Evolution of the Social Contract

submitted by

Daniel J. Taylor

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Computer Science

March 2014

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signature of Author ....

Daniel J. Taylor

SUMMARY

Humans act to bring about ends that are individually beneficial. Within groups, we also act concertedly to bring about ends that are beneficial to the group. From an evolutionary perspective, there is a tension between these two forces, because, more often than not, what is good for the individual is bad for the group, and what is good for the group is bad for the individual. In this dissertation, I argue that in order to resolve this tension between individual and group good, humans have evolved to utilise social contracts — sets of rules that structure interactions within our societies. These contracts, I argue, emerged in our hunter-gatherer ancestry as a means of solving problems surrounding their most basic and most important economic activity — food sharing. Once we became adapted to forming social contracts, we used them to solve increasingly complex social problems, leading, eventually, to our modern day economic organisation.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my partner Carly Stoneman for supporting me in so many ways while I was writing this dissertation. Life would have been hard without you.

I would also like to thank my Supervisor Joanna Bryson for all of the advice and patience that you have given me. I probably haven't been the easiest student to supervise.

Finally, I would like to thank my colleagues Marios Richards, Dominic Mitchell and Paul Rauwolf for all of the enlightening discussions we have had. I'm certain that this dissertation would not have been the same without you.

CONTENTS

CHAPTER 1
INTRODUCTION

I would like to see anyone, prophet, king or God, convince a thousand cats to do the same thing at the same time.
— Neil Gaiman *The Visionary Cat, in Sandman #18: "A Dream of a Thousand Cats"*

Evolutionary biologists have long believed that the process of evolution by natural selection is dominated by competitive forces. From Herbert Spencer's 'survival of the fittest,' to Tennyson's 'nature is red in tooth and claw[1],' this perspective has inspired many of the famous aphorisms that we have come to associate with natural selection. But this view is changing. We are now beginning to understand that, as well as competition, there is another potent force in evolution — the force of cooperation.

This dissertation is about cooperation in our own species, *homo sapiens*, the scale of which is unprecedented in nature. In most other species, cooperation is only very small in scale, perhaps only including a few individuals or a family unit. When it is large in scale, such as with some species of social insects, it occurs within groups whose members are closely genetically related to one another. But humans are different. Humans cooperation is vast, earth-spanning even, but occurs primarily between those who are not directly genetically related. But why is this a problem at all?

Human cooperation is problematic for evolutionary theory because it seems to contradict the most basic tenet of natural selection. We expect to find individual members of a species who have been subjected to the force of natural selection to act to maximise their fitness. But when humans act, we do not seem to maximise this quantity, at least not directly. Rather, humans often act to increase the fitness of other, genetically unrelated individuals. What is more, we cannot dismiss human cooperation as

[1]Strictly speaking, Tennyson's poem was written before the publication Darwin's Origin of the species, but it has since become synonymous with Darwin's theory.

a minor human oddity. Rather, cooperation is the fundamental reason for our success as a species, and that, in the past few thousand years, a blink in the eye of evolutionary time, we have colonised and prospered in many of the least hospitable habitats on earth.

But what exactly is 'large-scale cooperation?' Unfortunately, defining such a thing is like trying to establish an immutable definition of *species* — whenever one is attempted it either excludes cases that we want in, or admits ones that we do not. Fortunately, evolutionary biologists have long accepted that it is quite possible to proceed with evolutionary science without a perfect definition of a species that covers all cases. Sometimes theory must bend — nature is not neat, why should we expect it to submit to our neat frameworks? So rather than attempt a cast-iron definition of large-scale cooperation, I simply describe some of its important features and give some examples.

For cooperation to be considered large-scale, it must be performed by many individuals. Of course 'many' is not a great improvement, but when we consider Williams [1966]'s famous definition of an evolutionary gene, that 'the selected entity must have a high degree of permanence and a low rate of endogenous change, relative to the degree of bias,' we must admit that, although this is not a precise definition, it is clearly a useful one.

Large-scale cooperation also involves *concerted action*. When humans cooperate in groups, we do so with the understanding that others will also cooperate, and that our behaviours will complement one another as we work towards a common end. Therefore, the notion of 'common end' is also important for large-scale cooperation. For example, when we go to war, it is possible to describe human behaviour so that each individual works towards the larger goal of winning that war. Of course, this does not mean that we expect all behaviour to be aimed towards this at all times, or, as some of the theoretical models suggest, that all humans must act in the same manner. Rather, it is enough that we can use it to predict some behaviour some of the time.

It is important to conceptually distinguish large-scale cooperation from it's smaller-scale counterpart. Small scale cooperation takes a narrow view of human interactions and applies to ones that are largely exclusive. An individual trade or a dyadic relationship might both be considered small-scale cooperation. In these examples, we are primarily interested in the relationship between the two individuals and the actions they take, and we do not consider the effect on others within the group. But there is a great deal of human cooperation that does not conform to these conditions. To give a modern day example of this, consider two individuals cooperating within a company. A worker might never meet the person who deposits wages into their bank account. In the context of a dyadic interactions, this relationship is incomprehensible. It appears that some individuals within a community behave altruistically by giving away their property to other individuals who do not pay them back, or even know who they are.

Of course, we know that interactions between these individuals are not exclusive at all. We can only make sense of them by couching them in terms of what the rest of the group is doing. The money that was deposited into their accounts was not a 'gift,' since the money was never owned by the paymaster, but was simply handled by them. The money was not 'owned' by anyone at all, rather, its distribution is a consequence of some set of rules that the group has established. Presumably, the money was offered as remuneration for services rendered for the good of the group, however that is defined, and passes from 'group ownership' to the ownership of the workers through the ritualised process of 'payment.'

1.1 The Altruism Polemic

Is there an analogue of this in other species? Indeed there is — it is much the same for ants within a colony. More often than not, it makes little sense to investigate the isolated behaviour of an individual ant without considering it with respect to the wider context of the rest of the colony. Only by considering how an individual ant aids its nest as a whole can we comprehend the evolutionary logic of its actions [Wilson, 2012]. The organising principle for ants — the reason that their behaviour is maintained against selection and mutation — is genetic altruism. Because ants are closely genetically related to those who benefit from their behaviour, they are capable of evolving adaptations to sacrifice their personal fitness, for the fitness of the relatives. But humans also act for some higher ends, does this mean that we are also altruistic?

This question of altruism lies at the heart of a great many scientific debates about the fundamental nature of humans. Are we ultimately selfish or altruistic? This debate has raged for centuries in many different subjects, most famously pitting Rousseau, who argued that man is good if he is not corrupted, against Hobbes, who argued that man is bad if he is not tamed. While the content of this debate has changed little, the standard of evidence has, and now many behavioural economists travel around the world directly testing the hypothesis in laboratory conditions by giving people money, offering them the chance to give it away, and observing their behaviour [Fehr et al., 2002, Camerer and Fehr, 2006, Fehr and Schmidt, 1999, Fehr and Gächter, 2002, Gintis et al., 2003, Henrich et al., 2005, Herrmann et al., 2008, Bernhard et al., 2006, Fowler et al., 2005, Egas and Riedl, 2008, Marlowe et al., 2010, Chaudhuri, 2011].

This community speaks out directly against what they consider to be orthodox economic and evolutionary theory. They argue that both of these theories derive from the empirical assertion that humans are in some way fundamentally selfish — an assertion which they argue they have demonstrated to be false through their empirical testing. For example:

> Economic reasoning is typically based on the self-interested hypothesis,

> ie, on the assumption that *all* people are *exclusively* motivated by their material self interest... (D)uring the last decade experimental economists have gathered overwhelming evidence that systematically refutes the self-interested hypothesis and suggests a substantial fraction of people exhibit social preferences. [pp.1 Fehr and Fischbacher, 2002]

and

> The explanatory power of inclusive fitness theory and reciprocal altruism convinced a generation of researchers that what appears to be altruism — personal sacrifice on behalf of others — is really just long-run self-interest... However, recent experimental research has revealed forms of human behavior involving interaction among unrelated individuals that cannot be explained in terms of self-interest. [pp.153-154 Gintis et al., 2003]

The agenda is clear — standard economic and evolutionary theory predicts that humans ought to be selfish, but that when we observe the actual behaviour, they are often found not to be.

Is it true that this emerging science of human cooperation has falsified the standard doctrines of economic theory and evolutionary biology? I argue not. In economics, the fundamental organising principle is *subjective preference*, which can be either 'selfish' or 'altruistic' [Binmore, 1998, Samuelson, 1948, Binmore, 2005a]. Standard theory relies on the notion of revealed preference, which makes no prediction of what it is that humans ought to prefer. Rather, it assumes that we can discover what individuals prefer by observing how they act [Samuelson, 1948, Sen, 1973, Binmore, 1998, 2005a]. All it assumes is that, within some contexts, humans consistently prefer some outcomes over others, and that, all else being equal, will act to achieve those outcomes. It does not worry about what those preferences ought to be — the preference to be altruistic is entirely consistent with standard economic theory.

The traditional subject matter of economics is means and not ends. Economists investigate what means humans use to achieve our ends, not what our ends should be. Of course, there is no doubt that some economic theory assumes that individuals aim to maximise their material well-being in some situations, but that is only because we observe this in the real world. It does not assume that all individuals consistently act to increase their material well being all of the time. For example, it seems reasonable to assume that, all else being equal, an individual in the market place would prefer to pay less for a good than more. It is equally reasonable to assume that once this good is purchased, it will be given as a gift. Neither of these assertions undermine basic economic theory.

But these modern attacks on economic theory should be viewed in the contexts of debates going far back in time. Fifty years ago, Cook [1966] defended economic man against substantivist scholars, whose philosophy of anthropology derived from Polanyi [1944, 1957]. Knight [1941] defended economic orthodoxy, and engaged the substantivists in a debate, which is reviewed by Binenbaum [2005]. My discussions on method in this dissertation is inspired by this exchange.

In evolutionary biology, the organising principle is *inclusive fitness*. Inclusive fitness is the theory that organisms are evolved to favour those who are genetically related to them. A gene can propagate itself through two routes. The first is by increasing the likelihood that the body in which it resides will survive and reproduce. The second is by increasing the reproductive success of close relatives who also possess copies of the same gene. Inclusive fitness is the combination of both of these effects.

Inclusive fitness theory is the modern formalisation of the research program that began with Fisher [1930], and carried through Williams [1966], Hamilton [1964], Dawkins [1976] and more recently Gardner and West [2014]. In this framework, we try to understand the evolutionary process from the perspective of changes in gene frequencies across generations. One of its most important finding is that genes that correlate with altruistic behaviour can spread, so long as there is a sufficient probability that the gene is also present in the individual who receives the help. For example, consider the aforementioned ants — the only reason altruistic ant behaviour can evolve is because the benefit goes to other ants who are genetically related. This is why the quote above is confusing, when it describes self-interest, it seems to imply that this is from the perspective of the individual, rather than of the gene. But inclusive fitness theory is from the perspective of the gene, and not from the individual.

In response to these attacks on standard evolutionary and economic theory, many scholars have defended the orthodox by pushing back with a series of counter-claims. Some claim that the pro-altruism scholars have misinterpreted standard theory and have attacked a straw man [Binmore, 2005a, West et al., 2011, 2008], while others that there is insufficient empirical evidence that humans really are altruistic [Guala, 2012a, Burton-Chellew and West, 2013, Binmore, 2005a].

I argue that the emerging literature on human cooperation has made two major mistakes. The first has been in attacking orthodox economic and evolutionary theory. As it stands, I do not believe that these attacks, or the subsequent rebuttals, have achieved much to further our understanding of the evolutionary forces that have selected human cooperation. Of course, as with any new science, there are likely to be issues that need ironing out, and the science of human cooperation brings together several traditional disciplines of biology, anthropology and economics, so there is bound to be some, if not a great deal, of disagreement. But I do not believe that many of the exchanges that have been made so far have been as valuable as they could have been.

The second fundamental flaw with this literature has been its approach to linking the model with empirical data. My vision of a well working scientific community is one where there are close links between the empiricists and the theorists. Of course, it can be useful to divide our labour and specialise in different areas, but if this occurs, then we must be extra careful that the two sides of the community communicate with one another. In the evolution of cooperation literature, this has not occurred. The assumptions of many of the models of human cooperation emerge largely as a response to other theorists, rather than from some empirical issue that our theory currently cannot explain.

The travesty is that when compared with collecting empirical data, theory is relatively cheap and easy to produce. The flexibility of theorists means that it is they who should give way to the empiricists and not the other way around. Theory is infinite, so we must constrain it with empirical observation. Theorists best serve the scientific community by clarifying the consequences of commonly held assumptions so that the community can better make predictions to test theory, and also by ensuring that all empiricists within a scientific community are expressing the same idea, or different versions of the same ideas for comparison. Many modern models of cooperation achieve the opposite effect. Some are so complicated that it is impossible to understand the principles that drive the model, meaning that it is unclear how they clarify the consequences of commonly held assumptions, or how they help make predictions about the world.

To highlight the mismatch between the standard theoretical approach with empirically observed phenomena, take again the example of the individual working as part of a company. We would hope that the standard theory of cooperation might apply to this extremely common case. Yet it does not. The model assumes that human communities produce unexcludable public goods. But in general, companies produce private, excludable goods. The individual worker does not expect to suffer a net cost for turning up to work — they expect to be paid, meaning they expect a net benefit. But according to the theoretical model of public goods, the individual who cooperates with anyone *must* cooperate with everybody. That is, the good that they produce must go equally to everyone. *If* companies had to pay wages to all workers regardless of whether they actually turned up, then we would be justified in modelling this as a public good and questioning why humans produce it. But they do not. So I see no reason why the community has spent so much effort modelling public goods, and why solving this model has become the 'problem of cooperation,' since it does not seem to apply to the most basic building block of our modern economy.

This dissertation is a theoretical study of the evolutionary causes of human cooperation. Therefore, like the models outlined above, I too use mathematical models to this end. But rather than aiming the models towards other theorists in the effort to explain whether humans are really altruistic or not, I approach modelling by helping to better understand problems that the empirical literature highlights about human cooperation. In particular, I aim to explain some of the phenomena that anthropologists have observed in small-scale societies. Why solve these problems? Because it is impossible to directly observe our evolutionary ancestors, and the fossil and archaeological record can only give us indirect evidence of how they solved social problems, so I assume that modern hunter-gatherers are our best source of evidence for how human evolved to cooperate [Hill, 1982].

In modelling, my chief aim is clarity over complexity. Recently, models of cooperation have become increasingly baroque, often in the name of 'realism,' but have often confused, rather than clarified hypotheses. I believe that the purpose of modelling is clarity, and not the realism that these models so often claim. We make models to better understand the consequences of theory, so we can use it to predict behaviour. Of course, we do not expect to be able to predict all behaviour at all times. My goal is to create theory that predicts some behaviour within a given community much of the time. We must admit from the outset that there is very little that is 'realistic' about game theory models at all. They are simply tools to extend our deductive capacities beyond what is immediately apparent. Since my ultimate aim is in testable hypotheses, if it is possible to communicate something deductively using natural language, and that method is simpler, then I always do so. But when the complexity of the logic necessitates, I use formal, mathematical theory.

1.2 Social Contracts

My thesis is that humans have evolved to cooperate by using social contracts. Social contracts are solutions to problems in 'the game of life' — the theoretical set of all social interacts that humans must navigate in our everyday life. By 'solution,' I mean that the contracts suggest an action to all players who play the game of life — they tell us how we 'ought to' behave. But basic evolutionary theory tells us that, all else being equal, humans ought to behave in ways that are individually beneficial, and so this theory of the social contract also assumes that social contracts are also beneficial from the perspective of the individual.

There are two distinct processes that form contracts. First, there is what I call 'spontaneous formation.' This is achieved when a groups of humans come together to negotiate solutions to shared problems. Of course, by 'spontaneous,' I do not mean to imply that all culture is invented from scratch with a specific plan in mind. In most cases, new contracts are likely to be heavily reliant on older ones. The second type of process of cultural change is the 'gradual' process of cultural evolution [Boyd, 1988]. The theory of cultural evolution have recently been given a theoretical overhaul, inspired by standard models of biological evolution. In this theory, rather than genes being copied in subsequent generations by the process of sexual reproduction, it is strategies being copied by the process of social learning.

What is the chief difference between these two theories of cultural change? It is primarily in whether we expect change to occur gradually, like in evolutionary biology, or instantaneously. To give an example, suppose that within a modern company, individuals find that they are more productive when they perform a given task in a different way. They pass this information along to other individuals within the company, who adjust their behaviour accordingly. This would be an example of gradualistic cultural evolution. However, it is also possible that the company moves premises. This change occurs 'all at once;' we do not expect individuals to turn up to the old premises and have to learn by diffusion that the premises have changed.

While the difference between the two types of cultural change may seem obvious when they are spelled out like this, the fact is that the cultural evolution literature so far has embraced a gradualistic theory of change, almost to the complete exclusion of any spontaneous change.

1.3 The Wider Context of this Research

By applying evolutionary theory to the scientific study of human morality (which I consider equivalent to the social contract), evolutionary theory pits itself against centuries of received wisdom about the dual nature of man — his body and his mind. Evolutionary theory makes no such distinction between these two causes of human action and eschews all talk of 'free will.' It is interesting, then, that many of the methods employed by modern evolutionary biology originated in economics, which, as a theoretical science, relies primarily on the idea of free will and individual autonomy. Economic theory rests on the foundation of subjective desire, and assumes that humans act to maximise this quantity. Who decides what is best? Each individual is arbiter of their own good.

In many ways, evolutionary theory is a natural continuation of economics. It makes a firm distinction between 'proximate' explanations and 'ultimate' ones. Proximate explanations are immediate ones that make predictions in response to an immediate situation. For example, the empirical observation that individual humans try to make money is proximate observation. Ultimate explanations explain how behaviours lead to greater evolutionary fitness. Therefore, the 'subjective desire' that acts as an ultimate cause in economics can be easily enough replaced by 'evolutionary fitness' as a more ultimate cause, making micro-economic theory a branch of evolutionary biology.

It is when we attempt to predict human behaviour that we see the importance of evolutionary theory. The idea of a subjective desire does not predict anything. In fact, it is the opposite of prediction — it says that the source of all action is some entirely inaccessible and unpredictable entity. But while it is no doubt true that some aspects of human action are not predictable, there is a great deal that is. From the biological need to eat, stay sheltered, procreate, to our cultural behaviour, our language and so on. Evolutionary theory, along with a theory of culture, helps us to predict human behaviour.

Not only do social scientists seek to apply the theory of natural selection to understand the evolution of human behaviour from a biological perspective, but also many have sought to apply it to explain the evolution of human culture. It is not widely known that the term 'genetic' does not have its roots in evolutionary biology, but rather, in the study of the evolution of language and culture. To my knowledge, one of the first attempts to explain culture as an evolutionary process was by the philosopher Nietzsche [1887] in his book 'The birth of tragedy and the genealogy of morals.' Notice that the word 'geneology' is used here long before it was applied to any biological phenomena. But probably the most famous direct use of the theory of cultural evolution is by the philosopher Spencer [1895]—the scholar who coined the term 'survival of the fittest.' With typical Victorian zeal, Spencer argued that society was like a 'giant organism' that moved from a state of simplicity to one of relative complexity through the process of natural selection. He called this process 'social Darwinism.'

This highly influential view was subsequently adopted by many economists and philosophers, particularly those who lean towards a libertarian view. They, like Spencer, argue that governments should do as little as possible to interfere with the 'natural' process of the market, which they argue is the fundamental driving force of social change. In this view, natural selection is a 'tool' that we can use to create better societies. To highlight the modern importance of these theories, consider the fact that Margaret Thatcher was heavily influenced by philosopher and economist F.A Hayek, who himself wrote in this libertarian tradition.

I do not think that I am stick my neck out too far by saying that I believe that a great many of the investigations into the nature of human altruism have been made through the desire to counter Spencer's claim that selfishness is 'superior,' because it leads to more adaptive outcomes, by demonstrating that altruism can also be selected. If this can be shown then, by the same logic, altruism can be superior.

I know of no better demonstration of this than the theories of economist and philosopher Hayek [1988] and the current evolution of cooperation scholars Boyd and Richerson [2005]. What is striking is that these two sets of scholars both name their theories 'cultural group selection,' but as I show, they take precisely the opposite position on the question of cooperation. Hayek conceives of two types of morality — that which applies to our 'micro-cosmos' — our friends, family and close colleagues, which he argues is naturally altruistic — and that which applies to the 'macro-cosmos' — wider society — which he argues has evolved the morality of what he calls 'selfishness'. He concludes that:

> If we were to apply the unmodified, uncurbed, rules of the micro-cosmos (i.e., of the small band or troop, or of, say, our families) to the macro-cosmos (our wider civilisation), as our instincts and sentimental yearnings often make us wish to do, we would destroy it. Yet if we were always to apply the rules of the extended order to our more intimate groupings, we would crush them. So we must learn to live in two sorts of world at once. To apply the name 'society' to both, or even to either, is hardly of any use, and can be most misleading. [p.18 Hayek, 1988]

Hayek argues that we began as altruists in small group, and had to learn selfishness through the process of cultural group selection.

Boyd and Richerson argue the precise opposite of this [Boyd and Richerson, 2005]. They claim that it is altruism towards fellow members of the group that best explains human success, this then spread through the process of cultural group selection. Both sets of scholars claim that the respective morals of selfishness or altruism drove human cooperation towards its current state.

The major point of contention between these two sets of scholars is their view on what a community is. Hayek [1988] argues that humans are evolved to take part in small-scale communities, and that larger scale communities, like that on the scale of countries, are incomprehensible to us. To highlight this point, Hayek did not even like to use the word 'economy,' which derives from the Greek to mean 'household management,' to describe the order that is brought about by the use of money. Aristotle originally used the term as a metaphor of the state, implying that all peoples within a state have a congruent set of goals as they do within a household (350 BC). But Hayek preferred the term 'catallaxy,' which does not imply that a society has a congruent set of goals. Catallaxy means 'the order brought about by the mutual adjustment of many individual economies in a market.' But Boyd and Richerson do not share this vision. They believe that it is appropriate to represent entire communities as homogeneous units working towards common, community ends, as represented by the assumptions in their models [Boyd and Richerson, 2003].

But I argue that both sets of scholars are incorrect. Contra Boyd and Richerson, I do not believe that there are many social interactions that occur on a large-scale that require unconditional altruism. Of course, there are always the examples like stopping to give another individual directions. But is this sort of action really evidence of altruism? What is the marginal lifetime fitness benefit of not giving directions? Is it possible that the benefits of the very small chance that in giving directions, which costs essentially nothing, you make a friend, and that these benefits outweigh the costs? If an individual went out of their way to give out directions, voluntarily taking time out to help strangers at a real cost to themselves, then this would require explanation. But I do not believe that the vignettes about occasional acts of kindness such as this should be the focus of human cooperation studies. Contra Hayek, I do not believe that behaviour within a group can be considered 'altruistic.' This is because when interactions repeat, and future behaviour is dependent on past behaviour, then single actions within groups resist simple categorisations like 'selfish' and 'altruistic.'

1.4 The Specific Claims of this Dissertation

My thesis is that, in order facilitate cooperation, humans have learnt to enact social contracts. I divide this dissertation into four main chapters, along with a topic review, where each chapter supports this claim.

In chapter 2, I review the literature and outline the method that I use to investigate the evolution of social contracts.

In chapter 3, I argue that social contracts evolved initially to order to solve the social problems that our hunter-gatherer ancestors faced with respect to ensuring a consistent diet of highly nutritious meat. The best way for a group of individuals to do this is by having its members go out to hunt large-game and bring that game back to the group to share. However, from the perspective of the individual, we have to explain why this, and not a more selfish strategy, is selected. For example, if one individual out of a group decides not to hunt, but instead saves their energy by hanging around the camp, and then gets to eat some of the meat that others have worked to catch, then that individual would likely have a higher lifetime fitness.

I argue that the invention of social contract of property rights is vital in understanding the evolution of food sharing. Property rights underpin both the theory of reciprocal altruism, as well as some of the alternative theories of food-sharing such as tolerated theft. I argue that what we actually observe is not one, but a selection of different methods of food-sharing based on different social contracts. One important contract that has been widely observed is where taking part in a hunt gives property rights over the catch, even that of others, so that food is not shared dyadically, but through a group-wide process of distribution.

In chapter 4 I demonstrate how the ability to form social contracts based on the idea of property might have invaded from rarity, and also how humans have evolved to use property to encourage or dissuade individuals from engaging in pro or anti-social behaviour. I focus on an example of a system of fines that are in place for the Nuer — a population of desert pastoralists — who have adopted such a system.

In chapter 5, I examine ways in which humans have evolved to bond socially. While not strictly speaking about social contracts, I argue that this an important method that humans use when establishing new cooperative ventures, which is important due to the relative fluidity of human populations. I argue that we bond by jointly engaging in costly but otherwise evolutionarily pointless activities. The model in this section is a formalisation of Zahavi and Zahavi [1997]'s theory of bond testing.

Finally, in section 6, I bring all of these parts together to explain how humans learnt to use social contracts in order to facilitate large-scale cooperation. I argue that the adaptations that we evolved to use in food-sharing arrangements became adapted to a wider context of economic goods. However, it is not that we used the exact same contracts that we created in food sharing in order to establish more complex social contracts. Rather, I argue that we evolved to be plastic in our ability to establish contracts, and that we are capable of negotiating many types of contract.

The theory in each section is inspired by the ethnographic literature on cooperation. In chapter 2, I outline my method for validating theoretical models against ethnographic, which I use in every main chater in this dissertation.

CHAPTER 2
TOPIC REVIEW

Frustra fit per plura, quod potest fieri per pauciora — It is pointless to do with more what can be done with fewer.

— William of Occam

It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.

— Albert Einstein *On the Method of Theoretical Physics. The Herbert Spencer Lecture, delivered at Oxford (10 June 1933)*

Whereof one cannot speak, thereof one must be silent.

— Ludwig Wittgenstein *Tractatus Logico-Philosophicus*

2.1 Introduction

My purpose in this chapter is twofold. Firstly, I introduce the theoretical framework that I use throughout this dissertation. I use explicit examples to motivate the different models, and also to highlight any limitations they might have. Secondly, I explain how this theory helps us to understand human behaviour.

My thesis is that humans have evolved to form social contracts, and so all the parts of the framework that I describe helps me to develop this theory. My primary tool is evolutionary game theory — a formal, mathematical method for studying social interactions. Evolutionary game theory was initially used to describe biological evolution [Maynard Smith, 1974], but has subsequently been developed into a model of cultural evolution as well [Dawkins, 1976, Boyd, 1988, Binmore, 1998, 1997].

In the second part of this chapter, I explain how we can use the theory to help understand human behaviour. I argue that to evaluate a given model of social contract, we should work as closely as we can to the ethnographic literature. I outline the method that I use throughout this dissertation, which is to run through all of the assumptions that I make in the models, and find evidence from the ethnographic literature to either confirm, disconfirm, or say where there is insufficient evidence to validate each of them.

2.2 The Utility of Simple Models

Before I discuss any models at all, first I describe the process of modelling in general. The models that I analyse in this dissertation are much simpler than the phenomena that they represent. The individuals who populate the model are quite stylised - they usually have simple traits, time is discrete, stochastic effects are minimalised and so on.

Simple models like this never come close to capturing all of the detail in any particular situation. But what they lose in specificity, models gain in tractability. Models are like maps — they are most useful when they capture the details of interest and ignore the rest. As McElreath and Boyd [2008] put it ‘A map of subways help a person to navigate a subway. A map of the street helps them navigate the surface. A map of both subways and streets would be too crowded and confusing for either task.’ Models are like maps for understanding the consequences of adopting a small number of assumptions, and maps are always better when they contain less information. I think that Lewis Carroll makes the point the best:

> “What do you consider the largest map that would be really useful?”
>
> “About six inches to the mile.”
>
> “Only six inches!” exclaimed Mein Herr. “We very soon got to six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of a mile to the mile!”
>
> “Have you used it much?” I enquired.
>
> “It has never been spread out, yet,” said Mein Herr: “the farmers objected: they said it would cover the whole country, and shut out the sunlight! So

> we now use the country itself, as its own map, and I assure you it does nearly as well." [Carroll, 1893]

The models that I investigate in this thesis are only as complicated as they need to be to capture the structure of the problem. They always omit features that are important in reality — such omissions are purposeful.

2.3 Game Theory - Brief Introduction

Originating in economics half a century ago, game theory is a formal approach to modelling social interactions between rational agents. The idea is simple. We imagine that social interactions are like games played between the members of a society. Within these games, each player can make one of a number of moves, which represents the actions they can take in real life. The result of the game depends on the actions taken by all of the players within that game. We also assume that each player has an ordered list of preferences over the possible outcomes, with a high preference meaning that, all else being equal, they would choose that outcome over others. Normally, this information is represented by giving each player a *payoff*, represented by a number, for each of the possible outcomes.

The organising principle of game theory is called 'rationality.' Rationality simply means that we assume that each player in the game always plays to maximise their payoff. Because the payoffs always represent the subjective desires of the players, and not, as they are often confused to represent, material reward, this framework does not assume that humans are necessarily 'selfish.' If a player enjoys giving away money, then this would give them the highest payoff.

The idea of resting economic theory on this bedrock of subjective desire was first conceived by Austrian economist Menger [1871] to explain how prices form. Menger argues that the price of an economic good on the free market is primarily caused by the subjective desires of the consumers choosing that product over another, and not, as had been previously thought by Marx and Engels [1847] and others, that it was determined by the amount of labour that went into constructing it.

Is this theory of action tautological? The answer is yes it is. If we say that all action is performed because the performer wanted to do it, we have not said anything empirical about the world. However, it is important to realise that game theory is not a method for determining the subjective desires of people. Instead, the purpose of game theory is to predict how people should act *given* their preferences.

Neumann and Morgenstern [1947] formalised this and Nash et al. [1950] discovered the idea of an equilibrium. An equilibrium in game theory is a set of strategies that, assuming all other players followed their strategy, no player would benefit from unilaterally deviating. Subsequently, there have been many refinements to this equilibrium concept, including subgame perfect equilibria, where players are assume to maximise their utility in each subgame, trembling hand equilibria, correlated equilibria and more.

One refinement of particular interest to this dissertation is the *evolutionary stable strategy*, first conceived by Maynard Smith [1974], which is an equilibrium in an evolutionary game. In this model, rather than thinking about rational agents playing games to maximise their utility, we assume that the population is playing games to maximise their fitness. We assume that there is a pool of players from which we pick subsets to play games. These subsets engage in some game, receive their payoffs, and return to the population pool. However, we assume that their performance in the game determines the fraction of the population that come to play a specific strategy, so the more successful strategies are better represented in the following generation.

