 169-177.
- Adolphs, R., H. Damasio, D. Tranel, G. Cooper, and A. R. Damasio (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience*, 20(7), 2683-2690.
- Alexander, R. D. (1987). *The Biology of Moral Systems*. Hawthorne, N.Y.: A. de Gruyter.
- Arsenio, W. F. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, 59(6), 1611-1622.
- Arsenio, W. F., and A. Lover (1995). Children's conceptions of sociomoral affect: Happy victimizers, mixed emotions, and other expectancies. In M. Killen and D. Hart (Eds.), *Morality in Everyday Life: Developmental Perspectives* (pp. 87-128). Cambridge, UK: Cambridge University Press.
- Axelrod, R. M. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Aytekin, M., C. F. Moss, and J. Z. Simon (2008). A sensorimotor approach to sound localization. *Neural Comp.*, 20(3), 603-635.
- Bargh, J. A., M. Chen, and L. Burrows (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2), 230-244.
- Barkow, J. H., L. Cosmides, and J. Tooby (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.
- Barnier, A. J., J. Sutton, C. B. Harris, and R. A. Wilson (2008). A conceptual and empirical framework for the social distribution of cognition: The case of memory. *Cognitive Systems Research*, 9(1-2), 33-51.
- Barrett, L., P. Henzi, and D. Rendall (2008). Social brain, simple minds: does social complexity really require cognitive complexity? In N. Emery, N. Clayton and C. D. Frith (Eds.), *Social Intelligence: From Brain to Culture* (pp. 123-146). Oxford ; New York: Oxford University Press.
- Barsalou, L. (2003). Situated simulation in the human conceptual system. In H. Moss and J. Hampton (Eds.), *Conceptual Representation* (pp. 513-562). Sussex: Psychology Press.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-45.
- Barsalou, L. W., A. K. Barbey, W. K. Simmons, and A. Santos (2005). Embodiment in religious knowledge. *Journal of Cognition and Culture*, 5(1), 14-57.
- Bateson, P. (1991). Are there principles of behavioural development? In P. Bateson (Ed.), *The Development and Integration of Behaviour: Essays in Honour of Robert Hinde* (pp. 19-39). Cambridge: Cambridge University Press.
- Bem, S., and H. Looren de Jong (2006). *Theoretical Issues in Psychology: An Introduction*.

- London: SAGE.
- Blair, R. J. R. (1995). A cognitive developmental approach to morality: investigating the psychopath. *Cognition*, 57(1), 1-29.
- Bodenhausen, G. V., L. A. Sheppard, and G. P. Kramer (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology*, 24(1), 45-62.
- Boehm, C. (2000). Conflict and the evolution of the social contract. In L. D. Katz (Ed.), *Evolutionary origins of morality: cross-disciplinary perspectives* (pp. 79-102). Thorverton: Imprint Academic.
- Bowles, S., and H. Gintis (2003). Origins of human cooperation. In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation*. Cambridge, Mass.: MIT Press.
- Boyd, R. (2007). Cultural adaptation and maladaptation: Of kayaks and commissars. In S. W. Gangestad and J. A. Simpson (Eds.), *The Evolution of Mind: Fundamental Questions and Controversies* (pp. 327-331). New York: Guilford Press.
- Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6), 3531-3535.
- Boyd, R., and P. J. Richerson (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3), 337-56.
- Boyd, R., and P. J. Richerson (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13(3), 171-195.
- Boyd, R., and P. J. Richerson (2004). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
- Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47, 139-159.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies In Philosophy*, 4(1), 73-121.
- Buss, D. M. (1999). *Evolutionary Psychology: The New Science of the Mind*. Boston: Allyn and Bacon.
- Cacioppo, J. T., J. R. Priester, and G. G. Bernston (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of personality and social psychology*, 65(1), 5-17.
- Casey, O. C. (2008). Seeing what you hear: Cross-modal illusions and perception. *Philosophical Issues*, 18(1), 316-338.
- Chartrand, T. L., and J. A. Bargh (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76, 893-910.
- Chemero, A., and M. Silberstein (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of Science*, 75(1), 1-27.
- Chiel, H. J., and R. D. Beer (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, 20(12), 553-557.
- Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.
- Churchland, P. M. (1996). The neural representation of the social world. In L. May, M. Friedman and A. Clark (Eds.), *Mind and Morals: Essays on Cognitive Science and Ethics* (pp. 91-108). Cambridge, Mass.: MIT Press.
- Ciaramelli, E., M. Muccioli, E. Ladavas, and G. di Pellegrino (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, 2(2), 84-92.
- Clark, A. (1996). Connectionism, moral cognition, and collaborative problem solving. In L. May, M. Friedman and A. Clark (Eds.), *Mind and Morals: Essays on Cognitive Science and Ethics* (pp. 109-127). Cambridge, Mass.: MIT Press.
- Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, Mass.: MIT Press.

- Clark, A. (1998). Embodiment and the philosophy of mind. In A. O'Hear (Ed.), *Current Issues in Philosophy of Mind: Royal Institute of Philosophy Supp. 43* (pp. 35-52). Cambridge: Cambridge University Press.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345-351.
- Clark, A. (2000). Words and action: Reconciling rules and know-how in moral cognition. *Moral Epistemology Naturalized: Canadian Journal of Philosophy Supp.*, 26, 267-290.
- Clark, A. (2001). *Mindware: An Introduction to the Philosophy of Cognitive Science*. New York: Oxford University Press.
- Clark, A. (2004). Embodiment and the philosophy of mind. In A. Peruzzi (Ed.), *Mind and Causality* (pp. 35-51). Amsterdam: John Benjamins.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.
- Clark, A., and D. Chalmers (1998). The extended mind. *Analysis*, 58(1), 7-19.
- Clore, G. L., and J. R. Huntsinger (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393-399.
- Cosmides, L., and J. Tooby (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 163-228). New York: Oxford University Press.
- Cosmides, L., J. Tooby, and J. H. Barkow (1992). Introduction: Evolutionary psychology and conceptual integration. In J. H. Barkow, L. Cosmides and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (pp. 3-15). New York: Oxford University Press.
- Cowie, F. (1999). *What's Within?: Nativism Reconsidered*. Oxford: Oxford University Press.
- Cullity, G. (2004). *The Moral Demands of Affluence*. Oxford: Oxford University Press.
- Cushman, F., L. Young, and M. Hauser (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, 17(12), 1082-1089.
- Damasio, A. (1994 (2005)). *Descartes' Error: Emotion, Reason, and the Human Brain*. London: Penguin Books.
- Damasio, A. R., T. J. Grabowski, A. Bechara, H. Damasio, L. L. B. Ponto, J. Parvizi, and R. D. Hichwa (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3(10), 1049-1056.
- Damasio, A. R., D. Tranel, and H. Damasio (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, 41(2), 81-94.
- Darwin, C. (1872). The expression of the emotions in man and animals. Reprinted in *Darwin: The Indelible Stamp*. J. D. Watson. Philadelphia, NY, Running Press Book Publishers (2007).
- Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press, 1999.
- Dawkins, R. (1989). *The Selfish Gene*. Oxford: Oxford University Press, 1999.
- De Gelder, B. and P. Bertelson (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, 7(10), 460-467.
- de Sousa, R. (2008). Emotion. Reprinted in *Stanford Encyclopedia of Philosophy*. E. N. Zalta, Fall.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon & Schuster.
- Dennett, D. C. (1996). *Kinds of Minds: Toward an Understanding of Consciousness*. New York: Basic Books.
- Dreyfus, H. L., S. E. Dreyfus, and T. Athanasiou (1986). *Mind Over Machine: The Power of*

- Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Dunbar, R. I. M. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge, Mass.: Harvard University Press.
- Dunn, J. (2003). Emotional development in early childhood: A social relationship perspective. In R. J. Davidson, H. H. Goldsmith and K. R. Scherer (Eds.), *Handbook of Affective Sciences* (pp. 332-346). Oxford: Oxford University Press.
- Dunn, J. (2006). Early moral development and social interactions. In M. Killen and J. G. Smetana (Eds.), *Handbook of Moral Development* (pp. 331-350). London: Lawrence Erlbaum Associates.
- Dunn, J., and C. Slomkowski (1992). Conflict and the development of social understanding. In C. U. Shantz and W. W. Hartup (Eds.), *Conflict in Child and Adolescent Development*. (pp. 70-92). New York: Cambridge University Press.
- Dupoux, E., and P. Jacob (2007). Universal moral grammar: a critical appraisal. *Trends in Cognitive Sciences*, 11(9), 373-378.
- Dwyer, S. (1999). Moral competence. In K. Murasugi and R. Stainton (Eds.), *Philosophy and Linguistics* (pp. 169-190). Boulder, CO: Westview Press.
- Dwyer, S. (2006). How good is the linguistic analogy? In P. Carruthers, S. Laurence and S. Stich (Eds.), *The Innate Mind: Culture and Cognition* (pp. 237-256). New York: Oxford University Press.
- Dwyer, S. (2009). Moral dumbfounding and the linguistic analogy: Methodological implications for the study of moral judgment. *Mind & Language*, 24(3), 274-296.
- Ehrlich, P. R., and S. A. Levin (2005). The evolution of norms. *PLoS Biology*, 3(6), e194.
- Eisenberg, N., T. L. Spinrad, and A. Sadovsky (2006). Empathy-related responding in children. In M. Killen and J. G. Smetana (Eds.), *Handbook of Moral Development*. London: Lawrence Erlbaum Associates.