More recently, this model of evolutionary game theory has been applied to *cultural evolution* [Dawkins, 1976, Boyd, 1988, Binmore, 1998, 1997]. This model is based on the same set of assumptions as the biological model, but rather than the strategy being inherited via genetic transfer, it is inherited by cultural learning. In this model we assume that the 'fitness' of a strategy is still equal to its payoff, but we assume that a greater portion of the population choose to adopt the strategy with the highest payoff in the following generation, thus forming an evolutionary model of culture which is exactly analogous to the genetic type. Therefore this theory of cultural evolution is completely conceptually compatible with a rational theory of behaviour — it simply outlines the process through which members of the population learn that some strategies are more successful than others. That is, rational choice theory assumes that players just find the correct solutions, so that there is no scope of out of equilibrium behaviour. An evolutionary theory explains how a behaviour might be out of equilibrium at a particular time, but predicts that it will move towards the equilibrium as time passes.

2.4 Game Theory and the Social Contract

As described in the previous section, game theory describes human behaviour by assuming that within specific contexts, humans act to maximise their payoff, where the payoff represents their subjective desire for some specific outcome. My task here is to describe how we can use this idea to explain how social contracts remain stable over time in human societies. This method of social science has deep roots. As far as I am aware, the first scholar to use the deductive logic that is central to game theory is the Scottish philosopher David Hume [Binmore, 1998, Vanderschraaf, 1998]:

> Two men who pull the oars of a boat, do it by an agreement or convention, although they have never given promises to each other. Nor is the rule concerning the stability of possessions the less derived from human conventions, that it arises gradually, and acquires force by a slow progression,

> and by our repeated experience of the inconveniences of transgressing it. [Hume, 1738, pp. 490]

As Hume explains, conventions and norms govern a wide range of social phenomena, from the arcane table manners to our attachment to small bits of paper that we carry in our wallets (or even to numbers that flash up on a screen at the cash point). From the side of the road that we choose to drive to the vagaries of fashion. Life, according to Hume, is filled with coordination problems that we solve every day without thinking about it.

One approach to investigating these social norms was pioneered by mathematician, economist and philosopher Ken Binmore [1998, 1997, 2005b]. To describe the problems that require morality to solve, Binmore adopts the term *game of life* — the set of all human interactions that require our coordination and cooperation. Although we call it a 'game,' the theory is nothing at all to do with games in the usual meaning of the word. Rather, 'game' is used because it is morally neutral, and allows us to investigate human morality as a biologist studies a species, rather than as a censor. Life itself is not a game, but we study it *as if* it were a game — as if we didn't care about our conclusions.

In the game of life, players must learn to adapt to the choices of other players, just as other players must adapt to their own. The result of the game is always a combination of choices of the players, along with some rules of how the game operates. Where do these rules come from? They represent the external constraints to the problem at hand. For example, when two cars want to pass each other on the road, each must make a definite decision of whether to drive on the left, or to drive on the right. The result of this passing event is determined by the choices of the individuals driving the cars, as well as the physical constraints of the road, namely, that if two cars drive on opposite sides of the road (from the perspective of the driver), they will both crash into each other. Humans are free to drive on either side of the road, and there is a constant threat of another driver choosing to drive on the other side. Yet, not only do humans choose to drive, such incidents rarely end in head-on collisions. But why would this be? Why does every driver trust every other driver on the road? The answer, according to Binmore, and Hume before him, is that humans rely on the use of social contracts.

Game theory formalises our understanding of these interactions. First, we find a phenomenon we wish to describe. Next, we propose what possible actions the players in the game might perform. For example, in the driving game, it might be to drive on the left, or to drive on the right. Next, we propose a set of payoffs that exist over the possible outcomes. These payoffs represent what it is that the individuals want. At the moment, we can be quite relaxed about how proximate this 'want' is. In the driving game, for example, players want to pass one another on the road, which is equivalent to saying that they don't want to crash into each other, although ultimately, we want to link this 'want' to evolutionary fitness. Next we calculate the equilibria of the game, defined to be a set of strategies that no player benefits from deviating from. In this theory, the set of possible social contracts is the set of possible equilibria.

A Nash equilibrium is simply a set of strategies that no player benefits from unilaterally deviating from. Formally, if we call the game G, the strategy set S_i represents all of the strategies that player i can play. We call $S = S_1 \times S_2 ...$, and $\sigma_i \in S_i$ is the strategy that individual i plays, and finally $w_{\sigma_i}(\sigma_{-i})$ is the payoff that individual i gets for playing σ_i when the rest of the population play σ_{-i} (σ_{-i} just means all the strategies except for σ_i).

Definition 1. *In game G with strategy set S, the set of strategies $\sigma = (\sigma_1, \sigma_2 ...)$ is a Nash equilibrium when, for each i,*

$$w_{\sigma_i}(\sigma_{-i}) \geq w_{\sigma_i^*}(\sigma_{-i}),\ \forall \sigma_i^* \in S_i \tag{2.1}$$

For example, using the theory introduced above, we can build our model of the driving game. We say that each player chooses a strategy from the the following set $S = \{L, R\}$ (meaning 'left' or 'right'). Their payoff depends on their own choice and the choice of their partner and is represented in the following matrix.

	Left	Right
Left	1	0
Right	0	1

In this game, the players are indifferent about whether they pass on the left or the right, but they prefer either outcome to crashing into one another.

Proposition 1. *In game G, there are two Nash equilibria — $\{L, L\}$ and $\{R, R\}$*

Proof. Let $w_L(L)$ be the payoff of an individual who plays L when their opponent plays L (the subscript is the move of the focal individual). Then we can see that

$$w_L(L) = 1 > 0 = w_R(L) \tag{2.2}$$

Therefore, neither player benefits from deviating from L when their partner is playing L. The same proof can be used to show that R is a Nash equilibrium. □

This theory explains how a social phenomena — in this case the rules of the road — can be explained in terms of the individual desires of the drivers, along with some rules about how they ought to behave. In this case, the rule is represent by a Nash equilibrium, for example, both drivers drive on the left, or both drivers drive on the right. The idea of an equilibrium is that when playing the driving game, neither player benefits from deviating from the strategy that the social norm suggests they play.

However, this theory of human behaviour also has some limitations, and I believe it is best if I highlight these immediately. The main limitation is that it assumes that humans are always aware of the payoffs they will receive from playing one particular strategy over another. In many, and even most situations, this will simply not be the case. Rather, humans often act out of force of habit, and are unsure about the rewards of acting otherwise. The second limitation is in how we use this theory to interpret and ultimately predict human behaviour. This second limitation is sufficiently confusing that it requires its own section.

2.4.1 My Preference, Our Contract

Many of the social sciences tend to take a rather strong position on the question of whether people are products of societies, or societies are products of individuals. There are many examples of individual versus group wide explanations of social phenomena, but probably the two best know are Durkeimeian social facts and individual-level micro economics [Durkheim, 1895]. However, since the days of Darwin [1859], and in particular Fisher [1930], evolutionary biologists have understood that the answer must be both. Therefore, in evolutionary modelling, this classical dichotomy disappears. Individuals affect the group-level phenomena, but the group level phenomena also affects individuals. The two levels exists seamlessly in a single theory. In my own approach, the individual is holds the preference (or the gene), but the society holds the social social contract.

When interpreting human behaviour, it is easy to become confused over the difference between a contract and a preference. A preference is what an individual wants. It is what they aim for. But in order to obtain the best outcome, they must consider more than just their own preferences, they must also consider how other individuals around them will act, meaning that player must also consider the preferences of others. Recall from the model of the driving game above, that both players are indifferent about which side of the road they pass each other. This indifference is expressed in their preferences. But a theory that is based solely on preferences cannot predict or explain the fact that people in London drive on the left and people in Paris Drive on the right. In order to explain this, we also need the idea of the social contract. In Paris, there is a different social contract as there is in London. When we observe driving, we observe the contract, and not the preferences.

Preferences are held by individuals, social contracts by societies.

So in the driving game, we cannot infer that Parisians prefer right to left, simply because that is what we observe. Rather, it is quite possible that they are indifferent over the which side of the road they pass, and that it is the contract that is forcing them to make that choice. This fact is so obvious in the driving game that it hardly seems worth writing down, but for the fact that so many scholars, in particular, the group of scholars who argue that humans are *other regarding* get this point wrong in more complex settings (Guala [2012a], Binmore [2010] both review this, and I investigate these assumptions in section 6.2.2).

Part of the problem is in agreeing exactly what counts as an ultimate preference, and what counts as a contract. For example, many scholars argue that helping others is a preference and not a contract [Henrich et al., 2005, Fehr et al., 2002], while others argue the opposite [Binmore, 2005a]. I argue that one of the problems with the current framework of rationality is that there is no way out of this knot. If I give a present to another individual, that action is consistent both with me desiring to give out of the goodness of my heart (preference), and because I expect you to give me a gift in return later (contract). However, when we consider motivation from an evolutionary perspective rather than a subjective desire perspective, as I do in section 2.4.3, then we see that this problem disappears. We can theoretically measure the affect that an action has on lifetime fitness, in a way that we cannot with subjective desire. From a practical perspective, it might be difficult to measure the fitness effect of an action, but at least it is theoretically measurable, unlike subjective desire, which is theoretically impossible to measure.

2.4.2 Coordination and Punishment

The driving game is often called a *coordination problem*, because what is required is that both players coordinate their behaviour in order to solve a problem. But surely there is more to human morality than solving coordination problems — doesn't morality imply some sort of punishment for bad behaviour? The answer to this is both yes and no. Yes punishment is an important aspect of the problem. But no it does not require anything other than coordination.

Saying that other players *punish* here is somewhat misleading. The word 'punish' implies that the punisher somehow treats that as an end, that they 'want' to punish. Indeed, there are some scholars who argue that this is literally true — that humans have evolved to want to punish people who deviate from the norm [Mathew and Boyd, 2011, Boyd et al., 2010, Fehr et al., 2002]. However, although it might not seem like it, in the model of the driving game above, both players 'punish' each other for playing the wrong strategy. In this case, they are not doing it because they personally gain any delight out of hurting their partners, rather, they are simply attempting to maximise their own payoff. It just so happens that, by maximising their own payoffs, they are also punishing those who defect against the norm. Therefore, punishment does not need to be purposeful. It can be incidental.

So how does punishment apply in games other than the driving game? We can once again look to Hume for an explanation:

> I learn to do service to another, without bearing him any real kindness:

> because I foresee, that he will return my service, in expectation of another in the same kind, and in order to maintain the same correspondence of good offices with me or others. And, accordingly, after I have have served him and he is in possession of the advantage arising from my action, he is induced to perform his part, as foreseeing the consequences of his refusal. [Hume, 1738, pp. 521]

Aside from the archaic language, Hume's pronouncement might be straight out of a game theory textbook. He suggests that what holds cooperation in place is the implicit expectation of a favour returned. This logic has subsequently been formalised in the canonical model of human cooperation — the repeated prisoner's dilemma (RPD). The prisoner's dilemma is a game played between two individuals, each of whom can choose to play either *cooperate* or *defect*. In this game, cooperating increases your partners payoff by b at a cost of c to yourself. The payoff matrix is as follows:

	Cooperate	Defect
Cooperate	$b - c$	$-c$
Defect	b	0

where $b > c$.

Proposition 2. *In the prisoner's dilemma, the only Nash equilibrium is to defect.*

Proof. Let $w_D(C)$ be the payoff that an individual playing defect against another player playing cooperate, then we see that

$$\begin{aligned} w_D(D) = 0 &\quad > \quad -c = w_C(D) \\ w_D(C) = b &\quad > \quad b - c = w_C(C) \end{aligned} \tag{2.3}$$

So defecting always gives the highest payoff, regardless of what the opponent does. □

However, taking our cue from Hume, if we consider the repeated game, where after each round, the game is repeated with probability δ, and consider the two strategy *reciprocate* [Trivers, 1971] which cooperates until its partner defects, in which case it defects forever; and the strategy *always defect*, which always plays defection, then the game is transformed as follows

	Reciprocate	Defect
Reciprocate	$\frac{b-c}{1-\delta}$	$-c$
Defect	b	0

Proposition 3. *In the repeated prisoner's dilemma, defection is always a Nash equilibrium. Reciprocation is also a Nash equilibrium if*

$$\delta b > c \tag{2.4}$$

Proof. To show that defection is an equilibrium, notice that

$$w_D(D) = 0 > -c = w_R(D) \tag{2.5}$$

To show that reciprocate is an equilibrium, notice that

$$w_R(R) = \frac{b-c}{1-\delta} > b = w_D(R) \tag{2.6}$$

is true when the conditions in the proposition hold. □

The purpose of this section is to explain why it is important to conceptually distinguish a preference to punish from a contract where punishment emerges as an equilibrium. In the RPD, there is no explicit desire to punish anti-social behaviour. Rather, when such punishment occurs, it does so because that is the best strategy for that individual to perform. This is the first application of the discussion presented in the previous section — observing punishment does not imply a desire to punish. It might equally imply a contract to punish.

2.4.3 Do We Need the Concept of Rationality?

The theory of rationality based on subjective preferences expresses the belief that, in most places and at most times, humans do what they personally believe to be good for them, even though to an outsider what they do may seem bad. For example, smoking might seem bad because it is bad for your health, but we accept that people who smoke do not do so 'by accident,' rather we assume that they are capable of making decisions based on what they think is best. Why do this? Because theories based on rationality were originally used to explain how market prices form. Such a view must express that prices are formed for the demand for a certain good, where that demand is a manifestation of the subjective desires of the consumer. Prices do not form because something is objectively 'good,' but because individual consumers consider it thus.

However, from the evolutionary standpoint I take in this dissertation, there is really no need to have a bedrock of 'preferences' that individuals try to maximise. In fact, we avoid a great deal of confusion by abandoning the idea of utility altogether as a final cause and dealing exclusively with evolutionary fitness.

Of course we can still use utility as a *proximate* explanation for human behaviour [Tinbergen, 1963]. There are many situations where utility is acceptable, and where we do not profit from finding how a behaviour leads back to fitness maximisation. For example, if we assume that firms maximise profit, we do not need to explain how profit leads back to fitness in that particular situation. Likewise, in a model of the market, we might assume that various fashions exist that individuals desire to buy. There is likely to be some evolved reason why fashion is worth following, but the model of the market does not need to know this. It is acceptable, even though they do not explain the origin of fashion. But we use this method in biology. We may, for example, assume that a specific member of a species acts to maximise the amount of food they obtain over a long period. Of course, we also need to explain all of the processes that turn that food into energy, and how that energy is used to create more offspring and so on. But it is not necessary, and is probably confusing, to add these details into all explanations. This is essentially what we call the *proximate* and *ultimate* distinction in evolutionary biology. Therefore, I argue that rationality is a useful concept, as long as it complements, rather than contradicts, the theory of evolution.

2.4.4 Maximising What?

If we accept that 'utility' is just a proximate for evolutionary fitness, then we can rightly ask 'what is it that we maximise?' However, this question is quite difficult to answer. It amounts to creating a list of things that 'all people want.' There seems to be a wide spread belief that economics assumes that individuals will generally act to increase their own material wealth. While it might be true for some individuals some humans at some time, this is certainly not true of all individuals all of the time. In my opinion, there is no way that we will every get anything like a 'utility function' that applies equally to all humans at all times. Instead, I think the best thing that we can do is to observe a given group of individuals, and try to work out what it is that the seem to be working towards by observing them. For example, if a team of hunters go out into the forest and return with meat to eat, given what we know about human diets, it is safe to assume that they wanted the meat. If horticulturists plant crops and harvest them later, its reasonably safe to say that they wanted the crops. Questions like such as 'are humans altruistic' are difficult to test, and with the means that has been devised to do this, the numerous economic games that are being played [Fehr et al., 2002, Camerer and Fehr, 2006, Fehr and Schmidt, 1999, Fehr and Gächter, 2002, Gintis et al., 2003, Henrich et al., 2005, Herrmann et al., 2008, Bernhard et al., 2006, Fowler et al., 2005, Egas and Riedl, 2008, Marlowe et al., 2010, Chaudhuri, 2011], it is not clear whether they are measuring 'utility' or some aspect of the social contract [Dufwenberg et al., 2011, Cronk, 2007, Marsh and Kacelnik, 2002, Binmore, 2005a].

One problem that I have with the assumption that individuals simply tend to maximise their own material well being is that there seems to be the assumption that 'private property' is always simple and obvious. But in reality it is never as simple as that. Even in our own culture, where we tend to act as though private property is a natural and simple thing, reality is always more complex than this. Take, for example, the super market chain Tesco's—who owns that? The shareholders? The bank who loan them money? The management? The workers? What even is 'Tesco's'? Is it a building, or a plot of land, or is it a set of relationships between people? If so, how can anyone 'own' that. To give another example, when we interact with our friends, are we trying to maximise our material well being then? How does playing football on a Sunday make anyone wealthier? In a sense, football, or the resources that allow us to play it, *is* wealth. But football itself acts to form friendships which are useful from an evolutionary perspective. So calling sport 'pleasure' isn't necessarily right. Now we are back to the same problem as we had in the previous section—which of these is the 'true preference,' and which is a means of getting something else? The answer, I think, is that there is no answer. It is down to the whim of the modeller. Fitness is a universal currency, but it would be counter productive to describe how every single action relates back to this quantity.

We would find it difficult to explain the interactions between a husband and wife in terms of maximising individual private property. Indeed, it is even less clear what 'private property' is in such relationships. Likewise, when people live in small-scale societies, property—or the thing that most resembles property in the modern day Western civilisation sense of the word—is likely to be heavily reliant on the nexus of relationships and social contracts that any individual finds themselves in. There might be 'family property,' 'group property,' 'corporate property,' and so on. In section 4.7 I discuss Evans-Pritchard [1940]'s experience with property among the Nuer (pastoralists who live in Sudan), where, aside from cattle and a few household items that were strictly 'family property,' everything else was considered 'group property,' that any individual could take. What was unusual was that Evans-Pritchard himself was permitted to have more private property, and so other members of the group used him as a sort of deposit box, and so he describes how he had little pockets all over his tent with hidden tobacco inside, until he stopped doing this to live in 'poverty and peace.' This is a rather frivolous example, but it does serve to highlight the fact that there is nothing simple about property, because ultimately it depends on ones relationships with those around, which can be extremely complicated. Here I invoke Polanyi [1957]'s notion that the economy is an 'institutionalised' process.

One final word on this subject. If all behaviour is aimed towards maximising material reward, then how do we explain the invention of the norms of private property in the first place. It ludicrous to suggest that people invented private property in order to maximise it. The whole point of building the framework in this section is to be able to express notions such as 'private property' emerging not out of the private valuations of individuals, but as a feature of the group that has been invented to solve problems by balancing private desires.

2.5 An Evolutionary Model of Social Contracts

So far I have defined social contracts as a Nash equilibrium in a game of life. However, as described above, this theory makes many assumptions about human behaviour that simply do not seem realistic, for example, the assumption that all individuals know the payoff for taking a certain action, or are even aware that there is another possible way to act. Therefore, in this dissertation, I adopt an evolutionary approach which is more explicit about these aspects. Evolutionary game theory was first used by Smith [1982] in the context of biological evolution, but was later adapted by McElreath and Boyd [2008] to explain the process of cultural evolution. In the following sections, I outline this model.

Consider a population made up of two different types: σ_1 and σ_2. Suppose that there are n_{σ_1} type σ_1's, and n_{σ_2} type σ_2's (also define $n = n_1 + n_2$). Call p_{σ_i} the proportion of σ_i's in the population, so $p_{\sigma_i} = \frac{n_{\sigma_i}}{n}$. The payoff of a type σ_i, denoted w_{σ_i} is equal to the average number of progeny that type will leave in the following generation. Therefore, the number of σ_1's in the following generation is equal to $n_{\sigma_1} w_{\sigma_1}$. From this, we can calculate the proportion of type σ_1's in the following generation, which we call $\dot{p}$, which is equal to $\dot{p_{\sigma_1}} = \frac{n_{\sigma_1} W_{\sigma_1}}{n_{\sigma_1} W_{\sigma_1} + n_{\sigma_2} W_{\sigma_2}} = \frac{p_{\sigma_1} W_{\sigma_1}}{p_{\sigma_1} W_{\sigma_1} + p_{\sigma_2} W_{\sigma_2}}$ (this last step is achieved by multiplying the top and bottom of this equation by $1/n$). Let $\Delta p_{\sigma_i} = \dot{p_{\sigma_i}} - p_i$ be the change in frequency of type i. Then

$$\begin{aligned}
\Delta p_{\sigma_1} &= \dot{p}_{\sigma_1} - p_{\sigma_1} \\
&= \frac{W_{\sigma_1} p_{\sigma_1} - W_{\sigma_1} p_{\sigma_1}^2 W_{\sigma_2}}{p_{\sigma_1} W_{\sigma_1} + p_{\sigma_2} W_{\sigma_2}} \\
&= \frac{p_{\sigma_1} p_{\sigma_2} (W_{\sigma_1} - W_{\sigma_2})}{\bar{W}} \qquad (2.7)
\end{aligned}$$

where $\bar{W} = p_{\sigma_1} W_{\sigma_1} + p_{\sigma_2} W_{\sigma_2}$ is the average payoff of the entire population. This is normally written:

$$\bar{W} \Delta p_{\sigma_1} = p_{\sigma_1}(1 - p_{\sigma_1})(W_{\sigma_1} - W_{\sigma_2}) \qquad (2.8)$$

In these models, we are not generally interested in the change of the entire size of the population—only the change in frequency of a type. If we rescaled the fitness to be $w_i/\bar{w}$ rather than w_i, then the size of the population would never change.

To recap, this model represents the change in frequency of a type in a population. So far, there is no game theory here. In the following section, I introduce game theory into this model.

2.5.1 Evolutionary Game Theory

The model in the previous represents the evolutionary dynamics for any types σ_1 and σ_2. For the purposes of this dissertation, we can extend this model to include what happens the fitness of a type is dependent on the strategy that they play in a social interaction. In this model, we assume that each individual in the population is paired up with another individual at random to play their strategy. The payoff that an individual playing strategy σ_1 gets from interacting with an individual who plays strategy σ_2 is written as $w_{\sigma_1}(\sigma_2)$.

Now we want to create a set of conditions that tell us whether a given strategy, when common in the population, will resist invasion from another type. Suppose that σ_1 is the native strategy, and we want to see if σ_2 will invade. Suppose, therefore, that $p_{\sigma_1} = 1 - \varepsilon$ and $p_{\sigma_2} = \varepsilon$ where ε is small. What we want to check is that $\Delta p_{\sigma_1} > 0$ under these conditions. The idea is to check that if the population is perturbed by a small ε (which could represent the mutation of a new type), then evolution is expected to move the population closer to $p_{\sigma_1} = 1$. From equation 2.8 we can see that this occurs when $W_{\sigma_1} > W_{\sigma_2}$ (since $\bar{W}, p_{\sigma_1}(1 - p_{\sigma_1}) = (1 - \varepsilon)(\varepsilon) > 0$).

Now we need to calculate what the payoffs are. The payoff is obtained by assuming that each individual will meet another individual at random from the population, interacting with them socially using the strategy that they have evolved to use. If there are $(1 - \varepsilon)$ type σ_1's in the population, then the chance that an a random type will meet a σ_1 is $(1 - \varepsilon)$ and so on. Therefore, the average payoff for each type can be calculated as follows:

$$\begin{aligned} W_{\sigma_1}(1-\varepsilon, \varepsilon) &= (1-\varepsilon) w_{\sigma_1}(\sigma_1) + \varepsilon w_{\sigma_1}(\sigma_2) \\ W_{\sigma_2}(1-\varepsilon, \varepsilon) &= (1-\varepsilon) w_{\sigma_2}(\sigma_1) + \varepsilon w_{\sigma_2}(\sigma_2) \end{aligned} \tag{2.9}$$

Where $W_{\sigma_1}(1-\varepsilon, \varepsilon)$ is the payoff that strategy σ_1 expects to get when there are $(1-\varepsilon)$ σ_1's and ε σ_2's, and $w_{\sigma_1}(\sigma_2)$ is the payoff that individual of type σ_1 gets from playing the game with an individual of type σ_2 and so on.

We say that σ_1 is *evolutionarily stable* (*ESS*) if it can resist invasion from the type that plays σ_2. This means that the payoff of σ_1 when there are few σ_2's, which is $W_{\sigma_1}(1 - \varepsilon, \varepsilon)$, must be greater than the payoff for σ_2 in the same population, which is written $W_{\sigma_2}(1 - \varepsilon, \varepsilon)$. In other words:

$$W_{\sigma_1}(1-\varepsilon, \varepsilon) > W_{\sigma_2}(1-\varepsilon, \varepsilon) \tag{2.10}$$

Notice that if we take ε to 0, then we have the condition

$$W_{\sigma_1}(1, 0) = w_{\sigma_1}(\sigma_1) > w_{\sigma_2}(\sigma_1) = W_{\sigma_2}(1, 0) \tag{2.11}$$

which is nearly the same as the Nash equilibrium (it is called the strict Nash equilibrium).

When we check to see if strategy σ_1 is an *ESS*, we do not normally check condition

2.10. Instead, we use the fact that it implies that

$$(1-\varepsilon)w_{\sigma_1}(\sigma_1)+\varepsilon w_{\sigma_1}(\sigma_2) > (1-\varepsilon)w_{\sigma_2}(\sigma_1)+\varepsilon w_{\sigma_2}(\sigma_2) \tag{2.12}$$

for ε sufficiently small. Then we check the following conditions.

Definition 2. *σ_1 is an ESS when either;*

1. *$w_{\sigma_1}(\sigma_1) > w_{\sigma_2}(\sigma_1)$; or*
2. *$w_{\sigma_1}(\sigma_1) = w_{\sigma_2}(\sigma_1)$ and $w_{\sigma_1}(\sigma_2) > w_{\sigma_2}(\sigma_2)$.*

This is the standard definition of an evolutionary stable strategy [Smith, 1982, Britton, 2003].

2.5.2 Social Interactions Involving More than Two Players

In the model above, we assumed that players meet in pairs and play their strategies. However, for humans, there are many situations where multiple individuals must meet to interact. In this model, we suppose the population is divided randomly into groups of size $n+1$ (n other individuals plus the focal individual). If there are two strategies, and p_{σ_1} is the proportion of the population playing strategy σ_1, then the groups will be binomially distributed around p_{σ_1} (because each group is formed randomly). If we define the fitness of an individual playing strategy σ_1 in a group where i of the n other individuals are playing σ_1 as $W_{\sigma_1}(i)$, then their total expected payoff is equal to

$$W_{\sigma_1} = \sum_{j=0}^{n} \binom{n}{j} {p_{\sigma_1}}^j (1-p_{\sigma_1})^{n-j} w_{\sigma_1}(j) \tag{2.13}$$

To know whether a strategy is an ESS in this system, we must ask what occurs when $p_{\sigma_1} = 1-\varepsilon$, where ε is small, and we assume that the rest of the population play strategy σ_2 $p_{\sigma_2} = \varepsilon$. The payoff for strategy σ_1 here is

$$W_{\sigma_1}(1-\varepsilon,\varepsilon) = \sum_{j=0}^{n} \binom{n}{j} (1-\varepsilon)^j (\varepsilon)^{n-j} w_{\sigma_1}(j) \approx w_{\sigma_1}(n) \tag{2.14}$$

and the fitness type σ_2 is

$$W_{\sigma_2}(1-\varepsilon,\varepsilon) = \sum_{j=0}^{n} \binom{n}{j} (\varepsilon)^j (1-\varepsilon)^{n-j} w_{\sigma_2}(j) \approx w_{\sigma^*}(n) \tag{2.15}$$

Now we can use the same condition as 2.10 to get the following set of conditions.

Definition 3. *σ_1 is and ESS when $w_{\sigma_1}(n) > w_{\sigma_2}(n)$*

Note that this is *not* the complete definition for a ESS here, because it could be that $w_{\sigma_1}(n) = w_{\sigma_2}(n)$, in which case we would have to check whether $w_{\sigma_1}(n-1) > w_{\sigma_2}(n-1)$. If these two were also equal, then we would have to check that $w_{\sigma_1}(n-2) > w_{\sigma_2}(n-2)$ and so on. However, this does not occur in any of the models in this disseration, so we do not need to worry about this case.

2.5.3 Positive Assortment

One final addition to this model is *positive assortment*. Positive assortment means that the focal individual is more likely than chance to meet another individual of the same type. This could be due to limited dispersal, relatedness and so on. In this model, we assume that there is an additional r chance that an individual will interact with another who has the same type as them, and a $(1-r)$ chance that they will meet another individual at random from the population. This random encounter could also be with the same type, or it could be with another type. If we suppose that individuals meet in pairs, then the probability of a type σ_1 meeting another individual of type σ_1 is equal to

$$m(\sigma|\sigma) = r + (1-r)p_{\sigma_1} \tag{2.16}$$

Notice that r is not the chance that a type will meet its own type, but rather the *additional chance* that it will meet its own type. If those types are common in the population, for example if $p_{\sigma_1} = 1$, then it will always meet itself.

We can do the same for evolution in larger groups. Suppose that there is the probability that, given the focal individual in a group plays σ_1, then the probability that another individual plays strategy σ_1 is equal to $m(\sigma_1|\sigma_1) = r + (1-r)p_{\sigma_1}$. We call the probability that an individual of type σ_1 in a group of j other individuals playing σ_1 as $M(\sigma_1|j)$. This means that in a population where a proportion of p_{σ_1} are of type σ_1, then

$$M(\sigma_1|j) = \binom{n}{j}(r + (1-r)p_{\sigma_1})^j(1-(r+(1-r)p_{\sigma_1})^{n-j} \tag{2.17}$$

Positive assortment can have a significant effect on invasion. To see this assume that $p_{\sigma_1} \approx 0$, so that $m(\sigma_1|\sigma_1) \approx r$, then the probability that there are j individuals of type σ_1 is a group is equal to

$$M(\sigma|j) = \binom{n}{j} r^j (1-r)^{n-j} \tag{2.18}$$

We can therefore calculate the payoff for playing σ as being

$$W_{\sigma_1} = \sum_{j=0}^{n} M(\sigma_1|j) w_{\sigma_1}(j) \tag{2.19}$$

This can be written

$$W_\sigma = \sum_{j=0}^{n} \binom{n}{j} r^j (1-r)^{n-j} w_\sigma(j) \tag{2.20}$$

A resident strategy σ_2 is always going to be in the same of the same type because $m(\sigma_2|\sigma_2) = r + (1-r)p_{\sigma_2} \approx 1$ when $p \approx 1$. So

$$W_{\sigma_2} = w_{\sigma_2}(0) \tag{2.21}$$

With these assumptions we get the following result

Definition 4. *σ_2 is an ESS when $W_{\sigma_1} < w_{\sigma_2}$.*

Again, there are difficulties when these terms are equal, but I do not deal with such situations in this model.

The model introduced in this section extends the basic model of the Nash equilibrium. For the Nash equilibrium, we imagine pairs of players playing the game, and assume that they will act rationally to maximise their payoff. In an evolutionary model, we no longer assume that the players are rational, but rather, we assume that the strategies of the previous generation are inherited in subsequent generations, with the more successful strategies being better represented.

2.5.4 Non-Genetic Evolution

The method of studying evolution outlined above applies primarily to genes. However, there is no reason why we cannot apply it to other forms of replication as well. In many species, and particularly in humans, the process of cultural evolution is important, and can be captured using the same equations.

The purpose of this model is to show that, from a certain perspective, the process of cultural evolution is analogous to the process of genetic evolution. The idea is to replace the notion of a Nash equilibrium with a cultural evolutionary process by including the features of how strategies change through time. In this model, players chose their strategy, and then compare their payoff to other players, and will swap strategies when the other players do better than them.

Suppose that players play a game, and that after the game, all players inspect the payoffs of other players around them, and choose to play their strategy when it is higher. McElreath and Boyd [2008] show that this model is equivalent to the replicator equations via the following method. Assume that a proportion p of the population is playing strategy A and $1-p$ are playing B. Assume that if two player meet who have different strategies, then the probability that a player selecting strategy A is given by

$$Pr(A|A,B) = \frac{1}{2} + \beta(w_A - w_B) \tag{2.22}$$

We also know that the total fraction of the population playing different strategies that meet is equal to $2p(1-p)$ (the fraction of the population who meet that both play A is p^2 and the fraction that meet and both play B is $(1-p)^2$).

Then we can calculate the change in p is

$$\begin{aligned}\Delta p &= p' - p \\ &= p^2 + 2p(1-p)\left(\frac{1}{2} + \beta(w_A - w_B)\right) - p \\ &= 2p(1-p)\beta(w_A - w_B) \qquad (2.23)\end{aligned}$$

Notice that this is essentially the same equation as in equation 2.8, with the only difference being in the magnitude of change.

However, while this is the primary approach I use in this dissertation, it is important early on to consider two limitations. The first limitation is that this model describes cultural evolution only as a gradual process of evolution. That is, it assumes that the only way that a culture can change is by this gradualistic process, and that from this perspective, innovations are 'mutations.' I believe that while this is an important model, culture may also change by a different manner. To give an example, suppose that we have the driving game introduced in section 2.4 of this chapter. Recall that the payoff matrix is as follows:

	Left	Right
Left	1	0
Right	0	1

We can easily make this into an evolutionary game by supposing that, at each generation, one individual meets another individual at random from a population. If we assume that the proportion of players playing strategy L, which means to drive on the left, is p, then it is easy to see that both L and R are equilibrium strategies. This is because when every player in the population plays strategy L, it is best to play strategy L, when the strategy is R, it is best to play R. However, the model of cultural evolution in this section shows how driving on the left can be maintained by itself, however, it would also suggest that the side on the road on which we drive, once decided, can never change. But we know this is wrong. It has been changed many times before. How was it changed? Through political mandate. Governments are capable of negotiating and organising a change from one norm to another. This process is not selective, but occurs through the process of negotiation and consent. I discuss this in more detail in chapter 6.

The practical upshot of this is that I argue that, unlike with genes, we do not necessary have to show the invasion of a social contract from rarity. However, a social contract must be able to resist invasion from other possible strategies.

Another limitation to this model is that it assumes that every individual in the population necessarily benefits from copying the successful strategies. However, this might not be the case. Consider the fact that many humans form hierarchical organisations. Suppose that we try to apply this model to one of these organisations, say, a bank. Is it sensible to assume that, within this bank, a cashier will copy the CEO because they are paid more? Of course it is not. The reason for this is that a bank is not simply a set of individuals each individually attempting to maximise their own payoff relative to one another. Rather, it is concerted action, based on the division of labour and of responsibilities. In fact, the division of labour, a feature of humans that make us so successful, is completely absent from this model. Again, this point is expanded in section 6.

Finally, although it might not seem so, this theory of cultural change is more akin to rational choice than to evolutionary biology. It is best to think of it as a more realistic version of the Nash equilibrium. In the Nash equilibrium, we assume that players just jump to the equilibrium assumption. In this model, there is no such assumption. We assume that players know their own payoffs, but they do not calculate whether a certain move gives them this higher payoff until they see someone else doing it. Their evaluation of the situation never changes, only their appreciation of how well a strategy gets them to their desired outcome.

2.5.5 Negotiation of Social Norms

Suppose that some players of species X meet to play the game that has the following payoff matrix for both players:

	Left	Right
Left	b	0
Right	0	1

with $b > 1$. Now imagine that this species was currently adapted to playing the 'right' strategy. Here we have a classical problem of how to reach the higher peak. Any individual that attempts to play 'left' will receive a payoff of $0 < 1$ and will immediately be selected against. When there is genetic relatedness between the interacting members then, using the model , we can see that relatedness r would need to be $r > 1/b$.

But humans do not need to wait for genetic evolution to solve this problem. Rather, as I describe in the previous section, we are able to negotiate and to adopt new equilibria all at once. Why did we evolve this way? I argue that it is due to the number of different ways that have found to coordinate and cooperate. For example, humans are adapted to survive in many different habitats. It would be no good having a fixed adaptation for cooperating if the environment changed so much. We have evolved large brains that are able to understand the the potential mutual benefits from cooperating and to take advantage of them. Likewise, we have invented complex ways of communicating these ideas to one another.

However easy it is to describe this process, it is difficult to model precisely. This is due to the sheer number of situations where we negotiate contracts.

The only attempt that I am aware of to describe the formation of social contracts from an evolutionary perspective in any detail is the work of Binmore [2005b, 1998, 1997]. Binmore argues that humans have evolved negotiate social contracts from behind what is called the veil of ignorance [Rawls, 1971]. In this model, players negotiate about how a society should be organised as if they do not know what position they will take within that society. Under these conditions, we evolve to create fair social contracts. I discuss this model in more detail throughout this dissertation, and particularly in section 6.

But I do not follow Binmore's lead on this matter. Rather, I simply assume that there is some process that leads to the selection of a social contract. I also assume that humans have evolved to create 'fair' contracts. But do social contracts need to be fair? No. We only need to consider the history of slavery to see this. However, it so happens that most modern day hunter-gatherers are relatively egalitarian, and if our evolutionary ancestors are anything like them, and there is no great reason that I know of to assume otherwise, then the contracts of our ancestors were likely to have been egalitarian as well [Boehm, 2009].

2.5.6 Negotiation of Contracts and the Evolution of Language

This dissertation is about the evolution of social contracts, but this question is closely related to the evolution of language, and so I find it necessary to express my views on the question of how and why language evolved. Dunbar [1998] argues that language initially evolved as a means of social bonding. The idea is that, as humans came to live in larger groups, we needed a mechanism to be able to bond with many individuals at the same time, and that to do this, we used vocalisations that were a precursor to language. Language then evolved to include 'informational content' as an added extra.

But I argue that this logic is incorrect. In the context of communicating the status of a relationship, the most important feature of a signal is that it is costly. Only by being costly does it communicate information about the status of a relationship. The logic, which I describe in some detail in section 5, is that only individuals who value a relationship would be willing to pay the cost. Being able to transmit to many individuals at once undermines this cost, thereby rendering vocalisations useless as a means of transferring information about the status of a relationship.

However, I am more concerned with how language acquired a different kind of informational content. A great deal of theoretical work has been done to explain how social learning, which is key the evolution of language, can evolve. In this view, 'information' is something that exists in the environment, and it can be accessed by individuals, and presumably communicated through language. We use language to name a thing, begin associating qualities with it, and then we can communicate this to generate an adaptive advantage. However, I argue that this view of the evolution of language is incorrect.

Instead, I argue that language evolved as a means of negotiating and communicating the social contract. 'Information' is not so much something that exists in the environment, as much as it is created through conversation. This information has a great deal of adaptive value, because failing to learn the new moral code will no doubt result in punishment, or some other form of conflict. For example, suppose that a group establishes a system of property rights. This process involves them creating an abstract principle, the principle that a certain good will be exclusively used by an individual. It also involves naming other classes of goods, here, the things that can reasonably be 'owned.'

But isn't this arbitrary? Why is such a signal not cheated and broken? Surely natural selection would never select such a thing. This view communicates a common misconception that natural selection implies that 'betterness' is selected, where 'betterness' can be defined independently of the species that we are investigating. But why must this be the case? In fact, as we see in section 3.4.1, not only is it possible that otherwise arbitrary signals can settle a fitness enhancing dispute, it is quite common in nature.

So then how is language selected? I believe that the answer to this is that the type of social norms that we create, although arbitrary from the perspective of an individual dilemma, are generally created in order to promote pro-social behaviour, or equivalently, behaviours that make groups prosper, and to punish anti-social behaviour.

But I do not want to get ahead of myself here. All that I want to point out is that while I believe that the evolution of language is important, when we view it from the perspective of organising social contracts, it is not a difficult to explain as is commonly believed. I return to the question of language in sections 3.4.3 and 3.6.

2.5.7 Mobility of Hunter-Gatherers

One possible issue with the framework described in this section is the fact that it is supposed to apply to hunter-gatherers, but it assumes that once a 'group' of hunter-gatherers form, they remain with one another until the entire group disbands, and that there is no migration in the meantime. However, in reality, we know that hunter-gatherers are in fact quite mobile, which can cause some problems for the models [Eshel and Cavalli-Sforza, 1982].

This difficulty is not one that is particular to my own work, but to all work in this area, save those who use computer simulations in order to model the effect of migration. However, in defence of the orthodox, I argue that the models are already very abstract, and their purpose is normally to make specific points, rather than to accurately reflect reality. Of course we want to accurately reflect *some* aspects of reality, but we cannot hope to reflect all of it. That said, being able to 'walk away' from those who do not play according to the rules will obviously make a large difference, since it provides a relatively cheap method of punishing bad behaviour. But on the other hand, those who choose to stay together still have to devise a method of living, and not all transgressions are sufficiently costly that it is worth the entire group leaving, particularly if the transgressions are only against a particular individual.

One other point that is worth noting is that the effects of migration is likely to be very different for cultural evolution than it is for genetical evolution. For genetical evolution, when an individual changes group, they obviously carry their genes with them. However, with cultural evolution, when an individual leaves one group to join another, they do not necessarily have to play the same strategy when they join the new group. In fact, it is likely that they will quickly learn what is the 'best strategy' in the new group by observing what it is that they do and fitting in. The obvious example of this is the fact that the same person can drive on the left in Britain and on the right in France, maintaining the equilibrium in both countries.

One final point worth considering is that 'walking away' from a particular individual in response to their bad behaviour could itself be seen as part of the a social contract between the individuals who do that. For example, if one individual was bullying a second, then there may be no direct reason why a third individual might want make the special effort to walk away. However, such a contract would is relatively easy to 'self-police,' since those who did not leave with the rest of the group would be left with bully, which presumably is not the best strategy for them. It would be worth investigating the balance of costs and benefits for working away, but it is not something that I attempt in this dissertation.

2.5.8 Indefinitely Repeated Games

In this section, I describe how the model so far deals with indefinitely repeated game. The purpose of introducing indefinitely repeated game is to represent situations where individuals do not know when the interaction will end, but they do expect it it to end at some point. This is important, because if individuals knew when an interaction was going to end, then their best strategy is often to cheat on the last round, particularly in situations where the reason that they are performing a given action is because of what it will supply them in the future, for example, reciprocal altruism.

To deal with this, we assume that there is a possibility $\delta \in (0, 1)$ that an interaction is going to end between each round. So the probability that it will last at least until the next round is δ, and the probability that it will last at least two rounds is δ^2 and so on.

Likewise, the probability that it lasts exactly one round is $(1-\delta)$, and the probability that it lasts exactly two rounds is $\delta(1-\delta)$ and so on. The expected number of rounds is equal to

$$\sum_{i=1} \delta^{i-1}(1-\delta)i = \sum_{i=0} i\delta^{i-1} - i\delta^{i} = \sum_{i=0} \delta^{i} = \frac{1}{1-\delta} \tag{2.24}$$

There are also situations where we want to look at the expected number of interactions, *given* that a certain number of interactions have occurred. But this is easy. The expected number of interactions after the first round is $\frac{1}{1-\delta}$. The average after the second round is the same.

The last thing we need to know is represented by the following setup. Suppose on round 1 I get a payoff of A, and then on every other round I get a payoff of B. What is my expected payoff. This is equal A, plus the probability that I make it to the next round (δ), multiplied by the expected number of rounds $\left(\frac{1}{1-\delta}\right)$, multiplied by payoff ($B$). So the total expected payoff would be

$$A + \frac{\delta}{1-\delta}B \tag{2.25}$$

2.5.9 The Semantics of Cooperation: Mutual Benefit, Selfishness, Altruism and Spite

Throughout this dissertation, I use terms such as *cooperation* and *altruism* quite loosely to mean 'people help each other.' However, in the literature, because of certain disputes between scholars, these words have developed quite precise meanings, so in this section I describe the meaning of each one in detail based on the scheme that originated with Hartung [1999]. West et al. [2008] provides a good review of the confusions.

Suppose that one individual, our focal individual, meets another, the target individual, with whom they interact socially. The focal individual performs a behaviour that affects the payoff that they receive, and the payoff that the target individual receives. There are four possible outcomes:

1. Mutualism (+,+): The behaviour that the focal individual performs will be beneficial to both themselves and to the target.
2. Selfish (+,-): The behaviour that the focal individual performs is good for themselves and bad for the target.
3. Altruism (-,+): The behaviour that the focal individual performs is bad for them and good for the target individual.
4. Spite (-,-): The behaviour that the focal individual performs is bad for both them and the target individual.

Here the '+' in (+,-) is the effect of the action on the focal individual and the '-' is the effect on the target, so this would be 'Selfish' behaviour.

This definition becomes more complicated in models where interactions are likely to repeat. The reason for this is that if an individual performs an altruistic action, but *as a consequence of that action*, the target individual does an altruistic action in return (such as with reciprocal altruism), then the net effect is that the combinations of the actions are mutually beneficial. This becomes more complicated again when we have social interactions that go on indefinitely, for example, social interactions where there is a probability δ that the interaction will continue in the following round. Here we take the payoff over the *expected* number of outcomes.

As I discuss in section 3.4.6, it not always easy to classify a certain behaviour, particularly when the model contains multiple actions and repeated games. For example, take the model of reciprocal altruism. When a reciprocal altruist meets another reciprocal altruist, the relationship is mutualism. However, when they meet a defector, the relationship is altruistic. In fact, this means that for reciprocal altruism to invade a population of defectors, there must be a significant amount of positive assortment.

2.6 Game Theory and the Study of Human Behaviour

In this section I outline how I believe that game theory can help us to understand human behaviour.

2.6.1 Human Behaviour and Social Contracts

First let us discuss the kinds of measurement that evolutionary biologists use to investigate species other than humans. Biologists have also found it useful to adopt the 'means-end' vocabulary for describing animal behaviour. Nowadays, biologists posit a force that 'causes' the animal to 'want' the end — natural selection. Biologists do not use simple models of cause and effect models when investigating species, but instead use what we call 'population thinking' and statistical methods. However, the actual methods aside, when biologists posit theories about a particular species, they generally, but not always, generate theories that apply to the entire species. Of course, this also applies to humans — for example, we know that all humans have two parents.

However, in species that have other systems of inheritance, such as social learning, often we find that theories at the level of gene to be to coarse for accurately predicting behaviour. This, I argue, is exactly how we ought to think about human behaviour — there are large and important parts of our behaviour that are not predictable from the level of genes. The best example of this is language — we cannot possibly predict what language an individual will speak from their genetics alone. These rules of the game are what I have called the 'social contract,' which I construe as widely as possible to include all culturally influenced behaviour. Of course, adding social contracts into our understanding of the behaviour of a species does not mean that we can ignore evolutionary biology. Evolutionary biology is necessary for our understanding, but, in many important cases, it is not sufficient.

2.6.2 Human Behaviour and Prediction

So far I have established that often, human behaviour can only be comprehended relative to a set of abstract principles that I have called a social contract. In this section, I explain why I believe it is difficult to measure contracts.

There are four primary difficulties in investigating human behaviour from the perspective of social contracts. The first is the fact that the space of possible social contracts is enormous. The second is the complexity of such contracts. The third is that each contract is possibly unique. The fourth is the fact that, at least in social interactions, a great deal of action is what we call *symbolic*. Something is said to be symbolic when it represents, but is not related to, another thing. A word might represent an action, but the action has nothing to do with the sound of the word. So without knowing the language, all that we can observe is individuals making incomprehensible sounds with their mouth.

The same is true for other types of social phenomena. For example, suppose that an individual hands another individual a tool. The physical process is quite easy to describe, but 'handing over a tool' does not explain the social process that we need to truly comprehend the action and, more importantly, it does not, by itself, allow us to predict future action. Was the tool lent, sold, given as a gift, or was it passed in a wider context of cooperation, for example, two builders working for a company? But even identifying which of these social actions has occurred is insufficient, because we must also explain what each of these things mean. For example, a trade involves a complex social arrangement based on concepts such as exclusive property rights, which in turn might rely on not just an understanding between the two individuals involved in the trade, but also between them and other members of the group, since they might have to step in to resolve a dispute between two individuals. This in turn rests on who is willing to help, and what those arrangements rest on. The point is that none of these terms represent simple actions.

Another problem that makes measuring social contracts particularly difficult is that they often involve counter-factuals. To give an example that I discuss at length throughout this dissertation, a social contract might involve some form of punishment. As I discuss in detail in chapter 4, within human groups, punishment often begins with laughter being directed towards the individual who has transgressed against a moral rule. Among other things, this seems to be a tool of consensus building. The difficulty here is in determining how to measure such a thing. Was it a joke? A threat? A threat masked as a joke? A joke masked as a threat? Even if we can classify it, how can we say that we have 'observed' an implied threat if the threat is not carried out. The threat of punishment involves a counter-factual — what would happen *if* you break the rule. But if the rule is not broken and the punishment does not occur, then can we really say that we have 'observed' punishment here.

Yet, for all of these difficulties, we *are* good at predicting human behaviour. When we drive down the street, we expect everyone to drive on the left, and they normally do! When we walk into a shop, we expect that we will have to exchange money in order to purchase an item, and, again, this is normally so. We learn very quickly the abstract principles which guide actions within a group — the principles of the social contract. We also learn to communicate these principles. To do this, we name 'types' of behaviour, like trade, theft, gift, rape and others, and with these, we associate what 'ought' to happen in response to these actions. This forms the basis of how we predict the behaviour of others — the combination of the particular circumstances and the general principles. In fact, this is what I argue that language primarily evolved to do — to create terms that categorise different actions and to associate moral rules to them. This gave language one of its most important features — that it can be spoken to many different individuals at the same time.

But this is not simply a problem with comprehending human behaviour, but also for other types of social primate. In his book *Chimpanzee Politics*, De Waal [1982] outlines the standard ethological method at the time. The method was to catalogue behaviour, independently of time and context. They then used computer models to sift through the data to give information about bouts of violence, hunting, or what ever the researcher wanted to know. The problem, according to DeWaal, was that such information is useless when it comes to the political life of chimpanzees. For example, individual chimpanzees often 'gestured' to the alpha male, which involved them walking past him and performing some sort of behaviour. Each individual had their own specific type of unusual act, which made cataloguing them difficult. When there was doubt about the leadership, these gestures stopped. Then, when a new leader was in place, the gestures began again, this time directed towards the new leader. DeWaal's point was that the gestures clearly meant something important to the chimpanzees. The assumption was that the actions were not meaningless noise, as they had been treated by previous investigations, but that were purposeful action, and that the only way to understand the dynamics of the group is to observe them and to use them to try to predict behaviour. Through time, Dewaal developed a sophisticated view of chimpanzees, and was able to predict future behaviour based on quite subtle body language. But another point is also clear, DeWaal did not 'crack the code' of chimp communication — rather, he had simply gained insight into how the individuals within that particular group used symbolic communication.

In this dissertation, I argue that it is precisely the same for humans. We are capable of making qualitative predictions about what humans will do. However, I do not believe that there is an easy way around the problem of understanding the set of symbolic actions that a particular culture adopts. Taking the example of handing over the tool, there is no quick way to make a prediction about this — the only way to do it is to build up a knowledge of the local social contract.

2.6.3 The Uses of Deductive Game Theory

But by adopting this view, it seems that I have undermined the value of a deductive method like game theory. If all cultures are different, then what is the point of a generalisable theory like game theory? I believe that the answer to this is that it should play the same role in the study of culture as theoretical biology plays in the study of biological adaptations. In evolutionary biology, each theory only applies to a specific set of biological phenomena. For example, a theory of sexual selection in deer pertains only to the deer. If we want to find out how other species act, we have to observe them. The reason that we cast our theories in terms that are independent of a species is that, often, many different adaptations have similar adaptive logic. For example, the theory of why stags grow antlers may have a similar explanation as to why giraffes have long necks — that they are both used in order to fight and therefore attract mates. In this way, game theory is a labour saving device.

When we look closely at what game theory really is, we see that it is simply a language, based on as simpler a set of assumptions that we can postulate, which we use to explain human behaviour. These assumptions are that, within a population, individuals act to maximise some quantity, and that there are some constraints on what choices they have. Ultimately, evolutionary game theory assumes that what they are maximising is fitness, but as I have argued, it is often expedient to assume that they are maximising some other quantity, which we later argue relates to fitness in some way. The idea is that if our language is going to 'map' reality, then we can and should refine it into its most essential parts. We cannot 'discover' anything about the world with equations, we can only discover things about our own systems of thought, therefore I believe that it is important to remember that game theory is really a restatement and refinement of ideas that already exist.

Therefore, throughout this dissertation, I adopt a method where I work as closely as I can to the empirical literature on a given subject, and, wherever possible, I draw directly on the ethnographic literature. For example, I am interested to understand how our hunter-gatherer ancestors evolved to solve food-sharing problems, and therefore I consult the ethnographic and anthropological studies of food sharing and try to understand how the sharing works. Ultimately, this method relies on feedback between the theoretical work and the empirical work, because no doubt, some of the assumptions that I make in this dissertation will be incorrect. This is therefore the first step in a process of refinement that I hope will occur after my dissertation.

To recap what I have said here, I firstly argue that when interactions are expected to be repeated, as they are with humans and other social species like chimpanzees, then there is often no simple way to describe an interaction in a social context. Even interactions that we believe are simple often rely on a series of complex assumptions and involve many counter-factuals, the validity of which is inherently difficult to establish. Secondly, I argue that the social contract that govern each group is potentially different, and can evolve independently, so that to obtain knowledge about one particular group's contract, we must investigate that group specifically. Thirdly, I argue that game theory is a labour saving devise that we can apply to different contracts in different groups where there is a similar underlying set of principles.

2.6.4 The Method in Detail

In each of the four main chapters in this dissertation, I employ the following strategy. First, I review the broad empirical literature on the area of interest. Next, I review the popular theoretical explanations for the phenomena. Then I put forwards my own theory, expressed as a system of theoretical equations. Finally, I search the empirical evidence that either supports or refutes the assumptions that I find. My approach to selecting evidence is to select a number of the paradigm cases that each model refers to, and compare the assumptions of my model to these. This is unsystematic, because each literature is organised differently, and I have will no doubt miss one or two important examples.

This last part is, I believe, the most important. It is here where we receive the payoff of the model — because it helps us to understand the world better. But not only is it important to find supporting evidence, it is also important to find evidence that might conflict a theory, or at the very least, state what we do not know. In this way, empiricists should clearly be able to state where they believe that the assumptions of the model are incorrect so that the model can be revised and improved.

To present this evidence, I asked a series of possibly nested questions of the empirical literature. For example, in chapter 3, one of the assumptions that I make is that the marginal cost of fighting over a resource is less than the marginal value of that resource. In this case there is little direct evidence — there are few observed instances of fighting over food in the ethnographic literature, and certainly not enough to put an actual number. So in situations like this, I state that this is the case, and search for indirect evidence.

2.6.5 Economic Games and the Evolution of Cooperation

In recent years, the evolution of cooperation literature has married the empirical techniques of behavioural economics. Behavioural economists make different assumptions about human behaviour than I do, and so due to this and its prominence, I find it necessary to devote some time to discussing assumptions of this empirical school and the effect that the union has had on the way that we investigate human culture from an evolutionary perspective.

Firstly, what are the aims of behavioural economics? Behavioural economics is an experimental technique for measuring human behaviour. One particular goal of this literature is to counter the so-called 'selfish axiom:'

> Experimental evidence has strongly confirmed the doubts expressed ... concerning the adequacy of self-interest as a behavioral foundation for the social sciences... These experiments create an empirical challenge to what we call the selfishness axiom — the assumption that individuals seek to maximize their own material gains in these interactions and expect others to do the same. [pp. 1-2 Henrich et al., 2005]

This quotation articulates a position that is common in the evolution of cooperation literature — that it's goal is to overturn the 'selfish axiom' that it beleives are held in the social sciences. It does this by assuming that if it were true, then individuals ought to maximise the amount of money that they obtain in each social interaction. If they show this to be true, then the axiom is refuted.

There are several problems with this position, and here I deal with each of them in turn. The first is that the set of scholars who subscribe to the 'axiom of selfishness' is, as far as I have been able to find, empty [Binmore, 2005a]. The standard assumption in neo-classical economics is that individuals act to maximise subjective preference. If they were modelling saint Francis of Assisi, then they would have to say that he loved to help other people and would act to maximise that quantity.

The second problem is with interpreting behaviour in the games. The standard assumption is that individuals within the game know exactly how the game works, and therefore, if they select an outcome that does not maximise their material gain, then they do this because they 'prefer' that outcome. However, another interpretation of the results of these games are that we are not observing 'preferences' at all, but we are observing some manifestation of the 'values' of that society (I call the values of a society their contracts — see section 2.4.1 for a discussion about the difference between values and preferences) [Binmore, 2005a]. I would also like to add another possibility to this, that what we observe in economic games is humans attempting to communicate something to one another. I believe that a great deal of behaviour in these games is likely to be individuals 'gesturing' towards one another, albeit, with an extremely limited vocabulary.

This means that there is no guarantee that players are necessarily playing the game that the researchers believe they are playing — i.e. the one that is represented by the model. I know of an example of a dictator game that was played in a small-scale society, where the players gave some of their resources away in the dictator game and kept the rest. But when the game was concluded, all the participants immediately pooled all of their winnings and went to town to buy a pig which they cooked communally and ate. Here we have a clear problem — what does 'material gains' mean in this context? The only time when anything could be said to belong to any specific individual was the food in their hands before they ate, but that had to go through several cooperative processes of buying the meat, transporting it back, making a fire, cooking the meat and so on, beforehand. Clearly the individuals within the group did not think of it as 'their money,' and so we have failed to test their 'underlying preference' here.

I believe that the alternative views of the value of results from economic games stem from a fundamental set of beliefs about what we can infer from observing single, isolated actions. Behavioural economists assume that we can 'infer outwards' from the behaviour in the games. This somewhat mirrors the assumptions about regularity that natural scientists make — we can investigate behaviour independently of context and time. But I have argued against this assumption for two reasons. Firstly, in human sciences, we do not expect perfect regularity, and so we cannot know that regular behaviour in one context implies regular behaviour in another, and given that we know that humans act differently in different contexts in everyday life, it seems entirely likely that the same would be true in experiments. Secondly, interpreting the games at face value leaves no possibility for symbolic behaviour, which I have argued is vital for understanding human social interactions.

How to resolve these differences in opinion? One sensible way to do this is to consider the sort of hypotheses that such methods can generate and test. But here is precisely where I argue that traditional methods gain the upper hand. An anthropologist studying the food sharing norms of a particular hunter-gatherer society will be able to predict how food will be distributed within that society. But a behavioural economist who has performed a series of ultimatum games on the population will be unable to predict this. They may be able to predict how they will play in the ultimatum game under laboratory conditions, but this alone will not help us to predict the social contract that operates within that society.

However, in no way do I want to discredit economic games entirely. In fact, I believe that they are potentially an extremely valuable method. I believe that they would best be deployed to study situations that where other methods fail, namely, in asking counter-factuals, such as 'what would happen to an individual who stole,' since in reality, a researcher may have to spend a great deal of time waiting for a theft to occur. With this in mind, I argue that the empirical method of economic games would be more useful if they invoke, rather than remove, aspects of the social contract. The current philosophy is to make the environment as sterile as possible so that it does not invoke any normative behaviour. The problem is that I believe that some normative behaviour will always be invoked, and we do not necessarily know which (we may not even know what sorts of normative behaviour exists in a particular community). So my vision is that researchers construct the games along the lines of some aspect of the social contract, and use it to try to discover some aspects that contract that was hitherto inaccessible. Of course, this will make cross cultural comparison more difficult, but this alone is like criticising the use of different empirical techniques to study different species because they make a cross-species investigation difficult.

2.7 Conclusion

To conclude this section, I want to once again clearly state the method that I use in this dissertation. My fundamental assumption about human behaviour is that humans consistently act to bring about some positive evolutionary end. In order to do this, we have evolved to form social contracts. Social contracts organise a society. They make our behaviour predictable. They allow us to cooperate.

The formal theory that I use to study this is evolutionary game theory. This method is designed to explain how stable social phenomena can occur even when the motives of individuals in the group are antagonistic. This is important in humans because, unlike other social species, we are not genetically related to the individuals with whom we cooperate.

I also warn against taking the analogy between cultural evolution and genetic evolution too far. Some of the tools and ideas that are indespensible to biological evolution, like the idea of invasion from rarity, do not necessarily have applications for culture.

I also outline my broader scientific approach for studying human culture from an evolutionary perspective. I argue that the community ought to rely primarily on anthropological and ethnographic data, and that testing with economic games should work with, and not against, our wider understanding of given social contracts.

CHAPTER 3
FOOD SHARING AND THE EVOLUTION OF THE SOCIAL CONTRACT

Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his labour with, and joined to it something that is his own, and thereby makes it his property. It being by him removed from the common state nature hath placed it in, it hath by this labour something annexed to it, that excludes the common right of other men: for this labour being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to, at least where there is enough, and as good, left in common for others.

— John Locke, *The Second Treatise on Civil Government, Chapter V*

Communism deprives no man of the ability to appropriate the fruits of his labour. The only thing it deprives him of is the ability to enslave others by means of such appropriations.

— Karl Marx, *The Communist Manifesto*

3.1 Introduction

In this chapter, I introduce a novel model of food-sharing in hunter-gatherer populations. I focus on the question of property rights, which I demonstrate are fundamental to all of the current descriptions, and argue that we can obtain a great deal of insight if we consider this question from the social contracts perspective that I outline in the previous chapter.