- Ekman, P. (1999). Facial expressions. In T. Dalgleish and M. J. Power (Eds.), *Handbook of cognition and emotion* (pp. 301-320). Chichester, England ; New York: Wiley.
- Ekman, P., and W. V. Friesen (1971). Constants across cultures in the face and emotion. *Journal of Personality & Social Psychology*, 17, 124-129.
- Ekman, P., R. M. Liebert, W. V. Friesen, R. Harrison, C. Zlatichin, E. J. Malmstrom, and R. A. Baron (1972). Facial expressions of emotion while watching televised violence as predictors of subsequent aggression. *Television and Social Behavior*, 5, 22-58.
- Ellis, H. D., A. W. Young, A. H. Quayle, and K. W. De Pauw (1997). Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society B: Biological Sciences*, 264(1384), 1085-1092.
- Fehr, E., and U. Fischbacher (2004). Third-party punishment and social norms. *Evolution and Human Behavior*, 25(2), 63-87.
- Fehr, E., U. Fischbacher, and S. Gächter (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, 13(1), 1-25.
- Flanagan, O. J. (1991). *Varieties of moral personality : ethics and psychological realism*. Cambridge, Mass.: Harvard University Press.
- Fodor, J. (1981). The present status of the innateness controversy. In *Representations: Philosophical Essays on the Foundations of Cognitive Science* (pp. 257-316). Cambridge, Mass.: MIT Press.
- Fodor, J. A. (1975). *The Language of Thought*. New York: Crowell.
- Fodor, J. A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, Mass.: MIT Press.
- Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. New York: Norton.
- Frank, R. H. (2001). Cooperation through emotional commitment. In R. M. Nesse (Ed.),

- Evolution and the Capacity for Commitment*. New York: Russell Sage.
- Gallagher, S. (2005). *How the Body Shapes the Mind*. Oxford: Clarendon Press.
- Gallese, V. (2008). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. In N. Emery, N. Clayton and C. D. Frith (Eds.), *Social Intelligence: From Brain to Culture* (pp. 279-296). Oxford: Oxford University Press.
- Gangestad, S. W., and J. A. Simpson (2007). *The Evolution of Mind: Fundamental Questions and Controversies*. New York: Guilford Press.
- Giere, R. N., and B. Moffatt (2003). Distributed cognition: Where the cognitive and the social merge. *Social Studies of Science*, 33(2), 301.
- Gintis, H. (2000). Group selection and human prosociality. In L. D. Katz (Ed.), *Evolutionary Origins of Morality: Cross-Disciplinary Perspectives* (pp. 215-219). Thorverton, UK: Imprint Academic.
- Gintis, H., E. Smith, and S. Bowles (2001). Costly signaling and cooperation. *Journal of Theoretical Biology*, 213(1), 103-119.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11), 419-429.
- Goodin, R. (2009). Demandingness as a Virtue. *The Journal of Ethics*, 13(1), 1-13.
- Greene, J., and J. Haidt (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, 6(12), 517 - 523.
- Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2), 389-400.
- Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.
- Griffiths, P., and R. Gray (2005). Three ways to misunderstand developmental systems theory. *Biology and Philosophy*, 20(2), 417-425.
- Griffiths, P., E. Machery, and S. Linquist (2009). The vernacular concept of innateness. *Mind & Language*, 24(5), 605-630.
- Griffiths, P. E. (2002). What is innateness? *The Monist*, 85(1), 70-85.
- Gürerk, Ö., B. Irlenbusch, and B. Rockenbach (2006). The competitive advantage of sanctioning institutions. *Science*, 312(5770), 108-111.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814 - 834.
- Haidt, J. (2003). The moral emotions. In R. J. Davidson, H. H. Goldsmith and K. R. Scherer (Eds.), *Handbook of Affective Sciences* (pp. pp. 852-870). Oxford: Oxford University Press.
- Haidt, J., P. Rozin, C. McCauley, and S. Imada (1997). Body, psyche, and culture: The relationship between disgust and morality. *Psychology and Developing Societies*, 9, 107-131.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *J Theor Biol*, 7(1), 1-16.
- Harman, G. (2000). Moral philosophy and linguistics. In *Explaining Value: and Other Essays in Moral Philosophy* (pp. 217-226). Oxford: Oxford University Press.
- Harman, G. (2008). Using a linguistic analogy to study morality. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 345-351). Cambridge, Mass.: MIT Press.
- Harms, W., and B. Skyrms (2008). Evolution of moral norms. In M. Ruse (Ed.), *The Oxford Handbook of Philosophy of Biology* (pp. 434-450). New York: Oxford University Press.

- Harris, P. L., and C. Saarni (1989). Children's understanding of emotion: an introduction. In P. L. Harris and C. Saarni (Eds.), *Children's Understanding of Emotion* (pp. 3 - 24). Cambridge: Cambridge University Press.