Food sharing was likely to be the earliest problem in human evolutionary history that required a social contract to solve [Hawkes, 1992, Gurven, 2004a, Kaplan et al., 1985, Hill et al., 1993, Kaplan et al., 2005, Ridley, 1997, Marlowe, 2005]. Anthropologists have long recognised that food sharing is an almost universal human habit. In western civilisation, it remains important as a social ritual adhered to by almost all cultures, but in our hunter-gatherer past, as with hunter-gatherer societies today, it is likely to be the difference between life and death. The reason is simple. Humans have evolved to require a steady income of high-quality nutrition [Kaplan et al., 2000]. Acting alone, such an income is practically impossible to attain because of the variable returns from hunting. Even the ablest hunters may return home empty handed much of the time. But by pooling their risk, seeking larger prey, and sharing the rewards, hunters have learnt that sharing can be profitable.

Food-sharing in hunter-gatherer societies is striking because of the manner in which it occurs. Though there is a great deal of cross-cultural variation, the sheer scale at which it operates in most communities is astonishing — far wider than we might expect for social mammals with such a low level of genetic relatedness. Hunters regularly set out from their camp seeking large game which, individually, they have little chance of catching, and even if they did, because it is practically impossible to store food over a long period, their family and they would only be able to take advantage of a small portion of the meat they catch. But they hunt because they mean to feed others in the group, and when they return to camp, they can be confident that others will provide food for them, making up for any short falls.

To the casual observer, food sharing is unproblematic. Surely our ancestors were endowed with the same instincts to help their fellows as we are today? Would they not just use their kindness, or even adopt Kant's categorical imperative — to treat every individual as they themselves would be treated — as a universal moral law? Unfortunately for hunter-gatherers, natural selection operates by selecting the most successful strategies, rather than by satisfying Kant's metaphysics. While it may be true that humans have evolved the subjective desire or instincts to help those in need, this fact alone does not answer the question — it does not explain *how* those instincts evolved. A casual observer might respond that surely those communities whose members help one another do better than those that don't. But here again is an error. Natural selection does not operate primarily to maximise the efficiency of the group, but rather to maximise inclusive fitness of individuals within that group. In fact, it is unclear exactly what 'better group' means under these circumstances, unless it is a group that is made up of individually fitter members. Evolutionary theory shows us that only under the special circumstances that the members of a group are highly related, which is untrue for humans, is it sensible to treat communities as homogeneous and to neglect the strategies of the individuals who make up those communities. In terms of the problem of food sharing, this means that if food is shared, then it must be the best strategy. The task is to work out how.

The fact that food is shared can hardly be denied — there is overwhelming evidence that it is shared extremely widely. Neither, I think, can its importance to most, if not all, hunter-gatherer communities. However, the details of exactly *how* food is shared is contested in the anthropological and evolutionary biology communities [Hawkes, 1992, Gurven, 2004a, Kaplan et al., 1985, Hill et al., 1993, Kaplan et al., 2005, Ridley, 1997, Bowles and Gintis, 2011]. Many anthropologists would disagree with my description of food sharing. I use the word 'shared' as an empirical statement about what we observe — food is handed to other individuals within the group which they consume. However, this must not be mistaken with the verb 'given,' which is a theory about the social contract. To be given, the food must belong to the the giver. It must be their property. But in the estimation of many anthropologists, food is not given at all, but rather it is *taken* [Blurton Jones, 1984, Woodburn, 1998, Bird, 1997, Winterhalder, 1996, Hawkes, 1992, Polanyi, 1944]. For example, Woodburn [1998], in a chapter entitled 'Sharing is not a Form of Exchange,' argues that the 'giving' explanation places far too much emphasis on the giver, and that hunters generally do not rely on the kindness of others in food distributions. Giving food away is not considered a gift in these societies, and we only believe this to be so because of our own Western understanding of trade. The food is not given because it never belonged to them in the first place. Rather, when an individual goes to receive their share, they are simply enacting their right to the produce. They are collecting their property.

Formal theoretical models of this 'taking' theory of food sharing are somewhat lacking. The only attempt to square observation with standard evolutionary theory is the theory of *tolerated theft* [Jones, 1991, Blurton Jones, 1984, Bird, 1997, Winterhalder, 1996]. Backed up by references to the hawk-dove game of conflict in evolutionary game theory, tolerated theft argues that within many hunter-gatherer communities, food is considered a public good, because the individual who has less food is 'more willing to fight for it.'

But this theory raises some problems. If food is not given away, then why is there any motivation to go out hunting? Why not just stay in the camp and eat the food that others have collected? We can use a basic model from game theory to see why this is — if food is a public good, then we have the classic *tragedy of the commons* [Ostrom, 1990]. A tragedy of the commons is a social dilemma where each individual would like all other individuals to act a certain way, in this case, hunting, but that each does individually better by acting otherwise. Under normal evolutionary circumstances, we expect the personally beneficial outcome to be selected. Does this mean that the theories of 'taking' food must be wrong?

Here we have an observable phenomena — food sharing — along with two broad hypotheses for why it is shared. One is that humans intrinsically understand that food is the property of the hunter who caught it, and that the can do with it as they see fit. We share it because we expect others to do the same. The explanation that best fits this description is reciprocal altruism [Trivers, 1971]. The second hypothesis is that individual hunters have little control over the catch because it is perceived to be the property of the group as a whole until it is distributed to the members. This hypothesis seems to have problems explaining the motivations of the individuals.

In this chapter, I argue for an alternative theory of food sharing. Firstly I argue that in humans, the ability to recognise a good as being property is not a fixed adaptation that natural selection must work around. Rather, humans have evolved to understand property only through the lens of the social contract. In this sense, the quote from John Locke given at the beginning of this chapter is an inaccurate depiction of the world — there are no 'natural' rights to property. It is not the mode of procurement that makes property, but the social contract that operates.

Secondly, I argue that humans have evolved to negotiate social contracts, in this case property rights, in order to encourage pro-social behaviour, in this case hunting. Therefore I show that the theory of 'taking' the good is compatible with standard evolutionary theory without the tragedy of the commons dilemma. The reason for this is that the social contract connects property rights to previous actions, in the case of this model, whether an individual hunted or not. This model can also be extended to include, for example, different sexes, so that for women to gain access to meat, then they would have to take whatever action the social contract suggested to them in order to gain access to meat.

3.2 My Contribution

In this section, I describe the contribution that this chapter makes first in the context of the food-sharing literature, and then in the context of this dissertation.

In the context of the food-sharing debate, the question that is fundamental to both sides of the debate is *who has control of the food?* Producer control — the idea that food is controlled by the individual who originally hunted it — is premised by the theory of reciprocation. The lack of producer control, as proposed by tolerated theft, seems to provide no direct means of motivating individuals to hunt large game and then transport it back to the group. In this chapter, I investigate this problem of property using formal game theory modelling. Tolerated theft is more specific, since it tries to connect itself to the hawk-dove model of conflict studied by Smith and Price [1973]. However, I disagree with the main conclusion. In fact, taken at face value, the model would conclude that humans would *never* possess any power over what they produce, which seems absurd. I offer an alternative to both solutions. It is possible that rather than the food being controlled by the hunter or by the group as a whole, that its distribution is the consequence of the particular social contract that operates within the group. This theory focuses less on the strategy of the individual hunter maximising their payoff in every interaction, and more on their motivation to follow the rules established by the social contract.

In the context of this dissertation, this chapter contributes to our understanding of social contracts. I argue that food sharing in hunter-gatherers is likely to have been the first social contract. The model that I introduce here is fundamentally important for later chapters, because many of the aspects of the model that I introduce here apply to a wide range of modern day human cooperation.

To this end, in this chapter I do the following:

1. Review the empirical literature on food-sharing.
2. Review the theoretical models of reciprocal altruism and tolerated theft.
3. Introduce a model of food-sharing based on an extension of the hawk-dove model of conflict.
4. Demonstrate that tolerated theft is a single equilibrium of an infinite set of equilibria.
5. Produce a novel theory of food sharing based on the social contract that connects 'work' to 'property rights' within a community.

3.3 Food Sharing in Hunter-Gatherer Societies

Food sharing is an important part of almost all hunter-gatherer societies. Within such societies, a typical hunter, returning to their group carrying the carcass of a catch, will set about distributing the meat to other members of the group, or otherwise, members of that group will approach the hunter and ask for their share, and the hunter obliges. In the Ache, for example, the food may be handed over to an older and well respected male who divides up the carcass between the members of the group [Kaplan et al., 1985], or as in the Ifaluk, Pintupi, Washo, Mbuti, Aka, Efe, Shoshone and Paiute, Lamalera or Hiwi, it would go first to those who have take part in the hunt, and then would be later distributed out to others on an individual basis [Gurven, 2004a]. In the !Kung, it is the individual whose arrow made the kill, rather than the hunter, who distributes the meat among the various households [Boehm, 2009].

Of all of the food that is shared, meat is normally shared the widest. For example, it is normal for a Yanomamo of Venezuela to give away a great deal of any large game that is killed in the forest, but not to share small game or plantains grown in their garden nearly so much [Erdal et al., 1994]; for an Ache to give almost ninety percent of his meat away if he has caught a pig or a monkey [Kaplan et al., 1984]; for a Tiwi of Arnhem keeps only a small portion of a large catch, but a large portion of the smallest game [Hart et al., 1966]. More examples are reviewed by Gurven [2004a].

The fact that meat is more widely shared than other foods is significant. A hunter could, and women normally do, devote their time to catching food that is less calorie dense, but that provides a steadier income. Only when the hunter expects that food caught by others will be shared among the group is it individually profitable go out and hunt large game. This is because humans have evolved to require a steady income of nutrition [Kaplan et al., 2000]. Since there is no way for hunter-gatherers to store large game, particularly in warmer climates, then it is pointless for a hunter to spend time catching large-game, since they will waste most of it when it is caught, and spend the rest of the time hungry. But when the hunter expects that others will also be hunting larger game, and that the meat will be shared then it becomes a better strategy.

To give a sense of the typical social organisation of hunter-gatherer band, among foragers there are typically three levels of social organisation: the ethno-linguistic group (or tribe), which may never assemble in one place; the residential or local group (camp or band); and the daily foraging party [Marlowe, 2005]. Typically, the local groups split into smaller parties to hunt and forage, and then return to camp to share food and sleep. Groups tend to be relatively mobile, and also there is a great deal of fission-fusion (the coming together and separating of groups), that tends to occur on the basis of groups growing too large, or specific events such as quarrelling between families.

Marlowe [2005] estimates the average group size among the one hundred and seventy five "warm climate, non-equestrian" hunter gather groups most likely to have been similar to ancestral humans as being thirty seven. Although Bowles and Gintis [2011] argue that if we we take the average group that an individual finds themselves in[1], rather than the average group size, then this figure is more like seventy seven — a significantly larger number. This is susceptible to the fact that one or two outliers can heavily affect the sample, and Marlow argues that warfare may force groups together where they otherwise would not. He also notes that although rich resources may increase the density of the groups, it does not seem to increase their average size.

Before the seminal work of Kaplan et al. [1984], food sharing was mainly viewed by anthropologists in terms of the *group functional* explanations [Ridley, 1997]. It was generally argued that people in tribal societies shared food with one another in a deliberate egalitarian ploy to eliminate status differentials. The reason being that it stopped individuals attempting to gain status by hoarding food, thereby allowing

[1]If we imagine a group of 10 and a group of 100, then the average group size is 55, but the average size of a group that an individual finds themselves in is equal to $(100 \times 100 + 10 \times 10)/110 = 92$, which is significantly higher.

groups to remain in balance with their local environment.

But Kaplan et al. [1984] argued that this sort of reasoning was leading anthropology down a blind alley. Rather than focusing on societies as a whole, they argue, we should instead focus on the pressures and motivations of individuals within societies. After all, it is individuals who make the decision to share. Their work focused on the Ache of Paraguay [Kaplan et al., 1984, 1985, Hill et al., 1993], a small tribe of hunter-gatherers who, until recently, depended entirely on hunting and gathering. In the 70s, they were relocated by the Paraguayan government into mission camps; but in the 80s they still spent about one quarter of their time on hunting trips. During these trips, a small group will head out into the forest searching for game and honey. If they find honey, the women are called over to collect it. In the afternoon, the women set up camp and collect the food from the nearby forest. The men normally return with game such as monkeys, armadillos, pacas and occasionally larger game.

What is remarkable about the Ache is the basis on which food is shared. Although in the settlement they share mainly with the family, during extended hunting trips in the forest, they share freely with kin and non-kin alike [Gurven et al., 2002, Kaplan et al., 1984, 1985]. The man who returns empty handed from the forest is not left out of the feasting, but receives the same shares as every other individual in the group. Since the early studies on the Ache, a handful of ethnographic studies have been conducted on food sharing, 45 in total, collated by Gurven [2004a]. Gurven finds that while there are local variations in how sharing is organised, most of these examples follow a similar pattern — the hunters who return distribute the food extremely widely among kin and non-kin alike.

Kaplan et al. [1985] used to argue mainly that food is shared in exchange for future food. That is, individuals within the population give because they believe that, at a later time, they will receive in return. Previously they have focused on models of reciprocation, but more recently they have begun to consider sharing as a group-wide phenomenon, arguing that individuals do not share dyadically and strategically, but rather as share as part of a wider social norm [Kaplan et al., 2005].

However, there is also another theory based on an entirely different set of assumptions called *tolerated theft* [Jones, 1991, Blurton Jones, 1984, Bird, 1997, Winterhalder, 1996]. This alternative theory grew out of Isaac [1978]'s work on the origin of food sharing. Isaac argued that, like the behaviour of many other animals, food is not so much shared by humans, but that we simply tolerate thieves. For example, when lions and wolves feast on a kill, they are effectively scrounging of the kill that was made by another, and their presence is tolerated because it is too much work to chase them away. Jones [1991], Blurton Jones [1984], Bird [1997], Winterhalder [1996] carry this argument further suggesting not only did this dictate the lines over which food was shared in our past, but that food is still shared like this today. In this view, the food is like a ticking time bomb — an individual wants to give it away before a fight erupts, and it is this, and not sharing, that explains the wide distribution.

But if this is true, then why go through the bother of hunting at all? If food is a public good, as this theory implies, then why not just steal what everybody else has hunted and not hunt yourself? Hawkes [1992], who argues in favour of the 'taking' theory, turned to the work of economists Olson and Olson [2009] who argue that public goods could be provided if there are sufficient incentives to do so. What is this incentive? Hawkes argues that it is provided by the females of the group who prefer to mate with males who are good hunters. Hunting provides a 'theatre' where men can compete for attention and show off their skills. This theory relies on the a theory of sexual selection called *costly signalling theory* [Zahavi and Zahavi, 1997, Bird et al., 2001, Smith and Bird, 2000, McGuire and Hildebrandt, 2005, Hawkes and Bird, 2002]. The basic idea of costly signalling theory is that there is consistent variation in genetic quality among individuals within one sex of a species. But this quality is imperceivable. In order to display quality, males must perform some difficult task, something that a weaker individual would be unable to do. This provides the female with an honest signal of quality, one that cannot be faked. Hawkes believes that hunting is such a signal in humans. She further argues that the reason that food sharing is used in particular is because it draws attention to the hunter. Everyone likes receiving food and it will be easy to recall who it was that provided it. This theory is certainly a possible explanation for food-sharing, in this chapter I choose to focus on a different explanation. However, I return to it in the discussion.

As we have seen, the question of exactly why humans evolved to share remains contentious. Human sharing looks nothing like the patterns of food distributions found in other primate species, where distributions are heavily influenced by dominance. While dominance may still influence food distribution in some small scale societies [Turke and Betzig, 1985, Betzig and Turke, 1986, 1992, Turke, 1989], it is nowhere near as marked. So the question remains, why do people share food?

3.3.1 Current Theories of Food Sharing

I have mentioned the various different theoretical models of food-sharing throughout this chapter, and before I go in to investigate the mathematical models, I bring them together here for reference.

Kin Selection [Smith, 1964, Hamilton, 1964, West et al., 2008] Because biological kin have a higher probability of sharing identical alleles by descent, kin-selection predicts that food sharing should favour biased transfers toward relatives. This can explain a great deal, since and is likely to be important, given that small-scale societies tend to be organised around kin groups. However, the explanatory power of this hypothesis is ultimately limited, because in many populations food is shared widely with non-kin as well as with kin [Gurven, 2004a].

Reciprocal Altruism [Trivers, 1971, Axelrod and Hamilton, 1981, Gurven, 2004b] We saw the basic model for food sharing in the previous section. Applied to food sharing, reciprocal altruism predicts that food is given on the understanding that it, or other favours, will be received in return at a later date. Therefore, if food sharing were operating along these lines, then we would should expect future returns to be somehow contingent on past sharing. As a theory of food sharing, reciprocal altruism represent a sort of health insurance, where short-term costs are suffered in order to gain long-term benefits. These benefits largely accrue from the fact that humans have evolved to require highly nutritious food on a regular basis, and such a regular income of food cannot be achieved on the basis of individual hunting strategies. It should be noted that the strictly dyadic tit-for-tat model is not the only model of reciprocal altruism — it may also be a sort of indirect reciprocity, where individuals choose to share because they want to develop a positive reputation [Nowak and Sigmund, 1998]. Trade is also a kind of reciprocal altruism — one where different goods are given. Therefore explanations like 'meat for sex' [Gomes and Boesch, 2009] also fall into the category of reciprocal altruism.

Tolerated Theft [Jones, 1991, Blurton Jones, 1984, Bird, 1997, Winterhalder, 1996] Tolerated theft argues that food is shared because it is impractical to defend sizable resource packages. That is, when large game is caught, it often contains far more calories than an individual can consume before it turns bad. In this case, there is an asymmetry in the value of food between the individual who has caught the food and his peers. He therefore gives the food away before it is taken from him through violence. This model is very similar to how pack animals like lions share their food.

Costly Signaling [Zahavi and Zahavi, 1997, Bird et al., 2001, Smith and Bird, 2000, McGuire and Hildebrandt, 2005, Hawkes and Bird, 2002] The theory of costly signaling hypothesises that food sharing is not about the value of the food itself, but rather about showing off genetic quality by hunting difficult-to-catch animals. The subsequent distribution of food is a supplement of this strategy, rather than an end in itself. In this theory, distribution of the food is a means of drawing attention to the skill of the hunter, rather than about providing nutrition.

Collaborative Acquisition [Gurven, 2004a] This is the theory that food is shared because it is caught collaboratively, meaning that individuals do not know who the owner is. In many ways, this is the similar to the theory of tolerated theft — suggesting that individuals do not have individual control over the catch because they cannot defend it.

3.3.2 Data Unaccounted for by Existing Models

As previously mentioned, many anthropologists argue that food is not a form of exchange, and is not therefore covered by the model of recirocal altruism. Instead, they argue that when the meat is being distributed, the recipient is claiming their right, rather than the donor giving the food voluntarily. For example:

> Generosity is not stressed. We often think of sharing as deriving from generosity. The emphasis in these societies is quite different (the !Kung and the Hadza). Shares are asked for, even demanded... People believe that they are entitled to their share and are not slow to make claims. The whole emphasis is on donor obligation and recipient entitlement. The donor has no choice over whether the meat is shared... Typically the donor is not thanked. This is consistent with the notion that the donor is doing no more than he should. [pp.49 Hann, 1998]

It seems to be this feeling of 'right' that makes anthropologists feel that the food is not given strategically, and that the hunter has little control over where the food goes. It seems to be this consistent feeling of the emphasis on the recipient demanding their share rather than the donor giving that underpins the theory of tolerated theft.

Another example of this reasoning can be found in the following:

> Collective ownership of resources changes into individual ownership by the mode of procurement and characteristics of the resources. Resources that come in small units are accessible to any able-bodied adult on a regular basis, and have low risk of failure on each procurement episode (such as, but not limited to, vegetable foods); these are transformed from collective into individual ownership through foraging. Resources that come in relatively large units are differentially accessible by adults according to individual skills, and have high risk of failure on a given procurement episode (such as game animals); these resources are considered to be collectively owned through hunting. For the latter, ownership changes from collective to individual according to culturally specified rules of sharing that remove decisions about sharing from the individual hunter to the group as a collectivity. Among the Ju'hoansi, for example, the cultural rule is that the owner of the arrow that killed the animal (who need not have been present during the hunt) distributes the meat from the animal. Among the Netsilik Inuit, seals killed in winter hunting through their breathing holes in the pack ice were distributed in accordance with a culturally constituted

system of "sealing partners." Cultural rules like this make meat-sharing a group-level, not an individual-level, trait [Read, 2012].

In fact, I believe that many of the accounts of tolerated theft also fall into this category—they argue that hunter-gatherers share on an even basis because they have to. However, the problem with this tolerated theft/collective action model of food sharing is that it seems to imply that humans behave altruistically by providing the food in the first place. That is, if hunting is energetically costly, and each individual has equal access to all shares simply by hanging around the camp, then the question is why are any individuals motivated to hunt? Why not save their energy by hanging around the camp, saving their energy and eating the food that others have produced? This is why tolerated theft implies the collective action problem.

3.3.3 Conditional Property Rights — The Theory

In order to account for these instances of food sharing, a theory other than reciprocation is required. To introduce a third option, consider the following two scenarios. Firstly, two individuals meet and decide to cooperate with each in food sharing. Each gives food to the other individual when they require it, and the relationship continues indefinitely. They understand that if they do not give, then their partner will respond by not giving. Such a relationship might be modelled by the repeated prisoner's dilemma. Now consider this second scenario. Two individuals meet and decide that they want to share food. They enact a social contract that says that, as long as both individuals attempt to hunt food, they both have property rights over a portion of the others catch.

This model can explain why the individual who demands the food seems to be collecting their property, rather than asking for a favour — because in a sense, they are collecting their property. However, this description also avoids the problem with public goods, because it implies that no property rights are given to individuals who do not hunt. I call this theory *conditional property rights*.

My next task is to show how this theory is consistent with standard evolutionary theory. To do this, I first introduce the basic model of conflict, explain how this model fits with the theory of cultural evolution, before finally introducing the novel model of food sharing.

With the theory of conditional property rights, we have to consider two things. Firstly, what is the condition; secondly, what does the right guarantee. If we take our cue from cooperative acquisition, we see that it is the act of hunting that ownership of property is conditioned on. Just like how, in the bourgeois strategy of animal behaviour, being first to the site determines the territory, so in this model does hunting make property. The second problem is what exactly it gives rights over. The answer is some portion of the catch. Exactly how much of the catch might depend on whether you made the catch or not, or the particular characteristics of the social contracts

In order to go further, it is necessary that we ground this discussion in the theoretical literature. The fundamental problem is the question of who controls the catch and why. To this end, I now review the theoretical literature on the evolution of conflict.

3.4 A Model of Conflict

In this section, I begin by reviewing a model of conflict. I then show how this model can be can be brought into the social contracts framework that I introduce in section 2. Finally, I extend this model to show how it can be used to explain food-sharing in hunter-gatherer populations.

The fundamental problem of property ownership is captured by the *hawk-dove* game of conflict, which is sometimes called the game of *chicken* in game theory [Smith and Price, 1973]. Imagine there is a resource that two individuals will fight over but only one can use. With respect to food-sharing, this could be a theoretical unit of food. Individuals in this population may play one of two strategies: *hawk* or *dove*. The hawk strategy represents the aggressive strategy of an individual who is willing to fight, while a dove strategy represents a passive strategy who will not fight if their opponent challenges them. When individuals meet and play one these two strategies, we assume that the outcome of the game is as follows:

Hawk vs. Hawk : When two hawks meet, they both aggressively try to take the resource. Assume that they fight until on of them gets injured. Proximately there may be many reasons why an individual wins the resource, for example, they may be stronger and so on. However, since the result of the fight is uncorrelated with anything else in this model, we can assume that there is an even chance that either player will win.

Hawk vs. Dove : When a hawk plays against a dove, the hawk strategy plays aggressively and the dove plays passively, resulting in the hawk taking the resource and the dove getting nothing. So when these two strategies meet, no fight occurs.

Dove vs. Dove : When two doves meet, neither plays aggressively, so no fight ever occurs. We simply assume that the two individuals wait until the other leaves. Since the individual who leaves first is uncorrelated with anything else in the model, we assume that the resource is taken by an individual at random.

Formally, suppose that there exists a resource of value v over which two individuals contest. Each individual must unilaterally choose one of two strategies to play: *hawk*, which represents fighting, or *dove*, which represents retreating. When both agents play *dove*, then we assume all the resource randomly goes to one player, so the expected payoff to each individual when this occurs is $v/2$ (we can also assume that the resource goes to no player where they both receive a payoff of 0 without qualitatively altering the model). If one player plays *hawk* and the other *dove*, then the individual who played *hawk* gains the resource, giving them a payoff of v, and the individual who plays *dove* gets nothing, giving them a payoff of 0. Finally, if both players play *hawk*, then we assume that they fight until one player is injured. The injured player receive a payoff of $-c$ and the winner gains the resource gets a payoff of v. Therefore, each player gains an expected payoff of $(v-c)/2$ when they both play *dove*. The standard assumption in the classic hawk-dove game is that the cost of fighting is greater than the value of the resource ($c > v$). Were this not the case, then we simply have a prisoner's dilemma, another type of game. We can record this information in the following matrix

	Hawk	Dove
Hawk	$(v-c)/2$	v
Dove	0	$v/2$

where the number inside the box is the payoff received by the player on the horizontal.

It is easy to show that there are 2 pure Nash equilibria in this game. Recall that a Nash equilibrium is defined to be a pair of strategies where neither players will benefit from unilaterally altering their strategy. These are the strategy pairs $\{hawk, dove\}$ and the pair $\{dove, hawk\}$. There is also a *mixed equilibrium strategy*, where each player plays the mixed strategy σ_m: '*play* hawk v/c *of the time. Otherwise play* dove.'

So far this is not an evolutionary game, but a game based on rational choice. In order to make it an evolutionary game, we have to make some assumptions about how the population evolves. The original model adopted by Smith and Price [1973] assumes the following.

1. Assume a large population (strictly infinite) which is made up different strategies from the strategy set S.

2. Every round, each individual is paired at random with another individual from the population to play a hawk-dove game.

3. The average payoff from this game becomes the fitness of that type.

4. The population frequency of types in the population change according to the fitness of the different strategies.

5. The process repeats.

If we consider σ_h and σ_d which represent the strategies *always play hawk* and *always play dove* respectively, then we get the following result.

Proposition 4. *If $S = \{\sigma_h, \sigma_d\}$, then the only ESS is a polymorphic equilibria where a proportion v/c of the population play σ_h and the rest play dove.*

The proof of this is simple and can be found in Smith and Price [1973]. To get a basic sense of how this works, notice that in a population of all doves, and individual hawk has a higher fitness, since

$$w_{\sigma_h}(\sigma_d) = v > v/2 = w_{\sigma_d}(\sigma_d) \tag{3.1}$$

and likewise, in a population of all hawks, and individual playing dove does better, since

$$w_{\sigma_d}(\sigma_h) = 0 > (v-c)/2 = w_{\sigma_h}(\sigma_h) \tag{3.2}$$

Furthermore, notice that when there is a chance p that an individual will meet a hawk, where p is the frequency of hawks in the population, then the fitness for playing the two strategies is equal to

$$\begin{aligned} W_h &= p w_{\sigma_h}(\sigma_h) + (1-p) w_{\sigma_h}(\sigma_d) \\ &= v - p\frac{v+c}{2} \\ W_d &= p w_{\sigma_d}(\sigma_h) + (1-p) w_{\sigma_d}(\sigma_d) \\ &= v - pv \end{aligned} \tag{3.3}$$

and $W_h = W_d$ when $p = \frac{v}{c}$.

Neither a pure population of hawks nor a pure population of doves is evolutionarily stable, meaning that we should assume that the population will evolve to a stable state that is a mix of hawks and doves.

3.4.1 Uncorrelated Asymmetries

In the hawk-dove model of conflict, with the strategies σ_h and σ_d, we've seen that the only possible equilibrium is where the frequency of hawks is equal to v/c. In this section, we consider another type of strategy called an *uncorrelated asymmetry.* This is also known as a *bourgeois strategy* in the evolutionary literature.

The idea of an uncorrelated asymmetry, which was first conceived by Smith and Price [1973], has the following properties. Suppose that two individuals are about to play a hawk-dove game. Before the game begins, they both make an arrangement whereby they both observe an event and play their strategy based on this event. The important feature is that the events that both individuals observe may be correlated. An uncorrelated asymmetry is one where it is in neither player's interest to cheat and play a strategy other than what they arranged.

This model was first used as an evolutionary model to explain why it is that the individual who is defending a territory is more likely to win a contest [Smith and Parker, 1976]. Smith and Price [1973] found that in many species, being the first to find a piece of territory often settles disputes over territory. For example, butterflies monopolise patches of sunlight on the forest floor, and they use this mechanism to determine the owner. In this game, the asymmetry is 'who is defending and who is attacking,' and the strategy suggests that the defender plays hawk and the attacker plays dove. This was represented by first a random move by nature which determined who was the owner in each particular fight. The reason for this is that, from an evolutionary perspective, the strategy does not know which role it is going to take during the game, and there is an equal chance that it could be either.

The model proceeds as follows:

1. Assume a large population (strictly infinite) which is made up different strategies in from the strategy set S.
2. Every round, each individual is paired at random with another individual from the population to play a hawk-dove game.
3. Before the game is made, a random event occurs that names one player *owner*.
4. The players now play their strategy, which may be conditioned on the event.
5. The average payoff from this game becomes the fitness of that type.
6. The population frequency of types in the population change according to the fitness of the different strategies.
7. The process repeats.

Let σ_b be the strategy *bourgeois*, which plays *hawk* when the event names them the owner and *dove* otherwise.

Proposition 5. *If $S = \{\sigma_b, \sigma_h, \sigma_d\}$, then σ_b is the only ESS.*

Smith and Price [1973] gives a full proof of this, but to get an intuitive sense of how this works, notice that the payoff for a bourgeois playing against a bourgeois is always equal to

$$w_{\sigma_b}(\sigma_b) = \frac{1}{2}v + \frac{1}{2}0 \tag{3.4}$$

The strategy σ_h which always plays hawk gets the following payoff

$$w_{\sigma_h}(\sigma_b) = v - \frac{1}{2}c < \frac{v}{2} = w_{\sigma_b}(\sigma_b) \tag{3.5}$$

while the strategy σ_d gets

$$w_{\sigma_d}(\sigma_b) = \frac{v}{4} < \frac{v}{2} = w_{\sigma_b}(\sigma_b) \tag{3.6}$$

To see how it invades, notice that at a mixed equilibrium of hawks and doves, that $W_{\sigma_h} = W_{\sigma_d}$, meaning that any linear combination of σ_h and σ_d will all obtain the same payoff of every other strategy. However, when playing itself, σ_b obtains a higher payoff, meaning that we can use equation 2 to show that it invades from rarity.

3.4.2 Social Contracts

The model introduced in the previous section represents the emergence of property rights in nature. When two players meet, they both recognise an asymmetry and play their strategy according to this asymmetry. Does this model accurately represent human property rights? In some ways, human property rights do seem to work in precisely this way. For example, suppose that two strangers step into a train carriage, one of whom is carrying a bag. We could represent this interaction by a hawk-dove contest over the bag — only one individual will walk out carrying the bag, how do they decide who? Of course, it is not necessary that both individuals be aware that there is a conflict, all that matters is that the bag contains some useful resources that only one individual is going to keep. We know that the strategy that they will use to settle this conflict is likely to be that, all else being equal, the individual who arrived with the bag will leave with it. This has all of the hallmarks of the hawk-dove game — two random individuals using an uncorrelated asymmetry to settle the dispute.

There is, however, one major difference between the evolutionary theory of property rights introduced in the previous section and property rights in humans. In humans, property rights are not solely a biological phenomena. They are also a cultural one. It is not the physical properties of the bag, along with our biologically evolved sensory apparatus that determines whether we recognise something as property. Rather, the fact that we recognise the object as a 'bag' and that this can can legitimately excluded from other individuals are both properties of our culture. They are categories that we have invented and agreed upon.

To see how property rights can be maintained at equilibrium where it is culturally, rather than genetically determined, consider the following simple model.

1. Assume a large (strictly infinite) population.
2. This population is divided into subpopulations of size n.
3. Each subpopulation creates a social contract C, which, in each interaction, determines an uncorrelated asymmetry, naming one player as the 'owner.' Assume that each player has an equal chance of being the owner.
4. Each individual is randomly paired with another individual from his population to play a hawk-dove game.

5. The average payoff for each type is then calculated and the population evolves according to the standard model replicator dynamics in equation 2.8.

The assumption that the player meets another individual from the same subpopulation ensures that they are privy to the same social contract.

In this model, I do not consider the process of cultural evolution at all. Rather, I simply assume that when the social contracts are made, they are created in equilibrium. Clearly in practice this may not be entirely accurate. However, when we consider the rate of evolutionary change of cultural evolution when compared to genetic evolution, it does not make a great deal of difference if we added the initial few rounds until property rights go to equilibrium.

Consider the following two biological evolutionary strategies — *cultural* σ_c, and *acultural* σ_a. Cultural plays according to the social contract C, so long as he does not better by cheating. That is, they play hawk when they are owner and dove when they are not not. Acultural uses another asymmetry, or can play a pure strategy. This leads to the following novel proposition

Proposition 6. *If* $S = \{\sigma_c, \sigma_a\}$*, then* σ_c *is an ESS.*

Proof. In a population of σ_c's the expected payoff for playing σ_c is equal to

$$W_{\sigma_c} = \frac{1}{2}v + \frac{1}{2}0 = \frac{v}{2} \tag{3.7}$$

There are several different ways that σ_a could play. One is that they play to their own, different assymetry strategy. If we assume that the signal for this strategy is uncorrelated with the one that σ_c plays, then when is a population of σ_c's, it's payoff is

$$W_{\sigma_a} = \frac{1}{4}v + \frac{1}{4}\frac{v-c}{2} + \frac{1}{4}0 + \frac{1}{4}\frac{v}{2} = \left(\frac{v}{2} - \frac{c}{8}\right) < \frac{v}{2} = W_{\sigma_c} \tag{3.8}$$

If σ_a plays like σ_h or σ_d, then from equations 3.4 and 3.5 we get

$$\begin{aligned} w_{\sigma_h}(\sigma_c) &= v - \frac{1}{2}c \\ w_{\sigma_d}(\sigma_c) &= \frac{v}{4} \end{aligned} \tag{3.9}$$

□

The purpose of the model in this section is to demonstrate that evolving to judge property rights through the lens of a social contract is entirely consistent with standard evolutionary theory. What I believe often confuses the layman is the idea that the winner of a contest does not have a physical advantage, and that it is possible that evolution would select a strategy which relies on some arbitrary rule. Just because we have an evolutionary model does not mean that only the strongest are going to dominate every contest. What is more, it is not that different from what we observe in other species with respect to property. The only difference is that in other species, the rule that is established to solve the problem is a fixed biological adaptation, for example, defending territory or guarding a mate, but in humans the adaptation is to comprehend and follow the rules of the social contract.

3.4.3 Language and Property Rights

As a brief aside, I want to draw attention to the importance of language in the model above. The model demonstrates that it is possible for a otherwise arbitrary signal, in this case an arranged contract, to settle a fitness enhancing dispute.

The question is, how is the rule communicated? I believe that the answer to this is that we use language. To invent property, a population must first establish a category of what can be owned, for example, meat, a tree, a person etc, and what ownership means in this context — the model suggests some sort of exclusion.

This is what I mean in section 2.5.6 when I described the 'informational content' of language principally being information about the social contract, and not about the external environment. While humans clearly value information about the environment, I suggest that the selective advantage does not derive from increased information about the external environment, but from gaining predictive power of how conspecifics will behave.

But this alone does not explain why, of all the possible different arbitrary signals, language and contracts are selected. To preempt the model in section 3.4.7, I argue that social contracts are only arbitrary when we view them from the perspective of a one-shot encounter, such as the model above. But life is not a series of one-shot encounters. When encounters are repeated, humans can structure interactions so that they bring about positive evolutionary ends. Therefore, language and the ability to create contracts means that humans are able to create social environments that promotes pro-social behaviour.

3.4.4 Tolerated Theft and Producer Control

The theory of tolerated theft is an alternative model of food-sharing. Rather than concluding that humans can use a social contract to determine the distribution of food, it concludes that food *must* be shared evenly. This follows from the idea that the meat is not defensible [Winterhalder, 1996, Hawkes and Bird, 2002] Here I take an aside to investigate the claims of this model, which is based on a slightly more sophisticated version of the hawk-dove model. I claim that this model describes a possible world, but does not represent a necessary set of conditions, and therefore I go on to argue that the equilibrium that it identifies is not the one that humans use.

The basic logic of tolerated theft is that when two individuals compete over food, one individual desires the food more than the other. The model assumes that this individual would be more 'willing to fight' for that food because they value it to a greater degree. This in turn means that the other individual ought to play dove, in effect creating a pressure to give food away before it is taken and causes a fight. This is then coupled with the assumption that the value of food is subject to the law of decreasing marginal returns, meaning that the individual who has more food values each unit less than the individual who has less.

One problem is that it is not entirely clear what connects valuing a resource more to being 'more willing to fight,' to actually changing your strategy. For example, consider an evolutionary model where two individuals meet at random to play a hawk-dove game. In this model, we assume that the individual who is not the owner according to the bourgeois strategy gets a payoff $V > v$ if they win the contest, so that the payoff can now be written:

	Hawk	Dove
Hawk	$(v-c)/2, (V-c)/2$	$0, V$
Dove	$v, 0$	$v/2, V/2$

where the left hand of each entry is the payoff for player 1 who is the owner and the right hand for player 2 and $V > v$, who is not the owner.

Suppose that there are two strategies. The first is called *tolerated theft* σ_t, which plays the strategy *play hawk when they desire the food, otherwise play dove.* We also assume that there is the standard σ_b which plays the bourgeois strategy *play hawk when the owner, otherwise play dove.* Then we get the following result:

Proposition 7. *If $G = \{\sigma_t, \sigma_b\}$ then σ_t is an ESS, and σ_b is and ESS if $v > V - c$*

Proof. To show that σ_t is an ESS, notice that

$$\begin{aligned} w_{\sigma_t}(\sigma_t) &= \frac{V}{2} \\ w_{\sigma_b}(\sigma_t) &= \frac{v-c}{4} + \frac{v}{4} \end{aligned} \tag{3.10}$$

clearly $w_{\sigma_t}(\sigma_t) > w_{\sigma_b}(\sigma_t)$ when $c > v$. To show that σ_b is an ESS, notice that

$$\begin{aligned} w_{\sigma_b}(\sigma_b) &= \frac{v}{2} \\ w_{\sigma_t}(\sigma_b) &= \frac{V-c}{4} + \frac{v}{4} \end{aligned} \tag{3.11}$$

so $w_{\sigma_b}(\sigma_b) > w_{\sigma_t}(\sigma_b)$ when $V - c < v$. □

Of course, this model is completely contrived — there is no reason to assume that the individual who is the owner always prefers the good less than the challenger. The point is that *even if* challengers always value the good more, that is not reason to assume that they would necessarily play that equilibrium.

But the model of tolerated theft is more than just a single game of hawk-dove, but rather it is conceived as a series of many smaller games, and so we need to alter the model to reflect this fact. First I extend the model to accommodate this, but I show in the following section that there is a far neater way to represent this using a bargaining model. Suppose that rather than engaging in a single game of *hawk-dove*, we imagine that the good is divided into pieces of size ρ, and that each piece is fought as a separate *hawk-dove*, so the two players fight $1/\rho$ such games. For simplicity, assume that ρ is selected so that $1/\rho$ is a whole number.

To stay as close to the original theory of tolerated theft as we can, define a new game which is made up of $1/\rho$ separate hawk-dove games. In each of these games, a player plays either *hawk* or *dove*. If one player plays *hawk* and the other *dove*, then the player who played *hawk* gains a proportion ρ of the good. If both players play *dove* in any of the games, then the resource is split. If both players play *hawk* in a particular game, then we say that the game has descended into a fight. The payoff of each player depends on whether or not there was a fight. If there was no fight, then the payoff for individual i is a function of the proportion of the payoff that the individual received $v(p_i)$ where p_i is the proportion of the good that they won. However, if the game descends into a fight, then we say that each individual is equally able to win, and that the winner takes all of the good, and the loser is injured and gets a payoff $-c$. In that case, the expected payoff for fighting is $\frac{v(1)-c}{2}$.

However, this game is quite difficult to analyse, and so instead, I investigate a simpler version of this model in the following section.

3.4.5 Dividing the Dollar

The game of ‘dividing the dollar’ [Skyrms, 1996] represents the same essential idea as tolerated theft in a much neater way. The way to think about this problem is to consider what happens when ρ becomes very small. In this game, suppose that two players meet and they have got to decide how to divide up some resource. Assume for simplicity that one player gets x and the other $1 - x$. If they don’t agree on a split, then they fight over the resource, gaining an expected payoff of $v(1) - k$. This version of the model abstracts away all of the ‘mini-fights’ and represents the whole set of interactions as a single move in a game.

This model was analysed by Skyrms [1996] and I report the conclusions here. Skyrms’ model assumes that $c = v(1)$, and also that $v(x) = x$, so that the benefit each player gets is linear. Again, we can adopt this assumption without qualitatively altering the results. Skyrms considered three strategies — the first one he calls *modest*, σ_m, which plays $x = 1/3$ the second he calls *fair* σ_f, which plays $x = 1/2$ and the third he calls *greedy*, which plays $x = 2/3$. We can record the results that each individual gets in the following table:

	σ_m	σ_f	σ_g
σ_m	1/3	1/3	1/3
σ_f	1/2	1/2	0
σ_g	2/3	0	0

We can read straight off the payoff matrix that σ_f is an evolutionary stable strategy because to invade, a strategy would need a payoff of at least 1/2, but the other two strategies gain a payoff of 1/3 for σ_m and 0 for σ_g. However, there is also a mixed equilibrium between σ_m and σ_g where the frequency of each is 1/2. To show that σ_f notice that

$$\begin{aligned} w_{\sigma_m} &= \frac{1}{2}.\frac{1}{3} + \frac{1}{2}.\frac{1}{3} = \frac{1}{3} \\ w_{\sigma_f} &= \frac{1}{2}.\frac{1}{2} + \frac{1}{2}.0 = \frac{1}{4} \\ w_{\sigma_g} &= \frac{1}{2}.\frac{2}{3} + \frac{1}{2}.0 = \frac{1}{3} \end{aligned} \tag{3.12}$$

However, Skyrms goes a step further and argues that if there is sufficient positive correlation between the types of players who meet, for example from relatedness or limited dispersal, then it can be shown that σ_f, the fair strategy, is the only ESS.

So it seems that Skyrms' model supports the model of tolerated theft. However, neither the model of tolerated theft nor the model of dividing the dollar account for the fact that asymmetries in the disputees can tell them what to play. For example, consider the strategy that plays $x = 1$ when the owner and $x = 0$ when not the owner. We can assume as before that ownership is determined randomly before the dispute begins. Once again call this strategy *bourgeois*, σ_b, and it is easy to see that it cannot be invaded by σ_f

Proposition 8. *When $S = \{\sigma_f, \sigma_b\}$, then both σ_f and σ_b are ESS's*

Proof. To show that σ_f is an ESS against σ_b, notice that in population that is entirely made up of σ_f, then the expected payoffs are as follows:

$$\begin{aligned} W_{\sigma_f} &= \frac{1}{2} \\ W_{\sigma_b} &= 0 < W_{\sigma_f} \end{aligned}$$

and to demonstrate that σ_b is an ESS against σ_f

$$\begin{aligned} W_{\sigma_b} &= \frac{1}{2} \\ W_{\sigma_f} &= \frac{1}{4} < W_{\sigma_b} \end{aligned}$$

Therefore, although tolerated theft is an ESS, so is the 'dividing the dollar' version of the bourgeois strategy. □

As we have seen, these two models of tolerated theft and Skyrm's 'fairness' both predict that players ought to settle disputes fairly, but that in fact, the bourgeois strategy is also an ESS in these games, meaning that we cannot deduce that humans 'ought' to evolve to the 'no property' equilibrium.

Of course, we know that ultimately, Skyrms [1996]' model must be wrong, because it predicts that there is never any scope for any property to evolve. This is clearly incorrect for humans. The ESS brought about by the bourgeois strategy is also possible, as is any other number of equilibria.

3.4.6 Is Tolerated Theft Altruism or a Mutualism?

Recall in section 2.5.9, we defined four different types of social behaviour: mutualism (+,+), selfish (+,-), altruism (-,+) and spite (-,-). In this section, I ask the question 'which of these four is tolerated theft?' The answer to this is surprisingly complex.

One possible answer is that tolerated theft is not a direct social interaction at all, at least not from the perspective of the hunter, and therefore none of the four conditions hold. That is, if I kill an animal and then abandon it because it is useless to me, then from my perspective, no social interaction has taken place. This behaviour, if optimal to me, will evolve independently of what happens to the carcass afterwards.

If we decide that we *do* want to call this scenario a social interaction, then because the other individual eats the food that I leave behind, then we can view this interaction as a mutualism (+,+) because we both benefited from my having caught the food.

However, this is not how most scholars view tolerated theft. Instead, they argue that both individuals desire possession of the food, but because of the decreasing marginal value of a unit of food [Winterhalder, 1996], the individual who has less gains more than the player who has more. So in this model, anything that the thief gets lowers the payoff of the player who has more, which means that the interaction is (-,+), which is altruism. Granted, the '-' here is smaller in magnitude that the '+', but it is still assumed that it is greater than 0.

Yet another problem is that, because of the nature of the game, the payoff that a certain strategy gain depends on what the other player does, in a way that it does not with a model of the prisoners dilemma. For example, if a player $x = .7$ in the dividing the dollar game, then whether this is a mutualism (+,+) or spite (-,-), depends entirely on what they other player chooses to do.

3.4.7 Conditional Property Rights — The Model

To briefly review what we have seen so far. In section 3.4, I introduce a model of conflict based on the hawk-dove game. In section 3.4.2 I then show how one of the solutions to the hawk-dove game, the famous bourgeois strategy, can also be easily applied to a cultural setting. In section 3.4.4, I show that tolerated theft, the primary objection to the model that I put forwards, is only one of an infinite number of different equilibria, and that by itself it does not invalidate the model of property rights.

In the present section, I demonstrate how this basic model of property rights can be extended to explain the theory of conditional property rights introduced in section 3.4.7. Recall that I argued that in many populations of hunter-gatherers, a common means of organising economic activity is to implicitly connect property rights to hunting. The task at hand is to represent this as an equilibrium of a game.

Note here that I am only modelling men's shares here. Because there is typically a division of labour between males and females in human hunter gatherer groups, this model does not apply to female shares, however, I discuss later how the model can be easily reinterpreted to include things like female shares.

To formalise these ideas, first we assume that a population of n individuals form to play a hunting game. The first decision that the individuals must make is whether to hunt or not. Suppose, that doing so brings a personal cost k. We assume that the act of hunting creates some good of value G. Next assume that the contract C distributes property rights to the individuals within the population based on their action in the previous round. For simplicity, assume that each hunter is given G/n of the goods, where n is the number of individuals in the group.

C is going to be the contract that gives property rights to those who hunt. To test for stability, we need to check no alternative strategy will invade from rarity. To deal with this question, suppose that these property rights have been established. What we need to know here is does any individual benefit from from deviating from C? There are many ways that we could choose to represent this in game form. For example, we could assume that every individual meets every other individual, or that every individual meets a subset of the population and so on. I opt to assume the simplest setup — suppose that each individual plays two games of hawk-dove, one when they are the owner, and one when they are not. This represents the fact that everyone is a potential owner and everyone is a potential rival for the good. I also assume that if the individual did not hunt, then they do not play the hawk-dove when they are owner.

So the game goes as follows:

1. Form a population of $n + 1$ individuals.
2. Each individual chooses whether to hunt or not. Hunting costs k.
3. As a result of the hunting, there are goods of value G.

4. Assume a social contract where each individual who hunts receives property rights over a package of goods worth G/i, where i is the number of players who hunt.

5. All individuals who hunted now play a hawk-dove game against a random stranger from within the group where they are the owner.

6. All individuals now play a hawk-dove game where they are not the owner.

7. The payoffs from the two games and the hunting represent the total payoff, and the population evolves according to the assumptions of equation 2.8

Now we have to decide what the possible strategies are. Assume that the strategy bourgeois σ_b plays *hunt, play hawk when owner and dove when not*. Suppose also that there are strategies *thief* σ_t which plays *hunt and play hawk when not owner*, *coward* σ_c which plays *hunt and play dove when owner*, and *lazy* σ_l which plays *do not hunt*. The aim is to show that σ_b is evolutionarily stable out of these. For ease of reference, I show the different strategies in the following table:

Strategy	Hunt?	Owner	Incumbent
Bourgeois σ_b	yes	hawk	dove
Thief σ_t	yes	hawk	hawk
Coward σ_c	yes	dove	dove
Lazy σ_l	no	n/a	dove

Define $w_\sigma(i)$ is the payoff that an individual receives for playing σ in a population of i σ_b's, then we get the following result:

$$\begin{aligned} w_{\sigma_b}(n) &= G/n - k \\ w_{\sigma_t}(n) &= G/n - k + (G/n - c)/2 \\ w_{\sigma_c}(n) &= G/2n - k \\ w_{\sigma_l}(n) &= 0 \end{aligned} \tag{3.13}$$

Then we can easily prove the following result

Proposition 9. *If $S = \{\sigma_b, \sigma_t, \sigma_c, \sigma_l\}$, then as long as $G/n > k$ and $c > G/n$, then if we consider pairwise invasions, σ_b is an ESS.*

Proof. To show that σ_b is evolutionarily stable against the thieving strategy σ_t, notice that

$$w_{\sigma_t}(n) = G/n - k + (G/n - c)/2 < G/n - k = w_{\sigma_b}(n) \tag{3.14}$$

That is, the strategy σ_b is evolutionary stable against the invasion of σ_t. We can do the same for the other strategies in equation 3.13. □

Notice how this model operates. It does not assume that there is any round of 'giving,' where players must decide whether to share food or not. Instead, it assumes that food is 'claimed' by the fact that the individual hunted.

3.5 Empirical Validation

In this section, I validate the model against the ethnographic literature on food sharing. The method, as outlined in detail in section 2.6.4, is to take an ethnographically well-researched example of food sharing and try to estimate the relative magnitudes of the various factors. Because the magnitudes are based on 'utility,' which I have argued is essentially a proxy for evolutionary fitness, then ultimately, most of the important units would be in fitness or time. Theoretically it should be possible to measure everything in these units of fitness, but practically that is likely to be very difficult to do, so instead we must make estimations based on observation.

Also before I apply the method, I should explain why I validate the model against a single ethnographic example and not take a cross cultural view. The reason for this is outlined in section 2.6.3, that I treat each social contract as an independently evolving entity. This means that I am not seeking to explain the invariant properties of human behaviour, but the ones that are culturally variant. In turn this means that each time we apply the model, we must check each assumption is valid. Of course, one model may apply to two different cases, just as one evolutionary theory can explain two different biological examples, but I believe that it is better to err on the side of caution and assume that the model is incorrect until it has been checked. Aside from this point, the discussion running throughout this chapter already provides an overview of the general movements in the field. Here we are more interested in the particulars.

I apply this model to understand the specific example to explaining food-sharing in the example of the Ache of Paraguay. There are a number of studies on the Ache of Paraguay detail food-sharing norms among the Ache [Hill, 2002, Kaplan et al., 2009, Gurven et al., 2002, Kaplan et al., 1984]. All of the information that I give on the Ache are derived from these sources. The Ache are a population where meat sharing is particularly wide. Game makes up 60–80% of the hunter's diet, with honey, fruit and other less calorie-dense food making up the rest. On average an individual will share 74% of the total calories that they collect, with meat being the most widely shared single food item, with an average of around 85% shared. Most interestingly, many studies of the Ache found no significant evidence of direct contingency in giving and receiving between individuals within the same group, meaning that, within a hunting band, amount of food given at one time does not predict the likelihood of food being given in return.

Is the food distributed evenly accross all members of the group? (Does each individual gain $1/n$th of the good G?) Food sharing occurs after the day's hunting. in a series of detailed quantitative investigations, Kaplan et al. [1985] empirically test whether there is direct contingency between the amount of food given to a specific individual, and the amount received from that individual. What they find is interesting. With fruit, honey and other non-meat food types, the amount of food received from a specific individual was a statistically significant predictor of receiving food from them. This leads Kaplan et al. [1985] to conclude that is likely to be due to a strong reciprocal effect. However, with meat, the amount given previously was not a statistically significant factor, and so reciprocal altruism was ruled out. This means that there seems to be a stable social contract where all individuals share all of the meat evenly, that is not based on bilateral, reciprocal deals.

Another interesting point is the manner in which meat is distributed to the members of the group. Typically, the men return to the camp with their catches, which are then gathered together and distributed by a respected individual within the company, usually an older male. This means that due to the institutional arrangements, once the meat has been handed over, individual hunters do not have direct control over where their catch goes.[2]

Kaplan et al. [1985] also find that there are relatively wide variations in the amount of food that individual hunters catch. While all individuals seem to put the same amount of effort into hunting, the amount of calories they received differed significantly. Kaplan et al. could not be sure that this was not due to the limited time period and the variable returns from hunting (the study lasted several months, a more complete one would have to extend a number of years to establish this), however, but if we assume that there is variation in the skill of the hunters, then this seems to be strong evidence to support the assumption that all individuals receive an equal share, regardless of their particular relationship to the hunter, including being the hunter.

Are individuals who do not hunt left out of the distribution? This is a difficult question to answer because it involves asking a counter factual. That is, Kaplan et al. [1985] did not observe individuals choosing not to hunt, and so we do not have any direct evidence whether this is the case. However, there are two pieces of evidence that we can use to support this theory. Firstly, when asked what would occur if an individual did not hunt, the Ache said that they would be excluded from the sharing.

The second reason is more direct — teenagers sometimes do not hunt, but instead laze around the camp. When this occurs, the elders of the group collude to make sure that they are not given full shares. It is interesting to note that this is contingent on effort and not ability, for teenagers are generally bad hunters when compared to the

[2]This property is also the same with the !Kung, where the owner of the arrow distributes the food, rather than the individual who made the catch.

old individuals in the group. The Ache also say that they are ‘teaching the teenagers a lesson’ when they do this [Kaplan et al., 1990].

There many proximate questions left unanswered here. The main one is *how* are slackers excluded from the food? Is the decision taken unilaterally by the leader, or does it require ascent from the rest of the group? How is the information communicated to the slacker? Clearly these are the kinds of questions that ethnographers aim to provide answers for, but I believe that the use of theoretical modelling can help guide these questions by making it clear what we know, and what we assume.

Is the cost of fighting greater than the value of food? ($c > G/n$)? It is very difficult to estimate the lifetime fitness effect of fighting over going for a day without food. The best way to gain any purchase on this is to look for indirect evidence.

There are several reasons why it seems likely that the marginal fitness costs of fighting are higher than the marginal gains from the food in a single encounter. Firstly, humans are quite fragile when compared to other social primates like chimpanzees [De Waal, 1982]. Our bones break easily, and we risk infection. To make matters worse, we are adapted for using tools for hunting, such as bows and axes, and these tools greatly enhance our ability to cause damage to other individuals. These means that even individuals who are physically quite weak can still cause considerable harm to another.

A second reason why fighting is likely to be net costly is that humans generally rely on their mobility, and our ability to use tools and instruments, the sort of thing that a broken toe or finger is likely to render us unable to operate.

Obviously if another individual is starving, so that the alternative to fighting is death, then it may be that the benefits tip the optimal strategy in the direction of aggression. However, since food is shared equally in the Ache, to a first degree of approximation, everyone will be equally well fed, so that no one individual will be sufficiently starvings to fight.

Is hunting meat worth it? ($G/n > k$?) It is generally recognised that humans require a highly nutirtious diet in order to survive [Kaplan et al., 1985, 2000]. Meat is the most calorie dense food. However, it is unclear whether hunting meat would be an optimal individual strategy for a man adopting an entirely individual hunting strategy. The reason for this is that the most reliable way to collect food is to gather — even the best hunters return home empty handed around sixty percent of the time [Kaplan et al., 1985]. For the worst hunters this could be much higher. However, an individual does not simply want to maximise the number of calories that they gain over a long period. Rather, a steady income of food is far more likely to be important. Meat cannot be stored for a long time in the forest, and so the hunter and his family who catches a large animal must leave much of it uneaten. For example, an average peccary

— a breed of pig that is hunted by the Ache, contains tens of thousands of calories, so a hunter only requires a small portion of this to feed himself and his family.

The reason that this is important is that if there was no guarantee that the meat would be distributed throughout the group, then it is quite possible that hunting would not be the best strategy for an individual, and it may be that a more steady income of food would be optimal. However, when many individuals pool their risk, then it is likely that the constant income of food means that it is worth hunting and sharing.

3.6 Discussion

The idea of this chapter is to explain some of the accounts of food-sharing in small-scale societies that do not conform to any of the previous models. In particular, many anthropologists argue that the model of reciprocal altruism does not apply to some examples, because the food does not seem to be 'owned' by the hunter. However, the previous alternative — tolerated theft — implies that humans are evolved to behave altruistically towards non-kin, which seems unlikely because of the relatively low relatedness of the members of human groups. But I argue in favour of a third way — that food is not owned either by the individual who caught it, or unconditionally by the group. Rather, I argue that the distribution of the food is a consequence of the social contract that operates.

If this theory is correct, then it has some important implications. Theories such as reciprocal altruism focus primarily on food sharing from the perspective of the individual maximising his own income. But the theory of conditional property rights limits the amount of power an individual has to act. Instead, it highlights the importance of the rules that society has created. Of course, the individual must do better by acting according to the rules of the contract, which is what the model demonstrates.

I believe that the difference between the two processes can be likened to the modern day difference between cooperation in the free market, and cooperation within a company. On the market, it is sensible to assume that individual try to purchase as cheaper deals as possible by finding the best partners. However, within a company, this idea of maximising payoff in each interaction is far less applicable, since within companies, dyadic interactions are much less important. But we can still investigate the structure of these groups from the perspective of individual motivation, we have just got to be more careful that we have correctly identified the structure.

The most obvious way to extend this research would be to apply the model to different groups to see if the assumptions hold. The Ache are a particularly well studied group with a wealth of careful quantitative data for them, so they make an excellent test case. However, Gurven [2004a] reports that in other groups, there is a great deal more contingency between individuals.

One important point is that I only discuss hunting and gathering in terms of the meat that men produce through hunting. However, calories obtained through female gathering can account for a large portion (possibly even greater than 50%) of the total calories that members of the group can consume. However, the model is easily extended to include this point. In the model, we assume that there is a cost to hunting, which provides a benefit. Obviously there may be differences in the costs and benefits of any particular case, but this is no reason

3.7 Conclusion

The question that I set out to answer in this chapter is whether, according to standard evolutionary logic, it is possible to have a stable social contract for food sharing where the food is implicitly owned by every individual who goes hunting. The reason for this is to explain the observations in the anthropological literature about the nature of distribution. Many anthropologists argue that the popular model of reciprocal altruism is incorrect, because this theory places far too much emphasis on the individual hunter 'optimising' his distribution strategy to maximise his payoff. Instead they argue that hunters often have little control over their own catch, either for implicit reasons, like demanding a share, or explicit reasons, such as with the Ache where the meat is distributed centrally, or in the Ju'hoansi where the meat is distributed by the hunter who owns the arrow, rather than the one who made the kill.

However, I argue that the model that was put forwards as an alternative to this, the model of tolerated theft, was incorrect, because it could not account for the existence of shirkers.

Instead, I propose a new model. In this model, humans establish a social contract that determines how food ought to be shared. The idea is that humans use property rights to encourage pro-social behaviour. It is the contract that tells individuals what food they ought to defend and what they ought to leave alone. Human design the mechanism in such a way that it is in the personal interest of all to hunt.

CHAPTER 4

PUNISHMENT AND THE SOCIAL CONTRACT

Lets have faith that right makes might; and in that faith let us, to the end, dare to do our duty as we understand it.
— Abraham Lincoln, *The Cooper Union Speech*

The conch exploded into a thousand white fragments and ceased to exist
— William Golding, *Lord of the Flies*

4.1 Introduction

In the previous chapter, I argue that humans have evolved to use social contracts to encourage cooperation. I investigate the role of property rights, which are important to all models of cooperation that invokes exclusion from economic goods as an explanation. Rather than considering food sharing from the perspective of an individual distributing his property as he sees fit, the model invites us to consider food-sharing as a group level process that converts work into property. Within this system, there is relatively little scope for 'strategic sharing,' since the social contract dictates the lines along which food is shared. However, we must still do some more work to fully understand how food has come to be shared, and also to understand how this adaptation is adopted to solve other kinds of social dilemmas.

Imagine a population made up of individuals who are all adapted to an individual hunting strategy. One day, a hunter fells a large animal and brings it back to the camp. In the model in the previous section, I assume that the animal would be distributed out to the other hunters along the lines of the social contract, and that each hunter will then unilaterally defend their property. However, I judge this explanation to be incomplete, because it does not tackle the problem of how the good is defended before it is divided up. That is, why does an individual from the population who is not a hunter simply walk over and claim a portion of the good for themselves, against the will of the other members of the group?