- Hastings, P. D., C. Zahn-Waxler, and K. McShane (2006). We are, by nature, moral creatures: biological bases of concern for others. In M. Killen and J. G. Smetana (Eds.), *Handbook Of Moral Development* (pp. 483 - 516). London: Lawrence Erlbaum Associates.
- Haugeland, J. (1998a). *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, Mass.: Harvard University Press.
- Haugeland, J. (1998b). Mind embodied and embedded. In *Having Thought: Essays in the Metaphysics of Mind* (pp. 206-237). Cambridge, Mass.: Harvard University Press.
- Hauser, M. (2006a). The liver and the moral organ. *SCAN*, 1, 214-220.
- Hauser, M. (2006b). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco.
- Hauser, M., L. Young, and F. Cushman (2008a). On misreading the Linguistic Analogy: Response to Jesse Prinz and Ron Mallon. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2, The Cognitive Science of Morality: Intuition and Diversity* (pp. 171-180). New York: Oxford University Press.
- Hauser, M., L. Young, and F. Cushman (2008b). Reviving Rawls' Linguistic Analogy: Operative principles and the causal structure of moral actions. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2, The Cognitive Science of Morality: Intuition and Diversity* (pp. 107-144). New York: Oxford University Press.
- Hauser, M. D. (2009). The possibility of impossible cultures. *Nature*, 460(9), 190196.
- Henrich, J. (2006). Cooperation, punishment, and the evolution of human institutions. *Science*, 312(5770), 60-61.
- Henrich, J., and R. Boyd (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, 208(1), 79-89.
- Henrich, J., R. McElreath, et al. (2006). Costly punishment across human societies. *Science*, 312(5781), 1767-1770.
- Huebner, B., S. Dwyer, et al. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), 1-6.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, Mass.: MIT Press.
- James, R., and R. Blair (1996). Brief report: Morality in the autistic child. *Journal of Autism and Developmental Disorders*, 26(5), 571-579.
- Jones, M. (2007). *Feast: Why Humans Share Food*. Oxford: Oxford University Press.
- Joyce, R. (2006). *The Evolution of Morality*. Cambridge, Mass.: MIT Press.
- Kelly, D., and S. Stich (2007). Two theories about the cognitive architecture underlying morality. In P. Carruthers, S. Laurence and S. Stich (Eds.), *The Innate Mind* (pp. 348-366). New York: Oxford University Press.
- Kelly, D., S. Stich, K. J. Haley, S. J. Eng, and D. M. T. Fessler (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117-131.
- Keltner, D. (1996). Evidence for the distinctness of embarrassment, shame, and guilt: A study of recalled antecedents and facial expressions of emotion. *Cognition & Emotion*, 10(2), 155-172.
- Keltner, D., and P. Ekman (2000). Facial expression of emotion. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 236-249). New York: Guilford Press.
- Kirsh, D., and P. Maglio (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18(4), 513 - 549.
- Kochanska, G. (1991). Socialization and temperament in the development of guilt and

- conscience. *Child Development*, 62(6), 1379-1392.
- Kurzban, R., P. DeScioli, and E. O'Brien (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28(2), 75-84.
- Laible, D. J., and R. A. Thompson (2002). Mother-child conflict in the toddler years: Lessons in emotion, morality, and relationships. *Child Development*, 73(4), 1187-1203.
- Lakoff, G., and M. Johnson (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lakoff, G., and M. Johnson (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Laland, K. N. (1993). The mathematical modelling of human culture and its implications for psychology and the human. *British Journal of Psychology*, 84(2), 145.
- Laland, K. N., and G. R. Brown (2006). Niche construction, human behavior, and the adaptive-lag hypothesis. *Evolutionary Anthropology*, 15(3), 95.
- Laupa, M. (1994). "Who's in charge?" Preschool children's concepts of authority. *Early Childhood Research Quarterly*, 9(1), 1-17.
- Laupa, M., and E. Turiel (1986). Children's conceptions of adult and peer authority. *Child Development*, 57(2), 405-412.
- Leimar, O., and P. Hammerstein (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, 268(1468), 745-753.
- Lerner, J. S., and D. Keltner (2000). Beyond valence: Toward a model of emotion-specific influences on judgement and choice. *Cognition & Emotion*, 14(4), 473 - 493.
- Lerner, J. S., and D. Keltner (2001). Fear, anger, and risk. *Journal of Personality and Social Psychology*, 81(1), 146-159.
- Levin, P. F., and A. M. Isen (1975). Further Studies on the Effect of Feeling Good on Helping. *Sociometry*, 38(1), 141-147.
- Levy, N. (2004). Evolutionary psychology, human universals, and the standard social science model. *Biology and Philosophy*, 19(3), 459-472.
- Lieberman, D. (2008). Moral sentiments relating to incest. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 165-190). Cambridge, Mass.: MIT Press.