This is quite possibly the first public goods problem — every individual in the group would like the good to be defended since that maximises their own share, but also, no individual wants take part in defending — they would rather let everyone else do it. So here we see that tolerated theft once again seems like a possibility. Here I turn to the work by Olson and Olson [2009] on collective resource problems, who argues that public goods can be maintained if sufficient incentive can be provided for individuals to do so. Where does the incentive come from here? I argue that it is simply part of the social contract that connects 'work' to 'property.' In the previous section I assume that early humans simply connected hunting to property, but here I extend this and argue that they also made defending the good part of the contract. Any individual who does not engage in defending the good does not have property rights distributed to them.

But doesn't this assume the solution as part of the solution? In a sense it does, but as I demonstrate, this is one of the characteristics of repeated games in game theory that makes it a useful tool — solutions can include the solution as part of the assumption.

I compare this explanation to other models of punishment like 'costly punishment' and 'indirect reciprocity.' I argue that the mistake of both of these models is to assume that in humans, punishment strategies are an individual, biological adaptation. Instead, I suggest that we should investigate these strategies as features of the social contract, rather than from the individual perspective. This theory makes vastly different predictions about what we expect to observe in humans with respect to punishment.

I also seek to explore how punishment explains cooperation other than food-sharing. After all, if I am right that humans are not biologically adapted to using one specific form of punishment, then we expect to observe variations between communities. Therefore, I show how some of the basic observations of this model can be applied to understand a variety of different contracts. I choose to be specific, rather than broad, since as we will see, each group has its own peculiarities that cannot be understood using the broad strokes of game theory alone. There are many systems of punishment that I could choose to focus on, but one of the best documented in the ethnographic literature is Evans-Pritchard's work on the Nuer [Evans-Pritchard, 1940]. The Nuer are a population of pastoralists who have developed a system of fines that must be paid when they defect against many of their social norms. Since I am interested in economic exclusion, this is an excellent case-study for this model. I demonstrate how this model helps us to understand the motivations of all the individuals involved in the Nuer legal process.

4.2 My Contribution

I extend the model in the previous section by including the possibility of disputes that involve more than two players playing simultaneously. I suggest that humans use this to solve the first public goods problem — defending the good itself, then I go on to demonstrate how humans could use this same mechanism to later adapted to prohibit all sorts of anti-social behaviour. I compare this model of punishment against other theories, focusing on costly punishment and indirect reciprocity.

Finally, I use the results of this model to interpret a system of law found in the Nuer — a population of pastoralists who have enacted a social contract that includes a series of fines levied against individuals who break the contract. I argue that these fines follow the same logic as the model of conditional property rights introduced in the previous section and represent a means of enacting punishment that is not costly.

Therefore I will do the following:

1. Review the literature of social contracts and evolution, focusing on negotiation and cultural evolution.
2. Review the literature on punishment in game theory models.
3. Introduce a model of the invasion of property rights, focusing on the importance of moral coalitions.
4. Introduce some variations on this model to widen its possible number of applications.
5. Show how the model of property rights can be used to represent how a system of fines can be used to prohibit certain behaviour.
6. Apply this model to an ethnographic example of the Nuer.

4.3 Reverse Dominance Hierarchies and Moral Coalitions

In order to comprehend human interaction, it is vital to understand that it is not always dyadic, but often, it is multi-lateral. The most obvious example of this that I can think of is warfare. We cannot possibly understand warfare as a series of dyadic contests between individuals. Rather, we understand that warfare occurs between groups of individuals. In this section, I claim that conflicts within groups should also be viewed like this. Often, conflicts are not dyadic and exclusive, but include many individuals interacting simultaneously.

Conflicts between members of human groups are often settled by public opinion. However nebulous this force might seem, and however hard it is to measure, we all understand when it is used against us. It is likely, then, that humans have developed a specific adaptation both for forming public opinion, and also for detecting it. But what is it for? Here I argue that it primarily evolved as a means of promoting pro-social behaviour with the group.

But we are not the only species for which this is true — chimpanzees also seem to have this quality [De Waal, 1982, 2009]. However, in chimpanzees, the disputes tend to be primarily about who is dominant, and public opinion is basically an implicit promise to aid in a fight. This is evident from the fact that when disputes occur, they often involve third parties jumping to the aid of one of the disputees. But in humans, disputes seem to be qualitatively different. Humans tend to settle disputes along moral lines rather than through dominance [Boehm, 2009]. When humans dispute, we put forwards arguments, often in the form of abstract principles, and we work hard to ascertain who is 'right' in the dispute, and we punish those who have acted 'wrongly.'

That is not to say that dominance does not exist in humans at all, but compared to chimpanzees, it is markedly less pronounced. For example, in a chimpanzee troop, a dominant male might enjoy around eighty percent of all of the matings. I know of no human group where the figures are anything like this. In fact, on their list of moral transgressions, small-scale societies regularly place politically overbearing individuals near the top. These would-be dominants threaten the freedom and autonomy of the everyday life of other group members. He, or sometimes she, may become selfish when it is time to share food, he may be a leader who begins to issue direct orders, he may try to make off with another man's wife by force, or one of many more anti-social behaviours.

In his book *Hierarchy in the Forest*, Boehm [2009] argues that humans have evolved to solve these problems by evolving specific anti-dominance adaptations. This adaptation, performed primarily by individuals who would otherwise be victims of aggression, is to form coalitions with the expressed aim of stopping would-be-dominant individuals from dominating. Much social control is most preemptive and subtle. Boehm categorises five types of egalitarian sanction, each more serious than the last.

The first is criticism and ridicule. This may take the form of gossip, or, in more extreme cases, direct criticism. Ridicule is a particularly cutting form of criticism. When an individual becomes overbearing, they will often find themselves the victims of open ridicule from the other members of the group, with laughter being directed at them from other members of the group. As hurtful as ridicule is to the individual, from an evolutionary perspective, it should not be considered a form of punishment. The reason for this is that laughter in itself does not lower the fitness of the victim. Instead, I argue that it should be considered a form of consensus building within the community. If some sanction is going to occur, then it is vital that both the transgressor and the rest of the community are aware. Laughter is easily observable, loud, and all individuals engage simultaneously. This is why ridicule feels so bad. It is an indication that a coalition has been formed against you, and that you must be extremely careful in how you proceed, creating a great deal of anxiety. More often than not, punishment does not go beyond this, because the individual realises that they have lost and alters their behaviour accordingly.

Boehm's second type of social control is openly disobeying leaders who try to command. Leaders are important for hunter-gatherer societies for a variety of reasons, but it is always possible for a leader to overstep their bounds. In this case, ignoring them is a powerful tool. For example, migration decisions can be extremely important to survival in hunter-gatherers. These decisions need to be discussed within the group and a consensus made. Band members often look to their most able members for advice, but this has the potential for that individual to attempt to seize more power for themselves. One way to curb their power is to simply ignore their commands. The examples that Boehm gives are Geronimo's Apache, whose followers simply refused to follow, and the same for the Southern Ute. Among the Batek of Malaysia, members moved away from belligerent leaders. Among the Andaman islanders, a disenchanted minority would simply go their separate ways. But like laughter, disobeying orders is unlikely to significantly hurt the offending individual's fitness directly, but rather acts to signal to the aggressor and other members of the group that they are aware that a transgression is taking place and showing that they can act against it. Laughing and disobeying orders are symbolic behaviours designed to undermine certain individuals.

A third and yet stronger form of sanctioning, and one that is more likely to directly affect fitness, is social ostracism. Milder forms of ostracism manifest themselves as 'the silent treatment' — refusing to engage in conversation with that individual. But more serious forms of include being left out of food distributions, being left out of other favours, or in the most extreme cases, expulsion from the group. In small-scale societies where each individual is dependent on those around, ostracism is extremely costly.

A fourth type of sanctioning, and one that I investigate in more detail in later sections, is fines for bad behaviour. Among the Nuer, for example, there is a fixed fine in place for a series of transgressions ranging from thievery to murder [Evans-Pritchard, 1940, Hutchinson, 1996]. In the case of the Nuer, cattle are paid in reparations to the injured individual (or family of that individual in the case of murder). Other example include Enga war reparations, paid to the families of any dead individuals as a result of conflict [Wiessner et al., 1998].

The fifth and final sanction that Boehm identifies is assassination. Most of our evidence for assassination comes at the tribal level comes from New Guinea. This is unsurprising since, in other parts of the world, colonial powers who sanctioned against homicide pacified the population. But in the highlands of New Guinea, ethnographers arrived soon after contact, and they found several examples of leaders who had been murdered by members of the tribe [Pospisil, 1963, Godelier, 1986]. However, this is the most serious form of sanctioning and extremely uncommon when compared to the milder forms of punishment outlined above.

Boehm provides a great deal of evidence to support the claim that human egalitarianism is at least partly the result of coalitions against aggressive individuals. In this chapter I provide a model for these interactions. However, in this model, I move the emphasis away from the reverse dominance part of the explanation, on to what I believe to be the more fundamental idea — that humans have evolved to defend each others rights as guaranteed by the social contract.

When we express the theory in this way, we see how it implies its contra position — individuals who attempt to break a set of rules established by the group are likely to face sanctioning by its members. But it also applies to a wider variety of situations — not all disputes over the social contract are necessarily dominance moves. Recall that in the framework developed in the preceding chapters, dominance is not the absence of a social contract, it *is* a social contract, albeit one that benefits one individual to the expense of another. What Boehm is arguing is that humans have evolved social contracts that apply equally to every individual.

In this section I have not distinguished between hunter-gatherer populations, and other populations who are pastoralists, farmers and others. While there are clearly large differences between these different types of populations, I have chosen to investigate them together for the reason that they share an important commonality: All of the norms of these communities are, to a large degree, *contained* within that community. That is, in order for them to remain stable over time, they do not rely exogenous forces. The paradigmatic example of such a force is some kind of government. With the force of the government, many behaviours that would be otherwise adaptive can be held in check.

4.4 Repeated Games and Punishment

We have seen that small-scale societies employ many different types of sanctions against anti-social behaviours. In this section I review the current state-of-the art of models that have been put forwards with respect to punishment in small-scale societies. Recall that a social contract is a *self-policing arrangement between members of a society on a particular equilibrium in the game of life,* where the ‘game of life’ is the set of social dilemmas that humans must navigate in their everyday life first described in section 2.4 in this dissertation. The ‘self-policing’ aspect of this description has to do with what we consider an adequate explanation. Both evolutionary and rational choice theory begin with the assumptions that some quantity is being maximised — fitness in the former and utility in the latter. A 'self-policing' mechanism is one where all of the salient forces of the model are accounted for with respect to the maximisation principle. For example, suppose that we ask why individuals from a population perform behaviour X. We could respond that the answer is that it is policed, which creates either enough fitness incentive or dis-utility to make performing behaviour X optimal, according to whichever optimisation maximisation principle we use. However, this 'police' is an exogenous force to our model. It acts upon the population, but does not describe the evolutionary or rational motives of the police themselves. A contract becomes self-policing when we have included the motivations of all individuals who take actions.

In the previous chapter, the contract of the property rights was self-policing because of the structure of the interactions. When a social contract is put in place that connects individual property rights to the act of hunting, then all individuals simply acting in their own self interest will maintain the system, which is designed in such a way as to encourage cooperative sharing.

However, other theories also include the actions of third parties. For example, the theory of costly punishment assumes that in response to certain anti-social behaviours, individual members of a community will be willing to enact corporal punishment against the transgressor. Why should they punish? The theory is that individuals are also adapted to punish those who did not punish. If the cost of being punished is greater than the costs of punishing, then no individual benefits from not punishing. But what keeps this 'third order punishment' going? The answer is fourth order punishment. With these assumptions, it is easy to see that individuals can become adapted to perform behaviour that they would otherwise not.

4.4.1 A Model of Costly Punishment

To sketch this model, suppose that in a life cycle, individuals are placed in groups of size $n+1$ using the standard assumptions in equation 2.13 and that individuals have a choice between two actions A and B, and that they get a payoff of w_A if they play A and w_B if they play B, $w_B > w_A$. Suppose also that after they have chosen A or B, each individual may punish another individual, lowering their payoff by p/n at a cost of k/n to themselves, where n is the number of other individuals within the society. Suppose that the rounds of punishment continue with probability δ. The strategy *punisher* σ_p plays *play A, and then punish any individual not playing A. Also punish any individual who does not punish*, and the strategy *transgressor* σ_t which plays *play B and never punish.* Define $w_\sigma(i)$ as the payoff that and individual playing strategy σ gains when in a population where i individuals play σ_p, and the rest play σ_t

$$\begin{aligned} w_{\sigma_p}(i) &= w_A - \delta(n-i)\frac{k}{n} \\ w_{\sigma_t}(i) &= w_B - \delta i \frac{p}{n} \end{aligned} \tag{4.1}$$

From here we can see that

$$w_{\sigma_p}(i) > w_{\sigma_t}(i) \Leftrightarrow \frac{i}{n} > \frac{w_B - w_A + \delta k}{\delta\,(k+p)} \tag{4.2}$$

Proposition 10. *When $S = \{\sigma_p, \sigma_t\}$, then σ_t is always an ESS, and σ_p is an evolutionarily stable strategy when $w_A > w_B - n\delta p$.*

Proof. Suppose that $i = 0$, so the entire group is playing σ_t, then $w_{\sigma_t}(0) = w_B > w_A - (n-1)\delta k = w_{\sigma_p}(0)$. When $i = n$, so all individual within the group are playing σ_p, then we get $w_{\sigma_p}(n) = w_A > w_B - n\delta p = w_{\sigma_t}(n)$ according to the condition in the proposition. □

However, there is another problem here — the aforementioned problem of 'second order defection.' This problem is that it is possible for a type to invade the population who do not punish those who choose B. This individual will choose A, but will not punish any individual who plays B, thereby reducing their costs. However, this is dealt with by assuming that as well as punishing, individuals also punish those who do not punish. To show that this is evolutionarily stable against non-punishers, we must add a few addition assumptions. Suppose that with probability ϵ, an individual plays B instead of A. Also, assume that after the first round of punishment, there is a δ chance that there will be a second round of punishing the non-punishers. If i is the number of σ_p's we can write the expected payoffs as

$$\begin{aligned} w_{\sigma_n}(i) &= (1-\epsilon)w_B + \epsilon\left(w_A - \delta i\frac{p}{n}\right) - \delta^2\left(1-(1-\epsilon)^n\right) i\frac{p}{n} \\ w_{\sigma_p}(i) &= (1-\epsilon)w_B + \epsilon\left(w_A - \delta i\frac{p}{n}\right) - \delta\epsilon k - \delta^2\left(1-(1-\epsilon)^n\right)(n-i)\frac{k}{n} \end{aligned} \tag{4.3}$$

Here $(1-(1-\epsilon)^n)$ is the probability that there is at least one punisher in the first round. so

$$w_{\sigma_p}(i) > w_{\sigma_n}(i) \Leftrightarrow \frac{i}{n} > \frac{\epsilon k}{\left(1-(1-\epsilon)^n\right)\delta(p-k)} \tag{4.4}$$

Which leads to the result

Proposition 11. *If $S = \{\sigma_p, \sigma_n\}$, then σ_p is an ESS if*

$$\left(1-(1-\epsilon)^n\right)\delta(p-k) > \epsilon k \tag{4.5}$$

Proof. This derives straight from the equations 4.4. □

Of course we might continue this to investigate the strategy which plays 'third order defection,' but 'fourth order punishment' will keeps this in check and so on. And so the we eventually reach the strategy that punishes indefinitely. Any deviation from this will be punished.

If this is a cultural trait then we could stop here, assuming that it is adopted globally as a social contract. However, there are some who argue that this sort of punishment is genetically determined, and therefore needs to invade from rarity like other genetic traits. One solution to this problem is given by Boyd et al. [2010], who argues that if the action A hurt the individual at a personal level, but helps the group as a whole, then it is possible that between-group competition can support invasion. The idea is that individuals play repeated games within the group, and the resident defector population will respond to the 'threat of punishment' after the first round. So on the first round they defect, suffer the cost of being punished, and then in subsequent rounds they cooperate.

However, I do not believe that such a thing is needed, since I see little reason to assert that humans are biologically evolved to punish in this way. For example, suppose that we know that one of our colleagues cheats on a tax return. I know of few cases of individuals being physically assaulted by their peers for this. Yet this is what a biologically evolved theory would predict. Clearly such a system could exist, just that it is clearly not universal. What is more, as I found in my own collaboration, if we allow for the evolution of all such punishing strategies, then we find that 'anti-social punishment' is more likely to evolve [Powers et al., 2012].

4.5 Indirect Reciprocity and the Mutual Aid Game

So far we have seen three models of punishment in repeated games. The first is reciprocal altruism in section 2.4.2, which assumes that pro-social behaviour is held in place through individual threats of exclusion. The next is exclusion from economic goods in section 3.4.7, which relies on the institution of property rights being established by a population. The last is costly punishment described in the previous section, which relies on players repeatedly punishing those who do not conform.

Another model of this kind is that of indirect reciprocity, which was put forwards by philosopher Alexander [1987] and formalised in theories of Boyd and Richerson [1989] and Nowak and Sigmund [1998]. The idea of indirect reciprocity is simple — rather than individuals engaging in directly reciprocal acts, they instead form a social arrangement whereby they help one another in times of need. The idea is that individual A helps individual B not because they expect individual B to give return a favour from them directly, but because individual C might do them a favour, and so on.

There are several theoretical ways that we might represent indirect reciprocity formally, but I believe that the simplest is the theory of the *mutual aid game*. In each round of this game, an individual from a population of n individuals is selected to have a 'disaster.' Each individual in the population can then choose to help that individual at a cost of c to themselves to give a benefit of $\frac{b}{n}$ to the player who has the bad round. Suppose that each individual has a *reputation* that is either *good* or *bad*. Suppose that an individual gets a good reputation when they help another good individual, and a bad reputation every time they help another individual with a bad reputation, or if they fail to help another individual who has a good reputation. Suppose that we have the type *cooperator* σ_c who cooperates with anyone with a good reputation and defects against anyone else and *defector* σ_d, who always defects. Also assume that the game repeats with probability δ. If $w_\sigma(i)$ is the payoff of an individual who plays strategy σ in a group of i cooperators, then the payoffs for the two types are

$$\begin{aligned} w_{\sigma_c}(n) &= \frac{1}{1-\delta}\frac{b-cn}{n} \\ w_{\sigma_d}(n) &= 0 \end{aligned} \tag{4.6}$$

Proposition 12. *When $S = \{\sigma_c, \sigma_d\}$ then σ_c is an ESS when $b > cn$.*

How can this model be linked to punishing 'bad' behaviour? The answer is related to the work of Panchanathan and Boyd [2004]. In this model not only does refusing to help an individual in times of need give you a bad reputation, but so does performing some otherwise anti-social behaviour. Suppose that before the game is played, there is a round where each player must choose whether to cooperate in a public goods game. Doing so costs C. Not providing a public good gains you a bad reputation in the subsequent mutual aid game. Now suppose that σ_C cooperates in the the public goods game and plays like σ_c afterwards.

Proposition 13. *When $S = \{\sigma_C, \sigma_d\}$ then σ_c is an ESS when*

$$b > cn + (1-\delta)nC \tag{4.7}$$

and σ_d is always an ESS.

Proof. To show that σ_d is an ESS, notice that

$$\begin{aligned} w_{\sigma_C}(0) &= -C - c \\ w_{\sigma_d}(0) &= 0 \end{aligned} \tag{4.8}$$

from which it is clear that $\sigma_d > \sigma_c$. To show that σ_C is also an ESS under the conditions stated in the proposition, notice that

$$\begin{aligned} w_{\sigma_C}(n) &= -C + \frac{1}{1-\delta}\frac{b-nc}{n} \\ w_{\sigma_d}(n) &= 0 \end{aligned} \tag{4.9}$$

so $w_{\sigma_C}(n) > w_{\sigma_d}(n)$ when the conditions in the proposition are satisfied. □

Notice what these two models of punishment have in common — they both demonstrate that not only can the threat of future punishment incentivise a certain behaviour, it can also incentivise punishment itself. That is, if those who do not take part in the punishment are punished, then it can be better to just take part. In fact, Binmore [1998] points out that it is possible for strategies to evolve where the perpetrator punishes *themselves*, if the threat of punishment from the other player for not doing so is worse.

4.6 Crime and Punishment

Currently our model of the emergence of food-sharing does not require any additional punishment. The social contract guarantees property, and there is an economic incentive to claim it. However, I previously assumed that the food that is caught is automatically considered the property of a specific set of individuals within the group. However, in real life, the food does not become property until it is given out. This raises the question of how is the good defended until this time.

Suppose that the carcass of a killed animal is brought to the group for the first time, and an individual who did not take part in the hunt tries to take some. How does everyone else react? The carcass of the animal might be viewed as a common pool resource. That is, there is a valuable resource in the environment that it is not in any single individuals interest to unilaterally defend. So how is it defended?

Olson and Olson [2009] argue that it is possible to get individuals to provide a public good, which is another name for a common pool resource, if individuals are sufficiently incentivesed to do so. So what is the incentive? I argue that the incentive is rights over the meat itself. But does this solution assume that individuals can already defend it? i.e. Doesn't this solution already assume that the solution exists?

The two theoretical models above demonstrate why it is alright to assume the existence of the solution as part of the solution. Although the logic of this might seem confusing, first consider an unrelated example that uses the same logic. Suppose that there is a king and ten lords. The king has no army, but all of the lords do. The lords all believe that the other lords will do the what the kings says. The king tells the lords to go to war for him, and that any lord that does not will subsequently be attacked by the other lords, and any lord that does not take part in that attack will be attacked by all other remaining lords and so on. With these assumptions, each lord performs a cost/benefit analysis and calculates that it is better for them to go to war, because if they do not, then they will be attacked. But the king has no army, so why does everyone do what he says? The answer is because all the other individuals will do what he says, and they know that they will be attacked if they do not. In order for the solution to work, it has to be assumed in the model itself.

So back to the description of the common pool resource of food. Suppose that all of the individuals guard the food together as it is being divided up, and that any individual who does not guard it is excluded from the good, meaning that the rest of the population will stop him from accessing it. It should be clear that this uses this same logic — if everyone follows the rule, then each individual does a cost/benefit analysis and sees that it is better for them to defend the good, and for them not to steal it.

4.6.1 The Model

Recall the theory of conditional property rights introduced in the section 3.4.7. In this game, we assumed that pairs of individuals meet and contest over resources that both value that was represented by the following payoff matrix

	Hawk	Dove
Hawk	$(v-c)/2$	v
Dove	0	$v/2$

This model assumes that each individual is equally able to fight. Recall that it assumes that one individual wins the fight and gains a resource of value v, and that the other gets injured and loses the fight and pays a cost $-c$.

Picking up this model, we want to know how the strategy of hunting and sharing can invade from rarity. I partition this problem into two sub-problems. The first is the problem of how the good is guarded when it is the property of the group; the second is how it was guarded when it becomes the property of the individual.

Dealing with the first problem, suppose that there is a native population who do not play according to the social contract. The individuals in this population play by an individual strategy, hunting what they can and sharing only within their families. Now suppose that a group of cooperative hunters bring down some large game. The first question is, how do these hunters stop the native population from simply walking over and taking the good? How is it defended?

This problem might aptly be described as a common resource pool problem. Each individual within the group would like the good to be defended, but if doing so requires them to take a risk doing so, then they would like every other individual to do this. To formalise this idea, suppose that through the process of hunting, the population has created a good of size G, that they plan to divide n ways. Also assume that each individual who steals the resource also steals t of the good, so that if the thief steals then there is $(G-t)/n = G/n - t/n$, so that each theft takes t/i away from each individual. Recall that in the previous model I assume that once the goods had been distributed, they became the private property of the individuals and therefore are subject to the normal hawk-dove game. We retain this assumption for the current model, and so the question is how can the goods be defended.

The 'thief' in this model does not represent an actual player in the game, but rather an abstraction of them. It's purpose is to provide a reason for individuals to defend the good once it has been caught. Although I choose to represent it as a single event, it could represent many attempts to steal.

The solution that I propose is that the social contract is designed in such a way that each individual has two responsibilities. The first is to hunt, the second is to help defend the goods from theft. Failure to perform either of these responsibilities will result in the loss of property rights when the goods are distributed out.

We also need a model to represent the strategy for thievery. For simplicity, I assume that after the group good has been created, on the first round, there an individual from the population attempts to take some of the food. I assume that the thief manages to steal the food with probability x, and with probability $1-x$ he gets injured and suffers a cost c. In subsequent rounds, I assume that the thieves perform a basic cost/benefit analysis based on the previous round, meaning that I assume if $xG/n - (1-x)c < 0$, then the thief will attempt to steal with probability ϵ, which is small, and that if $xG/n - (1-x)c < 0$ then they will always steal in the following round. With these assumptions, the model proceeds as follows :

1. Form a population of with a proportion p of the population playing the native strategy σ.

2. This population is divided into groups of size $n+1$.

3. For invading types, we assume that there is positive assortment, so the probability of an additional individual of type σ being in a group where the focal individual is of type σ is $r + (1-r)p$.

4. Individuals choose whether to hunt collectively or individually. All those individuals who hunt collectively produce a resource of value G.

5. A thief will attempt to steal a portion of the good with probability 1 on the first round and probability 1 in subsequent rounds if $xt - (1-x)c \leq 0$ in the previous theft, and with probability ϵ if $xt - (1-x)c > 0$ during the previous theft.

6. If the thief attempts to steal, all individuals now get the choice of whether or not to defend the good. Defending costs each player d, and the thief gets a chance

$x(i)$ of stealing a portion of the good where x is assumed to be a function of i, where i is the number of defenders.

7. The property rights over all of the goods are now distributed to all players who both hunted and defended during the previous round.

For simplicity in this model, assume that once the good is the private property of the individual that it becomes subject to the standard hawk-dove game where the owner is the individual who holds the good. This is a reasonable assumption to make about the ancestral population — the native strategy must understand basic property rights, since if food could be taken out of the hands of the hunters, then it is unlikely that humans would have been a social species at all. Therefore we do not model this explicitly.

Assume that there is a native strategy σ_n that does not take part in the cooperative hunting or defending, but plays an individual hunting strategy. Also assume that there is the following invading cooperative strategy σ_c which plays *hunt, defend the goods, and then defend the individual property rights.* When placed in a group containing i other individuals that plays σ_c, this strategy gains a payoff of

$$w_{\sigma_c} = \begin{cases} D + \frac{\delta}{1-\delta} H & : x(i)t - (1-x(i))c < 0 \\ D & : x(i)t - (1-x(i))c > 0 \end{cases} \tag{4.10}$$

where $D = \frac{G}{i} - h - \left(d + \frac{x(i)t}{i}\right)$ and $H = \left(\frac{G}{i} - h - \epsilon\left(d + \frac{x(i)t}{i}\right)\right)$. Here the first $-h - d + \frac{G}{i} - \frac{x(i)t}{i}$ part of the sum represents the outcome of the first round, and the $\frac{\delta}{1-\delta}\left(-h - \epsilon d + \frac{G}{i} - \epsilon\frac{x(i)t}{i}\right)$ term represents the payoff for the subsequent rounds. Here h is assumed to be the marginal cost of taking part in collective hunting. I also assume that the payoff of the individual who hunts individually is assumed to be normalised to 0

The purpose in this model of assuming the change in the behaviour of the thief is that if we assume that an individuals attempt to steal every round, then the cost of defending is likely to be higher than the benefit. However, I ague that humans ought to be sensitive to fights where there is a significant difference between the strength of the contestants. Hammerstein [1981] showed that in dyadic interactions, when x, the probability that the focal individual would win, is greater that $\frac{c}{v+c}$, then the other player ought to play dove at equilibrium. In our case, where there are multiple individuals fighting and v/c is close to 1, then it is likely that we only require a few individuals defending for any potential thief to realise that they are too weak.

Assuming the standard model for positive assortment so that when $p \approx 0$, then the probability that there are j individuals of type σ is a group is equal to

$$M(\sigma|j) = \binom{n}{j} r^j (1-r)^{n-j} \tag{4.11}$$

We can therefore calculate the payoff for playing σ_c as being

$$\begin{aligned} W_{\sigma_c} &= \sum_{j=0}^{n} M(\sigma|j) w_{\sigma_c}(j) \\ &= nrD + \sum_{j=k}^{n} r^j (1-r)^{n-j} \frac{\delta}{1-\delta} H \end{aligned} \tag{4.12}$$

where k is the smallest integer such that $x(i)t - (1 - x(i))c < 0$

Proposition 14. *If $S = \{\sigma_c, \sigma_n\}$, then σ_c will invade from rarity when*

$$nrD + \sum_{j=k}^{n} r^j (1-r)^{n-j} \frac{\delta}{1-\delta} H > 0 \tag{4.13}$$

Proof. The average payoff of the σ_n strategy is equal to $W_{\sigma_n} = 0$, so the σ_c strategy will invade when $W_{\sigma_c} > 0 = W_{\sigma_n}$, which occurs when the conditions of 4.13 hold. □

This condition has a relatively simply interpretation. The first term nrD represents the cost that every σ_c in the first round has to suffer. Recall that D is equal to the value of hunting minus the cost of definitely having to defend. The second $\sum_{j=k}^{n-1} r^j (1-r)^{n-j} \frac{\delta}{1-\delta} H$ term is the probability that a σ_c will end up in a group with a sufficient number such that they can defend the good ($\sum_{j=k}^{n} r^j (1-r)^{n-1-j}$), multiplied that the payoff that they receive for repeatedly hunting ($\frac{\delta}{1-\delta} H$).

Remember that the native strategy gaining a payoff of 0 does not mean that they all die, it is just the base against which everything else is measured. We could add a base fitness of w_0 to both strategies, but this would in no way alter the dynamics.

The next question is, is the equilibrium stable against individuals who hunt but do not defend. Call this strategy σ_h. This is relatively easy to demonstrate. Suppose that the population is at the ESS where all players are playing σ_c. The payoff for an individual the two strategies are

$$\begin{aligned} W_c &= D + \frac{\delta}{1-\delta} H \\ W_h &= -h \end{aligned} \tag{4.14}$$

Recall that all individuals must defend on the first round, and so the strategy σ_h that does not defend must pay the cost of hunting, but does not receive property rights over any of the food. Following that, that individual is excluded from all food sharing.

Proposition 15. *When $S = \{\sigma_c, \sigma_h\}$, then σ_c is and ESS if $D + \frac{\delta}{1-\delta} H > -h$*

The proof follows directly from equations 4.14.