- Lieberman, D., J. Tooby, and L. Cosmides (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society B: Biological Sciences*, 270(1517), 819-826.
- Lieberman, D., J. Tooby, and L. Cosmides (2007). The architecture of human kin detection. *Nature*, 445(7129), 727-731.
- Linquist, S. (2007). Prospects for a dual inheritance model of emotional evolution. *Philosophy of Science*, 74(5), 848-859.
- Machery, E. (2006). Two dogmas of neo-empiricism. *Philosophy Compass*, 1(4), 398-412.
- Machery, E. (2007). Concept empiricism: A methodological critique. *Cognition*, 104(1), 19-46.
- Maglio, P. P., T. Matlock, D. Raphaely, B. Chernicky, and D. Kirsh (1999). Interactive skill in Scrabble. *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*, 326-330.
- Mameli, M., and P. Bateson (2006). Innateness and the sciences. *Biology and Philosophy*, 21, 155-188.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W.H. Freeman.
- Maynard Smith, J. M., and G. R. Price (1973). The Logic of Animal Conflict. *Nature*, 246(5427), 15-18.

- McElreath, R., T. H. Clutton-Brock, (2003). Group report: The role of cognition and emotion in cooperation. In P. Hammerstein (Ed.), *Genetic and Cultural Evolution of Cooperation* (pp. 125-152). Cambridge, Mass.: MIT
- Menary, R., Ed. (forthcoming-a). *The Extended Mind*. Aldershot, UK, Ashgate.
- Menary, R. (forthcoming-b). Introduction: The extended mind in focus. In R. Menary (Ed.), *The Extended Mind*. Aldershot, UK: Ashgate.
- Mesoudi, A. and P. Danielson (2008). Ethics, evolution and culture. *Theory Biosci.*, 127(3), 229-40.
- Mesoudi, A., and K. N. Laland (2007). Culturally transmitted paternity beliefs and the evolution of human mating behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 274(1615), 1273-1278.
- Mikhail, J. (2002). Aspects of the theory of moral cognition: Investigating intuitive knowledge of the prohibition of intentional battery and the principle of double effect. *Georgetown Law and Economics Research Paper No. 762385*, Available at SSRN: <http://ssrn.com/abstract=762385>.
- Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, 11(4), 143-152.
- Mikhail, J. (2008a). Moral cognition and computational theory. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3, The Neuroscience of Morality* (pp. 81-92). Cambridge, Mass.: MIT Press.
- Mikhail, J. (2008b). The poverty of the moral stimulus. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 353-359). Cambridge, Mass.: MIT Press.
- Milinski, M. (2002). Donors to charity gain in both indirect reciprocity and political reputation. *Proceedings of the Royal Society B: Biological Sciences*, 269(1494), 881-883.
- Milinski, M., D. Semmann, and H. Krambeck (2002). Reputation helps solve the 'tragedy of the commons'. *Nature*, 415(6870), 424-426.
- Miller, G. (2007). Sexual selection for moral virtues. *The Quarterly Review of Biology*, 82(2), 97-125.
- Moll, J., R. de Oliveira-Souza, and R. Zahn (2008). The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences*, 1124(1), 161-180.
- Moll, J., R. Zahn, R. de Oliveria-Souza, F. Krueger, and J. Grafman (2005). The neural basis of human moral cognition. *Nature Review of Neuroscience*, 6(10), 799-809.
- Monnin, T., and J. Liebig (2008). Understanding eusociality requires both proximate and ultimate thinking and due consideration of individual and colony-level interests. *Oikos*, 117(10), 1441-1443.
- Morris, J. S. (2002). How do you feel? *Trends in Cognitive Sciences*, 6(8), 317-319.
- Nadelhoffer, T., and A. Feltz (2008). The actor-observer bias and moral intuitions: Adding fuel to Sinnott-Armstrong's fire. *Neuroethics*, 1(2), 133-144.
- Nichols, S. (2002a). Norms with feeling: towards a psychological account of moral judgment. *Cognition*, 84, 221 - 236.
- Nichols, S. (2002b). On the genealogy of norms: A case for the role of emotion in cultural evolution. *Philosophy of Science*, 69(2), 234-255.
- Nichols, S. (2004). *Sentimental Rules: On the Natural Foundations of Moral Judgment*. New York: Oxford University Press.
- Nichols, S. (2005). Innateness and moral psychology. In P. Carruthers, S. Laurence and S. Stich (Eds.), *The Innate Mind: Structure and Content*. New York: Oxford University Press.
- Nichols, S., and S. P. Stich (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

- Niedenthal, P. M. (2007). Embodying emotion. *Science*, 316(5827), 1002-1005.