To recap, the model above argues that, from a population that plays an individual hunting strategy, a type may evolve that hunts and shares cooperatively. The manner that this strategy ensures that individual members receive a benefit is by providing property rights over the catch, meaning that no individual has the benefit from deviating by trying to take more, since in doing so, they will clash with other members of the group. However, in order for this type to be successful, it must be able to defend the catch from individuals who were not part of the hunt. This creates a classic common resource pool problem — every individual would like the good to be defended, but they would rather not do it themselves. The solution to this common resource pool problem is that individuals who do not help defending the good will not receive property rights over the goods, and will not be allowed to attend subsequent hunts. This provides sufficient incentive to hunt and defend, and also shows how the strategy can invade from rarity.

4.6.2 Mutual Defence of Individual Property Rights

So far I have assumed that clashes between individuals holding property rights are strictly dyadic. However, when the goods are the property of the group, I assumed that in defending them, individuals act concertedly. Does this possibility of having more than two individuals in a game at the same time make it harder for individuals to defend their property?

I argue that humans have learnt to form networks of mutual responsibility, so that other individuals within the community will help in the defence of property. What motivates this defence? The promise of defence in return. It is relatively easy to show that this model is very similar to the well known repeated prisoner's dilemma [Axelrod and Hamilton, 1981]. I sketch the model here. Assume that individuals are paired together to play a series of games that repeated with probability δ. In each game, one player is selected to be the *victim* and the other the *neighbour*. Assume that the victim's property is under threat and they must defend it. If their neighbour does not help them, they lose the property and gain a payoff of 0. If their neighbour helps them, they are strong enough to fight off the attacker and gain a payoff of $v - d$ where v is the value of the resource and d is the cost of fighting. As for the neighbour, they can choose to *help* or to *defect*. Helping costs d, the cost of fighting, defecting costs nothing.

Assume that there are two strategies, *defector* σ_d, which always defects, and *reciprocator* σ_r, which defends as the neighbour, so long as the other player defended the last time they were the neighbour. It is easy to see that the expected payoffs are as follows:

	Reciprocator	Defector
Reciprocator	$\frac{1}{1-\delta}\frac{v-3d}{2}$	$-\frac{2d}{2-\delta}$
Defector	$\frac{2(v-d)}{2-\delta}$	0

When the defector plays a reciprocator, then there is a half chance that they will be the neighbour first, meaning that they get a lifetime payoff of 0, and half a chance that they will be the victim, meaning that they get an immediate payoff of $v - d$. If we assume that their lifetime payoff is equal to S, then S is equal to the following

$$S = \frac{1}{2}\left((v-d) + \delta S\right) \tag{4.15}$$

That is, half a chance that they were victim gives them a payoff of $v - d$, plus the same expected payoff in the next round, multiplied by the chance that there is a next round. Likewise, for a recirocator playing a defector, the payoff is

$$S = \frac{1}{2}\left(-d + \delta S\right) \tag{4.16}$$

If we assume that $\frac{1}{1-\delta}\frac{v-3d}{2} > \frac{v-d}{2-\delta} > 0 > \frac{v-d}{2-\delta}$, meaning that pair of reciprocal strategies that meet gain the highest long-term payoff, then we get the following result

Proposition 16. *When $S = \{\sigma_r, \sigma_d\}$, then σ_d is an ESS, and σ_r is an ESS when*

$$v > d\left(\frac{4}{\delta} - 1\right) \tag{4.17}$$

Proof. For the first part, notice that

$$w_{\sigma_d}(\sigma_d) = 0 > \frac{-d}{2-\delta} = w_{\sigma_c}(\sigma_d) \tag{4.18}$$

For the second part, we see that

$$w_{\sigma_c}(\sigma_c) = \frac{1}{1-\delta}\frac{v-3d}{2} > \frac{v-d}{2-\delta} = w_{\sigma_d}(\sigma_c) \tag{4.19}$$

which occurs when the conditions in the proposition hold. □

The purpose of this simple model was to address the possible problem that the formation of coalitions could be dangerous to social contracts. I argue that in order to stop this problem, individuals learn to form networks of mutual aid. In this simple model, I assume that these networks contains exactly one other individual. However, in real life I expect that each individual would form many such arrangements.

4.6.3 Punishment and Politics

The model introduced at the beginning of this section shows that it is possible for a social contract to invade that links both hunting and communal defence of property rights to obtaining property rights. In this section, I show how this same mechanism of connecting property rights to actions can be used in order for communities to police other behaviours that it considers anti-social.

Suppose that each individual randomly meets another player from their community and can take one of two roles — A or B. A has the choice whether to act socially or anti-socially, gaining a payoff of w_S and w_A for doing so. Following this game, the players then play a game of hawk-dove over some of their property with that player of value v. Consider the following strategy *finer* σ_f, which plays *when A play socially; and then play hawk over the property, when B, play dove if A played socially, otherwise play hawk*

Proposition 17. σ_f *is an ESS when*

$$w_A > w_S - v \tag{4.20}$$

Proof. There are several different strategies to consider. The strategy that plays anti-socially gains a payoff of w_A, but then either plays dove to get a payoff of $w_A - v$, or hawk to gain a payoff of $w_A + \frac{v-c}{2}$ when they are A, which is worse that w_S according to the assumption in the proposition. The strategy that tries to steal when they are either A or B and not the owner gets a payoff of $w_x + \frac{v-c}{2}$, for $x = \{A, S\}$, which is worse than playing dove in either situation. □

What does this model show? The original idea of the social contract was to connect performing an action, originally defending a public good, with gaining property rights. Here I argue that this can be made more general by considering any behaviour that is considered to be anti-social. In this example, the community polices itself by offering a disincentive to behave anti-socially — the loss of property. In this model, the lost property becomes the property of the victim of the behaviour, although this is not strictly speaking necessary. The point is that this is a form of punishment that does not rely on humans having a genetically evolved adaptation to enact corporal punishment against the perpetrator of a crime that does not involve them. Rather, it shows how a community can negotiate a system of punishment that creates a disincentive to act anti-socially, but does not require any higher-order punishment. That is, the victim of the anti-social behaviour does not act out of fear of being themselves punished, but through their own self-interest.

4.7 Empirical Validation

As with all of the other chapters of this dissertation, here I apply these models to help understand the mechanism underpinning empirically established examples of punishment. Unlike the food-sharing example, the theory of costly punishment arose in the theoretical and not empirical literature, and so it is difficult to find a natural example. So my strategy here is to apply different parts to the well studied example of law and punishment in the the Nuer [Evans-Pritchard, 1940]. What I hope to do with this example is demonstrate firstly that the theories of costly punishment (reviewed in section 4.4.1) and exclusion from mutual aid (reviewed in section 4.5) do not apply, and secondly that my theory of mutual defence of property rights in section 4.6.2 and of prohibition in section 4.6.3 do apply.

It may be questionable to use the Nuer, who are predominantly pastoralists, to investigate behaviours that has deep evolutionary roots. We could, for example, use a modern corporation in London. But I believe that this line of thinking somewhat misses the point. Firstly, there are differences between pastoralists and modern economic behaviour that are not apparent between hunter-gatherers and pastoralists. The difference is that in hunter-gatherers and pastoralists, there are no exogenous forces acting upon these communities. In modern day corporations, we cannot ignore the force of the state that maintains certain behaviours. Another reason that it is worth investigating the social contracts of pastoralists is that it seems likely that their social contracts will be similar to the

The Nuer are a Nilotic ethnic group, mainly concentrated in Southern Sudan and southwestern Ethiopia. They are predominantly pastoralists, with the single most important economic good among being cattle. Cattle provide meat and milk, a means of trade, a means of paying reparations and settling disputes and a means of young men obtaining wives through the payment of bride-price [Evans-Pritchard, 1940].

Among the Nuer, order is maintained largely through a formalised system of fines. When a crime occurs, the victim of the crime is allowed to exact compensation along the lines of the social contract. These fines protect individual rights by creating a series of disincentives that make anti-social behaviour net costly. The most important tribal institution that protects the Nuer against violence is the *blood feud*, which is invoked in the case of murder. When it is invoked, the perpetrator must pay a fine — forty or fifty cattle in total — to the friends and family of the victim. Among the Nuer, this is a small fortune, and it is unlikely that all of the fine will be paid all at once, and so the debt may continue for many years.

Other types disputes are also common among the Nuer. If not resolved, these minor disputes may escalate into full blown blood-feuds. If a man commits adultery, he must pay indemnity of five cows and an ox. Fornication with an unmarried girl is compensated by payment, though they may turn a blind eye if the man has cattle and intends to marry her. If he does not possess cows, or she is already engaged and the girls kinsmen find out — they will go to his group and take a male and a female calf as punishment. Ten cattle are paid for a broken leg, skull, or the loss of an eye, and two for a girl's broken teeth.

There are several aspects of the Nuer social contract that is relevant to us here. Firstly, like in hunter-gatherers, the process of obtaining public opinion is vital. If you are 'right,' according to the social contract, then your kinsmen will support you, if you are wrong then they will not. Secondly, what counts as a transgression is culturally determined — there is no 'natural' moral law here. Thirdly, the type of punishment is also culturally determined. Individuals do not unilaterally determine how much they will punish, but rather, they adhere to the commonly held group norms. Lastly, there is an incentive for the punisher to see the punishment through, and likewise, there is an incentive for the punished to accept their punishment. In this case, the incentive for the punisher is that they personally gain cattle, and the incentive for the punished to accept this is that acting otherwise would likely lead to a fight.

Are property rights maintained through public opinion? There is a great deal of evidence that the Nuer aid one another. Evan-Pritchard observes that:

> The only quarrels within a village or camp about ownership of cattle that I have witnessed have concerned obligations of kinship or affinity and have eventually been settled by one party giving way on account of his relationship to the other. If a man seizes cattle from a kinsman or neighbour he walks into his kraal and takes them. If the owner has a strong case he may resist : otherwise he lets the cattle go, for he knows that the man will be supported by public opinion in the community. [p.165 Evans-Pritchard, 1940]

also

> The Nuer have a keen sense of personal dignity and rights. The notion of right, *cuong*, is strong. It is recognized that a man ought to obtain redress for certain wrongs. This is not a contradiction of the statement that threat of violence is the main sanction for payment of compensation, but is in accord with it, for a man's kinsmen will only support him if he is right. It is doubtless true that if a man is weak it is unlikely that his being in the right will enable him to obtain satisfaction, but if he is in the right he will have the support of his kin and his opponent will not, and to resort to violence or to meet it the support of one's kin and the approval of one's community are necessary. One may say that if a man has right on his side and, in virtue of that, the support of his kinsmen and they are prepared to use force, he stands a good chance of obtaining what is due to him, so long as the parties to the dispute live near one another. [p.165 Evans-Pritchard, 1940]

There is little more than I can say about this. The system that maintains it is mutual obligation, which presumably is maintained through repeated interactions (it is obvious that δ, the probability that individuals will be around another day is relatively high among the Nuer).

One interesting point that is not covered in the model is that Nuer law is always relative to the strength of the lineage of the injured party. Individuals from weaker lineages will find it difficult to exact compensation from a stronger lineage. Between lineages of equal size, disputes are settled by reference to the social contract, and an individual can rely on support from their own lineage when they exact their compensation. This means that, under ordinary circumstances, an individaul's lineage would help defend cattle, but in the context of compensation, the lineage of the taker would help them if they were strong enough.

This opens up the possibility that although Nuer law is theoretically equal, in practice it may not be. If this is true, then we should find that individuals from stronger lineages abuse this power by committing a higher frequency of transgressions. I cannot find any evidence of whether or not this is true.

Is Nuer property and fines determined by the social contract? (is there a clear signal sent to each individual in the dispute?) In all models up to now, I assume a perfectly working signal is sent to all players in the game, and that this signal clearly identifies one of the players as being 'right' and the other as being 'wrong.' There are many important assumptions here that are worth exploring in this context. What this model assumes is that there is a mutually accessible set of phenomena that members of the community can witness. This implies that the community has agreed on 'categories' of different actions, that are all of the same type. In a particular dispute, an individual is able to call on the mutually accessible event to invoke the rule. Why wouldn't an individual just ignore the rule? Because they do worse by fighting (at least, that is what we assume). Invoking the mutually accessible event implies an action in the game. This is the moral that is attached to it.

The question is, in practice, are the moral rules always as clear as the model assumes?

> When a Nuer speaks of a person having stolen (kwal) an animal, he means that he has taken it without permission and by stealth and by no means implies that he ought not to have taken it. Within a tribe an abductor of cattle always considers that he is taking what is owing to him. It is a debt (ngwal) which he is settling in this way, because the man who owes him cattle has not repaid them of his own accord. The legal issue, therefore, is whether he is right in assuming a debt and whether he should have taken the particular animals he took. So well is this practice of helping yourself

> to what is owing to you an established habit, that it may be said to be a customary way of settling debts. Thus the final cattle of a homicide-payment are often seized in the pastures, and it often happens that when a bridegroom and his kin do not hand over all the cattle they have promised, the wife's brothers try to seize the animals still owing to them. [p.172 Evans-Pritchard, 1940].

So it seems that there is indeed such a system in place. However, the system does not always work so smoothly:

> When we speak of a man being in the right we do not suggest that disputes are mostly a clear issue between right and wrong. Indeed, it would be correct to say that, usually, both parties are to some extent right and that the only question which arises is. Who has the greater right? To state the matter in a different way: a Nuer dispute is usually a balance of wrongs, for a man does not, except in sexual matters, wantonly commit an act of aggression. He does not steal a man's cow, club him, or withhold his bride-cattle in divorce, unless he has some score to settle. Consequently it is very rare for a man to deny the damage he has caused. He seeks to justify it, so that a settlement is an adjustment between rival claims. I have been told by an officer with wide experience of Africans that Nuer defendants are remarkable in that they very seldom lie in cases brought before Government tribunals. They have no need to, since they are only anxious to justify the damage they have caused by showing that it is retaliation for damage the plaintiff has inflicted earlier. [pp.171-172 Evans-Pritchard, 1940]

Here is where the model begins to break down. I have assumed that in a particular dispute, there is always an obvious individual who is right and wrong. But from this passage we see that any individual dispute is likely to be a question of the 'greater right,' implying some more complex system of hierarchy of right.

Presumably the Nuer have a good idea about this hierarchy of values, and may be able to express them. However, there are always likely to arise situations 'on the border' — situations where there is no moral precedent to invoke. How are such situations resolved? I imagine that they are done by some process of negotiation, and applying to other abstract principles in an attempt to justify action. Indeed, this is what Evans-Pritchard describes when they face tribunals. However, this is not entirely clear, and so we would benefit from further research.

As a brief aside, here I argue is an excellent demonstration of the value of using precise modelling to describe a situation. It allows us to state plainly what we do, and what we do not know about how a social contract operates, and why individuals are motivated to follow it. In doing so, it also points out new avenues of investigation.

Do the threat of fines keep people from behaving immorally? (is $w_A > w_S - v$ in section 4.6.3?) As with all of these investigations, trying to calculate the lifetime fitness costs and benefits of a single act is likely to be extremely difficult. However, we do know that cows are the single most important economic item among the Nuer, being the primary source of nutrition, as well as settling debts, and paying dowry and bridewealth. We also know that for blood feuds, the number of cows that are to be paid are so many that they must be paid back over a number of years. It seems likely, then, that these are significant costs.

Of course, it may well be that when these laws are created, they are originally intended to represent a sufficient cost, but end up being insufficient due to miscalculation. If this occurs, then our model would predict that, over time, such laws would be undermined. Indeed, this may well be the case with out of marriage fornication. The relative costs and benefits of this act may well be highly dependent on the individual circumstances of the individuals involved. For example, if a male had very little chance to find a marriage partner, then the benefits of fornication might be high. But if the Nuer are to have a fixed rule so that, at least in theory, applies to all members of the community, then it may well be that no matter how high the cost is set, some individuals will still find it beneficial. In these situations, the assumptions of the model would need to be revised, because not every individual within the community is facing the same cost/benefit question. However, I suggest that it is possible that the rules may have been revised throughout the history of the Nuer, so that they are set high enough that they dissuade what they consider to be anti-social behaviour a sufficient amount of time, without making the prospects of an individual so harsh that they do better by leaving altogether.

While these assessments are somewhat cursory, I hope that they suffice to show how game theory can be used to help anthropologists ask more precise questions when trying to understand the forces at that maintain social structures. Ultimately, this project relies on a dialogue between theoreticians and empiricists. In this dissertation, the dialogue is one way — I read the ethnographic literature and try to discern the essence of how the ethnographer believes a particular social system is organised. From this, I attempt to discern the sorts of biological strategies that these individuals are playing. Ideally, the assumptions of the model that I have described can be investigated by anthropologists and, where incorrect, can be corrected.

4.8 Discussion

The purpose of this chapter is to discuss models of punishment in human societies. The first model describes what I believe to be the solution to the first public goods problem that humans faced — how to defend large catches. My solution is that humans devise social contracts that not only connect hunting to property rights, but that also connect defending the meat to individual property rights. The second model discusses cooperative behaviour between dyadic pairs of individuals, which might exist within the wider context of group wide sharing. The third model demonstrates how the basic model of exclusion can be used in order to prohibit anti-social behaviour.

The take home message of this chapter is that humans have devised many mechanisms to punish anti-social behaviour. My thesis is that to understand how such punishment operates, we have to observe the mechanism in action and try to ascertain how that particular contract works. Models of 'universal punishment,' such as costly punishment, are demonstrably incorrect, since there are so many examples of punishment that are not maintained by the direct threat of violence. Like tolerated theft, I believe that this literate has identified one of an infinite set of equilibria and argued that all punishment be interpreted along those lines. For example, in economic games, evidence that individuals are willing to sacrifice resources to punish another individual is taken for evidence of costly punishment, rather than evidence for any of the other types of punishment that it might be.

4.9 Conclusion

My purpose in this chapter is to extend the theories of punishment. Most theoretical models of punishment either assume that individuals pay an individual cost to lower the fitness of others, or are based on some kind of reputational model.

In this chapter I take a different approach. By examining the ethnographic literature on cooperation, I find that there are many ways that small-scale societies enact punishment. In the particular example of the Nuer, punishment is enacted by the victim, who does not have to pay a cost for punishment, but actually individually benefits, because they get to keep the fine.

I argue that the assumptions of direct punishment and second order punishment (which together make the theory of costly punishment) have come from the modelling community, and generally seems bear little resemblence to the types of punishment that we generally observe in societies.

In this chapter I investigate three moral dilemmas that I believe are common, and demonstrated three possible solutions to these. The first is the scenario of a commonly owned resource, such as a large animal in a hunt. The question is, how can this be defended if that defence is a common pool resource. My answer is that it is only a common pool resource some of the time, and basic economic incentives can be explain how it is defended. Tthe model assumes that all individuals must take part in both actions. Strictly speaking, this is not necessary. It is quite possible for human groups to divide our responsibilities to have some individuals specialising in hunting, and others specialising in defending. This model also explains how such an adaptation can inavde from rarity.

The second model addresses the question of why, if individuals have to defend their private property alone, do groups of others not get together and steal their goods. The answer is that, within communities, individuals form reciprocal networks where they aid one another in defence of their rights. In this sense, the strength of a coalition is not simply the number of friends an individual has, but also what is the nature of the dispute. This point inspired me to include Lincoln's famous aphorism that 'might makes right' in the introduction of this section.

In the third model, I show how this idea of fining can be used to discourage a wide range of moral transgressions. To demonstrate the utility of this model, in the following section, I focus on how this applies in particular to a system of fines in place among the Nuer.

Hopefully my point has now been communicated plainly. Humans do not possess anything like a simple adaptation that makes us want to enact corporal punishment against those who we believe have morally transgressed. While we do *sometimes* enact such punishment, it is only one of the many types of punishment that we are capable of using. I focus on fines of economic goods, but more generally, in order to find what particular type of punishment has been utilised by a particular group, and how that group encourages all its members to behave in a certain way, I argue that we must go out into the world and observe the behaviours.

CHAPTER 5

BOND TESTING IN HUMANS AND OTHER SOCIAL SPECIES

It is not so much our friends' help that helps us, as the confidence of their help.

— Epicurus

This story shall the good man teach his son;
And Crispin Crispian shall ne'er go by,
From this day to the ending of the world,
But we in it shall be remember'd;
We few, we happy few, we band of brothers;
For he to-day that sheds his blood with me
Shall be my brother; be he ne'er so vile,
This day shall gentle his condition:
And gentlemen in England now a-bed
Shall think themselves accursed they were not here,
And hold their manhoods cheap whiles any speaks
That fought with us upon Saint Crispin's day.

— William Shakespeare, *Henry V*

5.1 Introduction

From chimpanzee grooming to play fighting in wolves, the biological world offers many examples of organisms that bond. Bonding is also an important part of the human experience. Although it feels natural enough for us to engage in such activities, from an evolutionary perspective, it is not so straight forwards to explain.

Bonding normally precedes some cooperative task. But the question is, why should two or more organisms who are about to engage in a cooperative activity first engage in the wasteful process of bonding? What exactly do these activities achieve? How is the cost recuperated? A popular explanation is that bonding is a mechanism used to develop 'trust' between two more conspecifics [De Waal, 1986, Smith and Bliege Bird, 2005, Zahavi and Zahavi, 1997]. But this opens up the question of 'what is trust and why is it required?'

Suppose that we observe a species of bird who pair-bond for the purposes of building a nest and raising young. For example, swans, bald eagles or turtle doves to name a few. In this hypothetical species, the male and the female must cooperate to produce enough food for their offspring, and to protect the nest from predators. However, before they do this, suppose we first observe them engaging in some elaborate and costly bonding display. If the interests of the two birds were already aligned, as it might seem that from the description, then the act of bonding seems like a frivolous waste of valuable resources.

Of course, when we dig deeper into the evolutionary logic of social bonding, we see that the interests of the two birds are probably only partially aligned. The male might benefit from having multiple nests with multiple females partners. This would hurt the female because the male produces less resources for her offspring. The female might benefit from tricking the male into bringing her food and other resources, and then building a nest with another male. As well as being aligned, the interests of the two sexes also conflict.

Social bonding is not just for the purposes of pair-bonding and nest building. Social mammals like humans and chimpanzees live turbulent social lives, and very often the biggest threat to a person comes from conspecifics. To protect themselves, members of these species form complex webs of social obligation for mutual protection. But as with pair-bonding, this type of cooperation is also open to deception. Ideally, every individual within these populations would like others to support them, while they would prefer to risk as little as they can.

My hypothesis, along with Zahavi and Zahavi [1997], is that the purpose of social bonding is to ensure that your partner is invested in the relationship. The basic idea of this hypothesis is that before cooperation commences, you require your potential partner to pay a significant cost to bond. This forces them to create an 'honest' signal of their intention. Those who do not value the relationship would be unwilling to pay the cost of the signal, therefore the cost of the signal must outweigh the benefits that your partner would gain if they cheated you. So although it is costly, the purpose of bonding is to *reduce* the overall costs of being cheated. We therefore expect to see social bonding utilised when the likelihood of finding an honest partner is low, and the cost of being cheated is high.

The actual act of bonding is normally two or more organisms engaging in some sort of *symbolic* activity. Although the word 'symbolic' might seem out of place when discussing animal behaviour, I simply use it as a metaphor. A symbol is something that is used to represent another thing that is not directly connected to the symbol. For example, a word can symbolise an action, even though the action has nothing to do with the sound of the word. Bonding is exactly this. It is a series of choreographed exchanges used to represent and communicate the status of the relationship. The act of bonding need have nothing to do with the activity that follows, which is what gives it its symbolic status.

Humans are no exception to this rule. There are many ways in which humans are unique in our social organisation, but the diverse ways that we have found to bond are among the most exaggerated and unusual. To humans, the 'symbolic' part of the bonding is no mere metaphor, we are often fully aware of their meaning of the bonding exercise. A wedding is not simply a group of people in a room, but a symbolic union of two individuals (and often families). It can be very expensive to the families of the bride and groom, sometimes necessarily so. If we did not know about the cooperation that follows the marriage, the activities in the wedding would seem strange indeed. Only when we understand the activities that follow can we comprehend the logic behind the extravagance of the wedding.

5.1.1 My Contribution

In this chapter I introduce a formal model the bond testing hypothesis described by Zahavi's famous 'testing the bond hypothesis' above. This theory has been investigated before, but had never been formally modelled, and so my contribution is therefore to demonstrate that such a mechanism is evolutionarily stable, and that it can invade from rarity. I also find some limiting conditions that were not known before, which may alter how 'testing the bond' is viewed.

The fundamental idea behind this model of the bond-testing hypothesis is that there is a group of two or more individuals (two in my model) who play a game of *imperfect information* against one another. In a game of imperfect information, the players can be one of a number of different *types*. Each player knows their own type, but not the type of their partner.

In this model, I assume that there are two types of player, where in any given game, each type is chosen by a *random move by nature*. One type wants to cooperate with their partner, the other type wants to take advantage of them. The random move by nature is a mechanism in game theory to represent a game where have only partial information about the game. In this setup, I assume that they do not know the 'intention' of their partner, as represented by type.

It is important to understand that the 'type' is not inherited, but is particular to the situation. For example, my desire to partner with another individual is dependent on whether or not I already have a partner. But even if I do, I may still benefit from you cooperating with me. Therefore, I will try to trick you to gain an advantage. However, this situation is not genetically inherited. At one point, I do wish to find a partner. Therefore it is not a case of simply evolving to be a 'cooperator' or a 'defector,' since I can cooperate with one individual, but gain nothing from cooperating with another.

My hypothesis is that social bonding is a means to discover the type of your partner. To do this, each individual requires their partner to engage in some costly activity. If they are willing to pay the cost which only individuals who want to cooperate will be, then this indicates the type of partner, which in turn indicates that it is safe to cooperate. So for bonding to be evolutionarily successful, it should cost more than the value of cheating, thereby ensuring its honesty.

The nature of bonding is extremely important here, because if an individual were to engage in costly social bonding with every individual they met, then it is possible that they will pay costs that could outweigh the benefits. To reduce these costs, I assume that bonding occurs simultaneously, so that both players know as soon as the other player stops. This allows them to immediately disengage themselves from the bonding activity, therefore saving the extra cost of wasted bonding that they would otherwise have paid.

In this dissertation I am most concerned with the evolution of social contracts in humans, and so I choose to analyse two examples of social bonding that I believe are important to small-scale societies, and are therefore likely to have been important in our evolutionary history. The first is the example of *hxaro* (har-ro) in the !Kung [Wiessner, 2009] — a system of elaborate gift giving that is used to cement long-term alliances with distant neighbours. The second example is ritualistic warfare that is seen in many small-scale societies [Keeley, 1996], and in particular, I focus on the example of the 'great ceremonial war ' among the Enga [Wiessner et al., 1998]. The first example is relatively straight forword. Because of an unpredictable climate, the !Kung enter into long-term dyadic reciprocal relationships where they help one another in times of need. However, before such relationships begin, the !Kung enter into an extended period where they send one another gifts over the period of about a year. These gifts are never practical, and always ornamental. The second example is of a particular kind of warfare, commonly referred to as 'ritualistic' [Keeley, 1996]. In ritualistic warfare, very few individuals tend to be killed, territory is not taken or lost, and loot is not taken. My hypothesis is that this is potentially an elaborate example of social bonding.

In this section, I do the following:

1. Review the theoretical literature on social bonding.

2. Introduce a theoretical model of social bonding by combining evolutionary game theory with games of imperfect information.

3. Review the empirical literature on gift-giving in the !Kung and ceremonial warfare

among the Enga, and demonstrate how the model can be applied to understand these two examples of social bonding.

5.2 Social Bonding

Signalling is an important adaptation found in many species; from song and aggressive posturing in birds to chimpanzees stomping around the camp and beating their chests. From an evolutionary perspective, it is particularly interesting, because it seems that if an individual was able to influence the outcome of a social interaction by simply sending a signal, and it would be to their advantage to do so, then it raises the question of why other individuals would evolve to take any notice.

Many scholars have argued that in order for a signal to evolve, it must be *honest* [Dawkins and Krebs, 1978, Krebs et al., 1984, Smith and Harper, 2003, Zahavi and Zahavi, 1997]. For a signal to be honest, it must have the property that it accurately conveys salient information about the sender to the receiver. This accuracy will be undermined by the 'arms-race' that occurs between sender and receiver. The sender wants to manipulate the situation to their advantage, the receiver will evolve to resist such manipulations. Zahavi and Zahavi [1997] argue that a common way of achieving this honesty is by making the signal costly in some way. For example, if a signal involves jumping around vigorously prior to some social interaction, then the act of sending the signal — vigorous jumping — is costly.

Social bonding is a specific type of costly signal. Although its logic can be applied to a diverse range of behaviours, its unifying feature is that it consists of two or more individuals engaging in some costly activity, in order to convey some information about the status of the relationship. Therefore, it normally precedes social interactions where their is some risk of being manipulated by your partner, thereby creating the necessary disincentive for your partner to cheat you. In the particular case of social bonding, the information that is being conveyed is something about the status of the relationship, which in turn gives the receiver information about how much they can 'trust' the sender.

To highlight the fundamental logic of social bonding, consider the following slightly unusual example of the wagtails — small European songbirds — offered by Zahavi and Zahavi [1997]. There are two stages to the wagtail life cycle. First, through a series of contests, the male wagtails establish some territory for themselves. With their territories established, the males then wait for the females to arrive. It is important to realise that, ultimately, the males want the females to stay in their territory, making what happens next seem somewhat counter intuitive. When the females arrive, the males proceed to attack them and drive them away. After some time has passed, some of the females return to the territory, where the same occurs again — the males drive the females away. This process repeats until, eventually, there is only one female left in the territory. At this point, the male relents in his attacks, and allows this last female to remain there.

The question is, if the male ultimately wants the females to stay in their territory, then why do they spend so much time and energy attacking them? What is the purpose of this costly behaviour? Why not just allow them to stay in the first place? Before Zahavi, there were several evolutionary hypotheses put forwards to explain this behaviour [Zahavi and Zahavi, 1997]. One of these is that the males cannot distinguish between the females and other males who were trying to take the territory. But Zahavi argues that this explanation falls short, because if such was the case, then we would expect there to be pressure for males to develop more sophisticated methods of identifying member of the opposite sex, or equally, for the females to make their sex more plain.

Zahavi proposes an alternative evolutionary explanation — that the purpose of such behaviours is to test the 'intentions' of the females that arrive on the territory. The fundamental idea is that the intentions of the males and females are not necessarily perfectly aligned. The females can profit from being allowed on the territory, even if they do not intend to stay there. They can find a temporary place of residence while searching for a more permanent one, and they can fool multiple males into accepting them as mates and take the best resources from both territories. The males, on the other hand, have a great deal to lose by allowing a cheating female to stay. The female may consume the best resources on that patch, making it less desirable for another female, or simply her presence might mean that the male misses an opportunity to find a female who is actually interested.