- Niedenthal, P. M., L. W. Barsalou, F. Ric, and S. Kraith-Gruber (2005a). Embodiment in the acquisition and use of emotion knowledge. In L. F. Barrett, P. M. Niedenthal and P. Winkielman (Eds.), *Emotion and Consciousness* (pp. 21-50): Guilford Press.
- Niedenthal, P. M., L. W. Barsalou, P. Winkielman, S. Krauth-Gruber, and F. Ric (2005b). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, 9(3), 184-211.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560-1563.
- Nowak, M. A., and K. Sigmund (2005). Evolution of indirect reciprocity. *Nature*, 437(7063), 1291-1298.
- O'Regan, J. K., and A. Noë (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(05), 939-973.
- Odling-Smee, F. J., K. N. Laland, and M. W. Feldman (2003). *Niche Construction: The Neglected Process in Evolution*. Princeton: Princeton University Press.
- Panchanathan, K., and R. Boyd (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432(7016), 499-502.
- Pecher, D., R. Zeelenberg, and L. W. Barsalou (2003). Verifying different-modality properties for concepts produces switching costs. *Psychological Science*, 14(2), 119-124.
- Pecher, D., and R. A. Zwaan (2005). *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge, UK: Cambridge University Press.
- Pinker, S. (1994). *The Language Instinct*. New York: W. Morrow and Co.
- Pinker, S. (1997). *How the Mind Works*. London: Penguin.
- Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Plotkin, H. C. (2007). *Necessary Knowledge*. Oxford: Oxford University Press.
- Priester, J. R., J. T. Cacioppo, and R. E. Petty (1996). The influence of motor processes on attitudes toward novel versus familiar semantic stimuli. *Pers Soc Psychol Bull*, 22(5), 442-447.
- Prinz, J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*. Cambridge, Mass.: MIT Press.
- Prinz, J. (2004). *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Prinz, J. (2005). Passionate thoughts. In D. Pecher and R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 93-114). Cambridge, UK: Cambridge University Press.
- Prinz, J. (2006). The emotional basis of moral judgements. *Philosophical Explorations*, 9(1), 29-43.
- Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- Prinz, J. (2008a). Is morality innate? In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 367-406). Cambridge, Mass.: MIT Press.
- Prinz, J. (2008b). Reply to Dwyer and Tiberius. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 427-439). Cambridge, Mass.: MIT Press.
- Prinz, J. (2008c). Resisting the linguistic analogy: A commentary on Hauser, Young, and Cushman. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 2, The Cognitive Science of Morality: Intuition and Diversity* (pp. 157-170). Cambridge, Mass: MIT Press.
- Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7(131-193).

- Pylyshyn, Z. W. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, Mass.: MIT Press.
- Ramachandran, V. S., and S. Blakeslee (1998). *Phantoms in the Brain: Human Nature and the Architecture of the Mind*. London: Fourth Estate.
- Ratnieks, F. L. W., T. Monnin, and K. R. Foster (2001). Inclusive fitness theory: novel predictions and tests in eusocial Hymenoptera. *Annales Zoologici Fennici*, 38, 201-214.
- Recanzone, G. H. (2002). Where was that? - human auditory spatial processing. *Trends in Cognitive Sciences*, 6(8), 319-320.
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10(3), 218 - 221.
- Rozin, P., J. Haidt, and C. R. McCauley (2000). Disgust. In M. Lewis and J. M. Haviland-Jones (Eds.), *Handbook of Emotions* (pp. 637-653). New York: Guilford Press.
- Rozin, P., L. Lowery, S. Imada, and Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4), 574-586.
- Rumelhart, D. E., and J. L. McClelland (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, Mass.: MIT Press.
- Saarni, C., and P. Harris (1989). *Children's Understanding of Emotion*. Cambridge: Cambridge University Press.
- Saarni, C., D. L. Mumme, and J. Campos (1998). Emotional development: Action, communication, and understanding. In N. Eisenberg (Ed.), *Handbook of Child Psychology. Volume 3: Social, Emotional, and Personality Development* (pp. 237-309). New York: John Wiley & Sons, Inc.
- Samuels, R. (2002). Nativism in cognitive science. *Mind & Language*, 17(3), 233-265.
- Schnall, S., J. Haidt, G. L. Clore, and A. H. Jordan (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34(8), 1096-1109
- Semin, G. R., and E. R. Smith (2002). Interfaces of social psychology with situated and embodied cognition. *Cognitive Systems Research*, 3(3), 385-396.
- Shapiro, L. (2007). The embodied cognition research programme. *Philosophy Compass*, 2(2), 338-346.
- Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(3), 229-243.
- Singer, P. (1993). *Practical Ethics*. New York: Cambridge University Press.
- Sinnott-Armstrong, W., Ed. (2008). *Framing moral intuitions*. Moral Psychology, Volume 2: The Cognitive Science of Morality. Cambridge, Mass., MIT Press.
- Skyrms, B. (1996). *Evolution of the Social Contract*. New York: Cambridge University Press.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, 52, 1333-1336.
- Smetana, J. G. (1989). Toddlers' social interactions in the context of moral and conventional transgressions in the home. *Developmental Psychology*, 25(4), 499-508.
- Smetana, J. G. (1997). Parenting and the development of social knowledge reconceptualized: A social domain analysis. In J. E. Grusec and L. Kuczynski (Eds.), *Parenting and Children's Internalization of Values: A Handbook of Contemporary Theory* (pp. 162-192). New York: Wiley.
- Smetana, J. G. (1999). The role of parents in moral development: A social domain analysis. *Journal of Moral Education*, 28(3), 311-321.
- Smetana, J. G. (2002). Culture, autonomy, and personal jurisdiction in adolescent-parent relationships. In H. W. Reese and R. Kail (Eds.), *Advances in Child Development and Behaviour* (pp. 51-87). New York: Academic Press.
- Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in

- children's moral and social judgments. In M. Killen and J. G. Smetana (Eds.), *Handbook of Moral Development* (pp. 119 - 153). London: Lawrence Erlbaum Associates.
- Smith, E. R. (2008). Social relationships and groups: New insights on embodied and distributed cognition. *Cognitive Systems Research*, 9(1-2), 24-32.
- Smith, E. R., and G. R. Semin (2004). Socially situated cognition: Cognition in its social context. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (pp. 53-117). New York: Academic Press.
- Smith, E. R., and G. R. Semin (2007). Situated social cognition. *Current Directions in Psychological Science*, 16(3), 132-135.
- Sodian, B., and U. Frith (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry*, 33(3), 591-605.
- Spurrett, D., and S. J. Cowley (2004). How to do things without words: infants, utterance-activity and distributed cognition. *Language Sciences*, 26(5), 443-466.
- Sripada, C. S. (2008a). Nativism and moral psychology: Three models of the innate structure that shapes the contents of moral norms. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 319-343). Cambridge, Mass.: MIT Press.
- Sripada, C. S. (2008b). Reply to Harman and Mikhail. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 1, The Evolution of Morality: Adaptations and Innateness* (pp. 361-365). Cambridge, Mass.: MIT Press.
- Sripada, C. S., and S. Stich (2005). A framework for the psychology of norms. In P. Carruthers, S. Laurence and S. Stich (Eds.), *The Innate Mind: Structure and Contents* (pp. 280-301). New York: Oxford University Press.
- Stel, M., and A. van Knippenberg (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, 19(10), 984-985.
- Sterelny, K. (2003). *Thought in a Hostile World*. Oxford: Blackwell.
- Sterelny, K. (2004). Externalism, epistemic artifacts and the extended mind. In R. Schantz (Ed.), *The Externalist Challenge* (pp. 239-254). Berlin: Walter de Gruyter.
- Sterelny, K. (2006). The evolution and evolvability of culture. *Mind & Language*, 21, 137-165.
- Sterelny, K. (2007). An alternative to evolutionary psychology? In S. W. Gangestad and J. A. Simpson (Eds.), *The Evolution of Mind: Fundamental Questions and Controversies* (pp. 178-185). New York: Guilford Press.
- Sterelny, K. (2008). Social intelligence, human intelligence and niche construction. In N. Emery, N. Clayton and C. D. Frith (Eds.), *Social Intelligence: From Brain to Culture* (pp. 375-392). Oxford: Oxford University Press.
- Sterelny, K. (forthcoming). *The Fate of the Third Chimpanzee*. Cambridge, Mass.: MIT Press.
- Sterelny, K., and P. E. Griffiths (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.
- Strange, D., M. Gerrie, and M. Garry (2005). A few seemingly harmless routes to a false memory. *Cognitive Processing*, 6(4), 237-242.
- Thagard, P. (2005). *Mind: Introduction to Cognitive Science*. Cambridge, Mass.: MIT Press.
- Thagard, P. (2008). Cognitive Science. *The Stanford Encyclopedia of Philosophy*, from <http://plato.stanford.edu/archives/fall2008/entries/cognitive-science/>.
- Thagard, P., and B. Aubie (2008). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*, 17(3), 811-834.
- Thelen, E., and L. B. Smith (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.
- Thompson, R. A., D. J. Laible, and L. L. Ontai (2003). Early understanding of emotion,

- morality, and self: Developing a working model. In R. V. Kail (Ed.), *Advances in Child Development and Behaviour* (pp. 137-171). San Diego: Academic Press.
- Thompson, R. A., S. Meyer, and M. McGinley (2006). Understanding values in relationships: The development of conscience. In M. Killen and J. Smetana (Eds.), *Handbook of Moral Development* (pp. 267-297). New Jersey: Lawrence Erlbaum Associates.
- Thomson, J. J. (1986). *Rights, Restitution, and Risk: Essays, in Moral Theory*. Cambridge, Mass.: Harvard University Press.
- Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift fur Tierpsychologie*, 20, 410-433.
- Tisak, M. S., J. Tisak, and S. E. Goldstein (2006). Aggression, delinquency, and morality: A social-cognitive perspective. In M. Killen and J. G. Smetana (Eds.), *Handbook Of Moral Development* (pp. 611-629). London: Lawrence Erlbaum Associates.
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.
- Tomasello, M., and A. Whiten (1999). *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.
- Tooby, J., and L. Cosmides (1989). Evolutionary psychology and the generation of culture: I. Theoretical considerations. *Ethology & Sociobiology*, 10(1-3), 29-49.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35-57.
- Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.
- Turiel, E. (1998). The development of morality. In N. Eisenberg (Ed.), *Handbook of Child Psychology. Volume 3: Social, Emotional, and Personality Development* (pp. 863-886). New York: John Wiley & Sons, Inc.
- Valdesolo, P., and D. DeSteno (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, 17(6), 476-477.
- Vasquez, K., D. Keltner, D. H. Ebenbach, and T. L. Banaszynski (2001). Cultural variation and similarity in moral rhetorics: Voices from the Philippines and the United States. *Journal of Cross-Cultural Psychology*, 32(1), 93-120.
- Vrana, S., B. Cuthbert, and P. J. Lang (1989). Processing fearful and neutral sentences: Memory and heart rate change. *Cognition and Emotion*, 3, 179-195.
- Waldmann, M. R. (2007). Combining versus analyzing multiple causes: How domain assumptions and task context affect integration rules. *Cognitive Science*, 31(2), 233 - 256.
- Waldmann, M. R., and J. H. Dieterich (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, 18(3), 247-253.
- Wallbott, H. G. (1991). Recognition of emotion from facial expression via imitation? Some indirect evidence for an old theory. *British Journal of Social Psychology*, 30, 207-219.
- Wells, G. L., and R. E. Petty (1980). The Effects of Over Head Movements on Persuasion: Compatibility and Incompatibility of Responses. *Basic and Applied Social Psychology*, 1(3), 219 - 230.
- Wheatley, T., and J. Haidt (2005). Hypnotic disgust makes moral judgements more severe. *Psychological Science*, 16(10), 780-784.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.
- Wilson, R. A. (2000). The mind beyond itself. In D. Sperber (Ed.), *Metarepresentations: A Multidisciplinary Perspective* (pp. 31-52). Oxford: Oxford University Press.

- Wilson, R. A. (2001). Group-level cognition. *Philosophy of Science*, 68(s1), S262.
- Wilson, R. A. (2004). Recent work in individualism in the social, behavioral and biological sciences. *Biology and Philosophy*, 19(3), 397-423.
- Wilson, R. A. (forthcoming). Meaning making and the mind of the externalist. In R. Menary (Ed.), *The Extended Mind*. Cambridge, Mass.: MIT Press.
- Wilson, R. A., and A. Clark (2009). How to situate cognition: Letting nature take its course. In P. Robbins and M. Aydede (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.
- Wimsatt, W. C. (2007). *Re-engineering Philosophy for Limited Beings*. Cambridge, Mass.: Harvard University Press.
- Winkielman, P., P. M. Niedenthal, and L. Oberman (2008). The embodied emotional mind. In G. R. Semin and E. R. Smith (Eds.), *Embodied Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches* (pp. 263-288). Cambridge: Cambridge University Press.
- Wohl, M. J., and G. D. Reeder (2004). When bad deeds are forgiven: Judgments of morality and forgiveness for intergroup aggression. In J. P. Morgan (Ed.), *Focus on Aggression Research* (pp. 59-74). Hauppauge, N.Y.: Nova Science Publishers.
- Wright, R. (1994). *The Moral Animal: Evolutionary Psychology and Everyday Life*. London: Abacus.
- Yau, J., and J. G. Smetana (2003). Conceptions of moral, social-conventional, and personal events among Chinese preschoolers in Hong Kong. *Child Development*, 74(3), 647-658.
- Yau, J., J. G. Smetana, and A. Metzger (2009). Young Chinese children's authority concepts. *Social Development*, 18(1), 210-229.
- Young, L., F. Cushman, M. Hauser, and R. Saxe (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20), 8235.
- Zahn-Waxler, C., and M. Chapman (1982). Immediate antecedents of caretakers' methods of discipline. *Child Psychiatry and Human Development*, 12(3), 179-192.
- Ziemke, T. (2003). What's that thing called embodiment. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 1305-1310.
- Zwaan, R. A. (1999). Embodied cognition, perceptual symbols, and situation models. *Discourse Processes*, 28(1), 81-88.
- Zwaan, R. A., R. A. Stanfield, and R. H. Yaxley (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2), 168-171.