Following this logic to its conclusion, this asymmetry in interests means that it can be adaptive for a male wagtail to test the intention of the female by physically attacking them. In the following section, I investigate how a similar logic underpins bonding in social primates.

5.2.1 Social Bonding in Humans and other Social Primates

The term 'social bonding' is not normally associated with the fighting mechanism of the wagtail discussed in the previous section. The term normally conjures up images of hugging, sharing, engaging in joint activities, grooming and so on. Indeed, these types of bonding are common in many species, particularly social mammals. The wagtail example serves to encourage us into separating our own subjective feeling of bonding from the evolutionary facts. If Zahavi's theory is right, then even though the phenotypic expressions are quite different, both human and wagtail social bonding are explained by the same evolutionary logic.

Compared to humans and other primates, the wagtails live a relatively simple social life. They test their partners, they settle in a territory, then they leave. Within the typical chimpanzee or human group, social life is constantly changing [De Waal, 1986, Boehm, 2009]. It may be that new members of the group arrive, new adults mature, allies are killed and so on. Any and all of these events can significantly upset the balance of social life, requiring individuals to act quickly to form new alliances.

Within chimpanzee colonies, acts of social bonding are often used to cement alliances. To take an example from De Waal [1982]'s chimpanzee politics, at one point, the most physically able but least politically savvy male was forced to share its dominant position with an older, smaller, but more intelligent male. The larger male was unable to establish female coalitional support — a vital instrument for the successful alpha. The smaller male, on the other hand, did have support from the females, but was unable to beat the larger male in a physical fight.

Complicating matters was the existence of a third male, who was able to both rally female support *and* was large enough to match either of the other males in combat. However, this third male could not face the other two males as a coalition, and was therefore held in last place by an alliance. During this period, the two males in the alliance spent a disproportionate amount of time together, sometimes grooming, sometimes playing, and sometime simply sitting near each other. DeWaal notices that when the older and third male — the one who was left out of the alliance — attempted to sit next to each other, the alpha male acted quickly to chase them apart.

The alpha seemingly recognised the importance of social bonding. Without it, the second and third males could not be sure of the intentions of the other. However, in this case, the two males eventually did manage to bond for a sufficient amount of time, and with the relationship cemented, they worked together to depose the alpha.

Cooperation and coalitions are also extremely important in human groups. As we have seen in sections 3 and 4, there is a great deal of evidence that humans rely heavily on the support from their social network for both economic reasons, and to defend one another against aggressors. Although I argue that these coalitions rely on institutional arrangements in the form of rules, they also require that individuals are actually able to form relationships — the institution by itself will not. The question here is, at any given time, how do you know exactly who will help you?

If networks of support are important to the social contract, then we expect humans to be adapted to being able to establish bonds of trust in order to ensure that those who we would help would be willing to do the same for us. Empirically, this seems to be a relatively accurate picture. Humans do not tend to engage in helping activities with strangers, at least not at a significant cost to ourselves, but are rather more willing to help our friends when they are in need, even risking death during violent episodes.

In the latter part of this chapter, I give some specific examples of human social bonding that I believe to be both interesting and important. These are a form of gift giving that is well documented among the !Kung, and a form of ritualistic warfare, of which I give a specific example among the Enga. These are but two of many different examples I could have chosen to highlight the behaviour — I fully expect that every human community will have their own methods for social bonding.

5.2.2 The Logic of Signalling

Before introducing the full model in the following section, I review the logic that underpins signalling. All models of signalling assume that individuals engage in some social activity, which here is represented as a game. In particular, they assume a game where there are a variety of different 'types' of player. In most, but not necessarily all, signalling games, we assume that players know their own type, but they do not know the type of their opponent.

A signal is a means of conveying your type to your partner, which in turn can influence their behaviour. To give a simple example of such a game, suppose that two individuals are about to pass one another down a hall. One individual A has their hands free and can signal which way they want to pass, while the other individual B is carrying something and cannot. Suppose that individual A can be one of two 'types,' they can be type t_l or type t_r. Type t_l has the following payoff

	Left	Right
Left	1	0
Right	0	0

and type t_r has the payoff

	Left	Right
Left	0	0
Right	0	1

We assume that individual B is indifferent, so they have a payoff matrix of

	Left	Right
Left	1	0
Right	0	1

Needless to say, this example is contrived, but it serves to makes the point.

Suppose that before the game begins, there is a 'move by nature' which determines whether player A is t_l or t_r, and that there is half a chance that they could be either. In this game, player A knows their own type, but player B does not. I assume that after the move by nature, player A is going to have the chance to signal player B, and that they may say either *left* or *right*. Assume the following evolutionary model:

1. Form a population of infinite size.

2. Pairs of individuals are picked to play the 'passing game.'

3. With half a chance, the given player is the *sender*, and the rest of the time they are the *receiver*.

4. With half a chance, the *sender* is t_l, and with half a chance they are t_r.

5. The sender can send a signal *right* or *left* to their opponent.

6. Players play their strategies and receive their payoff according to the payoff w_σ matrices above.

7. The frequency p of individuals of type σ changes according to the following

$$\bar{w}\Delta p = p(1-p)(w_\sigma - w_{\sigma^*}) \tag{5.1}$$

Suppose that there is the strategy *signaller and receiver* σ_{sr}, which plays the strategy *when the sender, if type t_l, signal left and play left. If type t_r, signal right and play right. If the receiver, go which way the sender signals.* The expected payoff for σ_{sr} when it plays itself is equal to

$$w_{\sigma_{sr}}(\sigma_{sr}) = \frac{1}{2}\left(\frac{1}{2}.1 + \frac{1}{2}.1\right) + \frac{1}{2}\left(\frac{1}{2}.1 + \frac{1}{2}.1\right) = 1 \tag{5.2}$$

It is not difficult to see why σ_{sr} is evolutionarily stable against a whole host of other strategies that we might play it against. For example, if we consider the strategy *liar* σ_l, which plays *if sender, signal left if t_r and right if t_l. When receiver play as σ_{sr}*, gains a payoff of

$$w_{\sigma_l}(\sigma_{sr}) = \frac{1}{2}.1 + \frac{1}{4}.0 + \frac{1}{4}.0 \tag{5.3}$$

Likewise, the strategy σ_o, which plays *as signaller, play as σ_{sr}. If receiver, go in the opposite direction that the signal suggests.* This gets a payoff of

$$w_{\sigma_o}(\sigma_{sr}) = \frac{1}{2}.0 + \frac{1}{4}.1 + \frac{1}{4}.1 \tag{5.4}$$

All of these strategies do worse than σ_{sr} which gets the payoff of one when playing itself. Although the result of this example is relatively obvious, it includes all of the techniques that I use in my full model in the following section.

Zahavi produced two models of animal signalling which, although seemingly similar, are actually different in important ways. The first type of signalling refers to signalling some genetic quality about the individual. The most famous example of this is the *handicap principle* in sexual selection models. In the handicap principle, the hidden type of player is the *genetic quality* of the partner. This genetic quality is assumed to be unobservable, but that it can be inferred indirectly through some actions. For example, a peacock might be extremely healthy, but there is no way to demonstrate their health to a female except by carrying around a giant tail that hinders them in all their activities. The idea is that the cost of the tail would be prohibitively expensive to those of lower genetic quality, and therefore indicates that the male is of high genetic quality. Notice that in this example, the trait in question is heritable.

Although similar, the logic of costly signalling is different to that of bond testing as described in the previous section. With bond testing, the hidden characteristic is is not a heritable genetic trait, but rather a phenotypic trait. In the example of the wagtails, the hidden trait is the valuation that the particular female has made regarding the quality of the male. In the chimpanzee and humans, it is the value of that coalition partner in that particular circumstance. Although the criteria on which the valuation was made will be passed onto the offspring, the valuation itself will not be, and what is of immediate interest to the receiver is valuation of the partner in their current context, not the genetic criteria.

5.3 A Game-Theoretic Model of Social Bonding

The theory of social bonding rests on three basic assumptions. Firstly, that there exists some social activity that two or more individuals would like to engage in for mutual benefit. Secondly, that there exists the possibility that one or more of the individuals involved in the activity could cheat their partners. Thirdly, that it is costly for individuals to be cheated. Together, these assumptions create a dilemma — individuals want to cooperate, but how do they know that their partner is really willing?

Social bonding solves this problem. In order to facilitate cooperation, it demands that individuals produce some costly signal for their partner. The signal is designed in such a way that it reveals their intentions to their partners, thereby creating the trust necessary for cooperation to occur. The signal is designed so that it would only be used by an individual who benefits from the cooperation. This means that the most important quality of the signal is that it must be *honest*.

As we will see, it is this 'honesty' that gives social bonding its counter-intuitive property, where a pair of individuals who intend to invest in a cooperative venture begin their relationship by engaging in some counter-productive activity. The activity *must* be counter-productive for it to be honest. If it did not infer a cost to the sender, then it would not provide any information to the receiver — all individuals would send it.

5.3.1 Must Signals be Costly?

The driving force behind models of costly signalling is the simple idea that in social interactions where there is a conflict of interests, the cost is a necessary component.

The reason for this is that if another individual is able to produce the same signal at no cost, then there is the possibility that they will produce the signal, but then cheat in the social interaction. This is called creating a 'dishonest,' or, 'fake' signal. If this type of strategy was associated with a gene, then we would expect this 'cheating gene' to spread through the population [Dawkins and Krebs, 1978, Grafen, 1990, Krebs et al., 1984, Dawkins and Guilford, 1991, Smith and Harper, 1995].

However, Zollman et al. [2012] showed the possibility of what they called 'hybrid equilibria.' Such an equilibrium is one where an individual plays a mixed strategy of sometimes signalling, and sometimes responding to a signal if it is sent. This theory has important empirical implications, because if an experiment demonstrates that some signalling is occurring, we cannot conclude that it is necessarily the full costly signalling equilibrium—it could equally be one of the hybrid equilibrium.

5.3.2 The Model

The standard game-theory model of cooperation is the repeated prisoner's dilemma. In this model, we assume that two players meet to play a series of prisoner's dilemmas. The conditions for this game are described in section 2.4.2, so I do not repeat them here. In this model, I use the repeated prisoner's dilemma as the framework.

It is standard practice in the repeated prisoner's dilemma to assume that each player in the game has an identical payoff matrix, and that each player is privy to this information. In my model, I drop this assumption of certainty. Instead, I assume that players are uncertain about the payoff matrix of their partner, and are therefore uncertain about what move they might play. I represent this uncertainty with a *random move by nature*. As we have seen, a random move by nature is a way abstracting a game of incomplete information. Before the game begins, we assume that nature is going to play a move, and that this move is going to decide the *types* of the players in that particular game. The player's type determines their payoffs, and therefore affects what strategy they will play in that game. I assume that each player always knows their own type, but not the type of their partner.

There are many reasons why two individuals that meet might not want to engage in a mutually beneficial relationship, but I believe that the most important one is the *opportunity cost* involved in doing so. Engaging in cooperation with one individual necessarily reduces the amount of time and resources available for cooperation with another. Even if an individual can identify a potential cooperative enterprise in which they might engage, they cannot know whether their partner has something better they could could be doing elsewhere. Given that there are often many more possibilities to enter into cooperative relationships than there is time to do so, the question becomes what is the best strategy for ensuring that your given partner intends to cooperate with you or not. To answer this problem, I propose the following model:

1. A pair of individuals are selected at random from a global pool to play a game G.

2. Before the game begins, there are two random moves by nature, E_1 and E_2 (each one occurs with probability p). Whether event i occurs determines the *type* of player i. A player's type will determine their payoff in the game.

3. Each player can signal their partner. Signalling is represented by both players simultaneously choosing a number k to represent the maximum cost that they are willing to spend on a signal. Each player pays a cost equal to the lowest number.

4. In each round of the game G, both partners choice of whether to *cooperate* or *defect* against their partners. These choices can be affected by the type of player, the signal that they send, and the past play of their partner.

5. Any act of cooperation on the part of player i gives a higher payoff to player $-i$ (where $-i$ means the other player), while giving a lower payoff to player i.

6. After each game G, there is a probability δ that another game is played.

7. When the last game has been played, each individual sums their payoffs from all of the games to receive a final payoff.

8. The proportion of a population playing a particular strategy evolves according to differences in the average payoff that strategy receives. If strategy X is at frequency x, has an average payoff of w_x, and strategy Y is at frequency $1 - x$ and gets and average payoff of w_y then the change in x between generation is

$$W \Delta x = x(1 - x)(w_x - w_y) \tag{5.5}$$

where W is the average payoff of the entire population.

The events E_i determine the type of player i. Player i's type determines whether the partner they have been paired with in this round is one which they would benefit from cooperating with. In our game we consider two types. The first type, which we call *partner preferred* (PP). A PP type sees that there exists a mutually beneficial cooperative arrangement with their partner. The payoff of a PP type each round is the standard prisoner's dilemma.

	Cooperate	Defect
Cooperate	$b - c$	$-c$
Defect	b	0

So they pay a cost c in order to bring a benefit to their partner of size b.

When event E_i does not occur then we say that individual is type *not preferred partner* (NPP). When an individual is type NPP, then we say that they can benefit more by spending their time elsewhere. We interpret this in our model by saying that they have a higher cost for playing the cooperative strategy. When player i is of type NPP then they have the following payoff matrix.

	Cooperate	Defect
Cooperate	0	$-b$
Defect	b	0

Here we see that they can still benefit from having their partner cooperate with them. But in order to represent the changing opportunity costs in the two cases. I assume that there is an added cost, which transforms $c < b$ into $c = b$.

If we assume that $p = 1$ and that no players signal, then we have the standard model of the *repeated prisoner's dilemma* that follows Axelrod and Hamilton [1981]. We can define the strategy *reciprocation* σ_R which will cooperate on the first round and will cooperate in subsequent rounds until their partner defects, in which case it always defects. We also define the strategy *always defect* σ_{AD}, which always defects. The payoff for each type is as follows:

	Reciprocator	Always Defect
Reciprocator	$\frac{b-c}{1-\delta}$	$-c$
Always Defect	b	0

Proposition 18. *In game G, when $S = \{\sigma_R, \sigma_{AD}\}$, and $p = 1$, then σ_{AD} is always an ESS, and σ_R is an ESS if*

$$\delta b > c \tag{5.6}$$

Proof. σ_{AD} is always an ESS because $w_{\sigma_{AD}}(\sigma_{AD}) = 0 > -c = w_{\sigma_R}$. $w_{\sigma_R}(\sigma_R) = \frac{b-c}{1-\delta} > b = w_{\sigma_{AD}}(\sigma_R)$ when $\delta b > c$. □

Now suppose that $p < 1$. Define the strategy *Risky Reciprocator* (RR) that plays like an R when they are type PP, but plays like an σ_{AD} when they are type NPP. We can record their payoff, depending on their type in the following payoff matrix.

	PP RR	PP AD	NPP RR	NPP AD
PP RR	$\frac{b-c}{1-\delta}$	$-c$	$-c$	$-c$
PP AD	b	0	0	0
NPP RR	b	0	0	0
NPP AD	b	0	0	0

Given that we know that the chance that a given player is of type PP is p, we can record the expected payoffs of the two strategies in the following matrix

	Risky Reciprocator	Always Defect
Risky Reciprocator	$p^2\frac{b-c}{1-\delta} + 2p(1-p)(b-c)$	$-pc$
Always Defect	pb	0

Proposition 19. *In game G, when $S = \{\sigma_{RR}, \sigma_{AD}\}$, then σ_{AD} is an ESS when $p > 0$, and σ_{RR} is an ESS when*

$$p > \frac{c}{b-c}\frac{1-\delta}{\delta} \tag{5.7}$$

Proof. To demonstrate that σ_{RR} is an ESS, recall from equation that σ_{RR} is an ESS if

$$w_{\sigma_{RR}}(\sigma_{RR}) > w_{\sigma_{AD}}(\sigma_{RR}) \tag{5.8}$$

which occurs when the conditions of 5.7 hold. To demonstrate that σ_{AD} is an ESS, we simply see that

$$w_{\sigma_{AD}}(\sigma_{AD}) = 0 > -cp = w_{\sigma_{RR}}(\sigma_{RR}),\ \forall p > 0 \tag{5.9}$$

□

I also show that there is an unstable equilibrium. Recall that if we suppose that a proportion x of the population play strategy σ_{RR}, then there is an internal equilibrium when

$$W_{\sigma_{RR}} = xw_{\sigma_{RR}}(\sigma_{RR}) + (1-x)w_{\sigma_{RR}}(\sigma_{AD}) = xw_{\sigma_{AD}}(\sigma_{RR}) + (1-x)w_{\sigma_{AD}}(\sigma_{AD}) = W_{\sigma_{AD}} \tag{5.10}$$

Proposition 20. *In game G, when $S = \{\sigma_{RR}, \sigma_{AD}\}$, then there is an unstable equilibrium when*

$$x = \frac{c}{(b-c)\left(\frac{1}{1-\delta} - p\right)} \tag{5.11}$$

Proof. Putting the payoffs into equation 5.10 and solving for x gives the condition in equation 5.11. □

My interpretation of this model is as follows. When p is low, so the chance that any individual will meet a preferred partner is low, then a strategy that cooperates with its preferred partner is not an ESS.

5.3.3 Costly Signalling

In this section I demonstrate how when p is low enough so that σ_{RR} is not an ESS, then signalling can be utilised by a population to facilitate cooperation. The idea behind bond testing is that after the types of players have been chosen by the random move by nature, the two players will decide whether to pay a cost to signal their partner. This round of signalling occurs by the following process. Each player secretly decides how much time and effort (translated into the common currency of payoff) they are willing to sacrifice to signal. I also allow individuals to play conditional bonding strategies. For example, they they can play the conditional strategy *signal up to cost s, unless the partner stops signalling before that, in which case break off signalling.* As we will see, this sort of strategy will allow individuals to save resources by not wasting valuable bonding time on partners who are not interested in cooperating with them.

With this in mind we consider the strategy *bond-tester* (σ_{BT}). The σ_{BT} strategy will play the following signalling strategy *signal up to strength s unless the other player has stopped signalling, in which case, immediately stop signalling.* During the round of cooperation, it plays the following *play like R if your partner signalled up to cost s. Otherwise, play like AD.* We can compare the payoff of this strategy with the strategy σ_{AD} that never signals and never cooperates by the following payoff matrix

	PP BT	PP AD	NPP BT	NPP AD
PP BT	$\frac{b-c}{1-\delta} - s$	0	0	0
PP AD	0	0	0	0
NPP BT	0	0	0	0
NPP AD	0	0	0	0

The only non-zero entry in this matrix is when two bond testers meet. The payoff that they receive is equal to the benefit from the repeated prisoner's dilemma, and the cost is s. Also notice that the σ_{BT} strategy does not pay a cost to signal to those who do not signal back, and so do not waste any resources. The payoff for strategy σ_{BT} is recorded in the following payoff matrix:

	Bond Tester	Always Defect
Bond Tester	$p^2\left(\frac{b-c}{1-\delta} - s\right)$	0
Always Defect	0	0

Proposition 21. *In game G, when $S = \{\sigma_{BT}, \sigma_{AD}\}$, then σ_{BT} is an ESS whenever*

$$\frac{b-c}{1-\delta} > s \tag{5.12}$$

and σ_{AD} is an ESS whenever the opposite inequality holds.

Proof. The proof for this is straight forwards, since the only payoff that deviates from zero is when a bond tester plays another bond tester and the conditions given in equation are when this is greater than zero. □

Recall the three basic assumptions that we made at the beginning of this section. Firstly, there must be a cooperative activity that in which both individuals would like to engage. Secondly, that there exists the possibility that an individual could be cheated. Thirdly, that being cheated entails a cost. These assumptions are formalised by a variant of the repeated prisoner's dilemma. We have seen that under the correct conditions, we expect a strategy to evolve where, before the cooperative activity commences, both individuals engage in some costly, but otherwise useless social bonding. In the following section, I show how this result is robust when we include strategies that try to subvert the signal to their own ends.

5.3.4 Cheaters

In this section I demonstrate how the cost of the signal can be selected such that it is not individually advantageous for an individual who is not a PP to send it. Consider the strategy *liar* (L) that always signals up to strength s, but only cooperates when they are of type PP. When this strategy plays against the σ_{BT} type then they get the following payoffs

	PP BT	PP L	NPP BT	NPP L
PP BT	$\frac{b-c}{1-\delta} - s$	$\frac{b-c}{1-\delta} - s$	0	$-s-c$
PP L	$\frac{b-c}{1-\delta} - s$	$\frac{b-c}{1-\delta} - s$	0	$-s$
NPP BT	0	0	0	0
NPP L	$b-s$	$-s$	0	$-s$

Therefore the two strategies get the following payoffs:

	Bond Tester	Liar
Bond Tester	$p^2\left(\frac{b-c}{1-\delta} - s\right)$	$p^2\left(\frac{b-c}{1-\delta} - s\right) - (1-p)p(c+s)$
Liar	$p^2\left(\frac{b-c}{1-\delta} - s\right) + p(1-p)(b-s)$	$p^2\left(\frac{b-c}{1-\delta} - s\right) - (1-p^2)s$

Again, using the condition from equation 5.3.2, we see that σ_{BT} is an ESS when:

$$w_{\sigma_{BT}}(\sigma_{BT}) > w_{\sigma_L}(\sigma_{BT}) \tag{5.13}$$

This leads to the following result:

Proposition 22. *In game G, when $S = \{\sigma_{BT}, \sigma_L\}$, then σ_{BT} is an ESS whenever*

$$s > b \tag{5.14}$$

and σ_L is never an ESS.

Proof. To prove the first part, we need to show that

$$w_{BT}(\sigma_{BT}) = p^2\left(\frac{b-c}{1-\delta} - s\right) > p^2\left(\frac{b-c}{1-\delta} - s\right) + p(1-p)(b-s) = w_S(\sigma_{BT}) \tag{5.15}$$

To prove the second part, we see that $(1-p^2) > p(1-p)(c+s)$ whenever $s > pc$, but since $s > b$, this will always be true. □

This is the main, novel result of this section. It means that if the cost is set sufficiently high, in particular higher than the marginal advantage that a cheater would get from lying about their type, then it will not be in their interest to send the signal. This in turn means that the signal can be trusted.

5.3.5 Too Costly?

I briefly discuss a question that may present a difficulty to this model — how much higher do we expect s to be than b? Indeed, it is possible to imagine a population where s is significantly larger than b. Suppose that we have a stable population playing σ_{BT} where s is larger than b. Is there any way for a strategy σ^* to invade, that plays like σ_{BT}, except with a much smaller s, call it s^*? If it uses the same signal but just signals to a lesser degree, then the answer is no. To see this, we just have to notice that the single invading σ^* gains a payoff of $-s^*$ when it meets a σ_{BT} type, because that type will never cooperate with it. Even if there was positive assortment r according to equation 2.5.5, so that the probability of meeting the same type is equal to $r+(1-r)x$, where x is the frequency of the invading type in the population, in order to invade, then

$$r > 1 - \frac{(s-b)(1-\delta)}{b-c} \tag{5.16}$$

meaning that the degree of positive assortment would need to be large.

There is, however, another way that it could invade, which was first suggested by Robson [1990] investigated a game with the following payoff matrix:

	u	d
u	1	0
d	0	2

The question here is whether it was possible for a d strategy to invade a population of u's. Robson found that it could, via strategy m, which played *send a costless signal before the interaction begins, if the same signal is received, play d, if it is not, play u.* To see how this can invade a population of u's, inspect the following payoff matrix:

	u	m
u	1	1
m	1	2

From which it is obvious that m is the only ESS. What maintains the honesty of the honest signal? In this case, we do not need to worry about honesty, because there is no benefit from cheating. In fact, the cheater hurts themselves. In our model, the payoff matrix looks like this:

	σ_{BT}	σ^*
σ_{BT}	$p^2(A-s)$	$p^2(-b)$
σ^*	$p^2(-b)$	$p^2(A-s^*)$

where $A = \frac{b-c}{1-\delta}$. It should be clear that this model has the same structure as the game with u versus d studied by Robson [1990], meaning that the same conclusions apply, so that we expect σ^* to invade if it can send a single before hand that only other σ_*'s can receive.

What this means to our model is that suppose there is a population at one equilibria for social bonding. If the cost of using that bonding mechanism is particularly high, then it can be invaded by a type that, before it bonds, it first sends a signal to determine what sort of bonding it will do. If it receives the same signal back, it will conduct the cheaper, and not the more expensive bonding mechanism. Of course, if the new mechanism is too cheap, so that $s < b$, then we expect another individual to invade this population that sends the free signal all of the time, thereby undermining the integrity of the signal. The type of costless signal does not matter. It could be a special call, or a dance, or even emitting a scent.

5.3.6 Summary

The hypothesis that I have investigated here is that social bonding acts as a means of discovering the intentions of social partners. The idea is that each individual tests their partner by observing them engage in some otherwise evolutionarily pointless activity. If their partner pulls out, then they have signalled that they are not interested in forming a relationship. The reason that bonding must be costly is to protect against cheating. Even if a particular individual does not wish to cooperate with their partner, they mights still benefit from having their partner cooperate with them, and so bonding protects against this parasitic strategy. The important point to remember is that the cost of bonding must be greater than the value of cheating.

5.4 Empirical Validation

Humans bond a great deal. The model outlined in this chapter shows that it is evolutionarily logical to do this when humans create long-term social bonds for mutual benefit. In presenting evidence for this theory, I find that I have many different examples to choose from. So my criteria for selecting examples in this section are firstly because they are controversial, secondly because they are important, and thirdly because I find them interesting.

The first example that I investigate is formal gift giving in the !Kung — a population of bushmen who reside in the Kalahari. The !Kung use gifts in order to establish long-term relationships [Wiessner, 2002, 1982, 2009]. The second example is that of ritualistic warfare in small-scale societies. Many anthropologists believe that ritualistic differs qualitatively from other forms of warfare due to its symbolic nature. Here I investigate the possibility that this form of warfare acts as a bonding mechanism between members of the group.

5.4.1 Hxaro and the !Kung

In many small-scale societies, it is very important to have a wide social network. These networks are not just limited to other individuals within the same group, but also to individuals from different groups. They are important for a variety of reasons — they offer a place to go if staying within the same group is unfeasible due to tensions between individuals, they can offer shelter in times of famine or war, they might offer places to store important economic goods like cattle and many more.

In this section, I investigate the ethnographic literature on the !Kung — a population of Kalahari bushmen — as a case study of this model of bond testing. The !Kung are hunter gatherers who live in northwest Botswana, northeast Namibia, and southeast Angola. They live in a regions called n!ores. Each n!ore has enough food and water to sustain at least one band throughout an average year. All persons inherit one n!lore from their mother, and one from their father. They do not inherit the land itself, but the right to exploit the resources along with others who hold a similar right. Over their lifetime, an individual will select an n!ore in which to live, and we say that they develop a strong hold over this n!ore. Children, in turn, will inherit the n!ore which each parent held strongly, although in some cases they may trace their rights back to their grandparents.

Because each member !Kung generally hold the rights to hunt in 2 n!ores, and can live in their spouse's n!ore, the population typically distributes itself with minimal conflict. A person will not utilise the n!ore of others without having the correct ties to ask for permission, the system allows the !Kung to plan their yearly behaviour on the basis of accurate information about who has rights to utilise the resources.

On a good day in the Kalahari, the environment provides good returns on hunting and farming. However, draughts are common, and these returns are subject to high variation. Drought and other risks mean that hunting may not always be fruitful. The !Kung have learnt to reduce their individual risk through social obligation — a formal relationship known as hxaro (pronounced 'har-ro').

Hxaro is a relationship based on mutual responsibility which is formed between two individuals or families. The arrangement is to help those families during times of need. Deciding whether or not to accept a particular family as a hxaro partner is an important decision. If a relationship comes out as expected then both parties will come out ahead as the burden of helping during good times is often much less than the value of receiving when it is needed.

Hxaro relates to the model of social bonding described above through the way in which relationships are established. To initiate a hxaro relationship, a gift is sent from one family to another. To accept, a gift is returned. After a trail period of around a year, in which a series of gifts are transferred back and forth, the relationship is considered formed. Once secured, the relationship is maintained by a constant flow of gifts and is only cancelled through loss of interest indicated by a reduction of volume of gifts sent and received.

What the !Kung expect their partner to do depends entirely on their abilities and location. If they are in the same camp the relationship generally smooths over income disparities. For example, they are likely to be in the 'first wave' of meat sharing, and a person who for some reason cannot gather vegetables can count on a hxaro partner for help, even though vegetables are only usually shared within the family unit. Hxaro partnerships in neighbouring camps lead to frequent visits, a share of the meat of a large catch and access to a partner's n!ore during bad times. Because hxaro is not a contractual arrangement with an institution to enforce the relationship, it can be difficult to avoid exploitation.

The !Kung enjoy a relatively short working-week. The reason for this short working week has been the subject of speculation. *Prima face*, there is no reason why women could not gather more nuts at the end of the wet season for storage, and why men do not help gather more when they are not hunting and so on. Wiessner's theory is that it is a combination of the fact that any excess would be expected to be given away, and therefore is of little personal use, and secondly, because of the amount of effort that is expected to be put into maintaining hxaro relationships. Many hours of the week are spent making and remaking gifts, gathering information about who is in need and who has resources, and ensuring that they get their fair share of other's resources. Taken in sum, Wiessner argues, if the business of maintaining social relationships are taken into account, a 14 hour week may quickly become a 40 hour week.

5.4.2 Hxaro as Bonding

Here I briefly review the assumptions of the model before comparing them to the empirical literature. The model assumes that two individuals meet, and that there is a chance p that either one desires to enter into a relationship. If they both decide that the relationship is mutually beneficial, they begin bonding with one another. This bonding occurs simultaneously, and so one can pull out if the other pulls out immediately. The cost of the bonding activity s must be greater than the benefit of cooperation that a cheater might gain b. Once the bonding is over, the pair of individuals will cooperate according to the standard assumptions in the repeated prisoner's dilemma, that individuals pay a cost c , and receive a benefit b, where $b > c$. Here I apply my method of justifying each assumption with reference to the ethnographic literature.

Is there a small chance that any particular individual that the !Kung meet will be willing to enter into a relationship? (p small?) Accepting a hxaro partner is a serious decision for the !Kung. Typically, a husband and wife will have to consider a great many factors before they decide to settle on a specific partner. For example, they may have to consider their physical proximity, their skills, reputation, and so on. Since the !Kung's goal in creating relationships is to give themselves a diverse set of options during times of need, then the value of a specific partner is not simply a feature of that partner themselves, but it is also a feature of the relationships that the family already has. This also means that there is no way for a potential partner to know whether they are serious about the relationship without the gift giving. It is practically impossible, or at least highly impractical, to find out how much a potential partner values you by discovering their current social network.

Why is p likely to be small? Because there are so many potential partners that the !Kung can have. In fact, potentially, any other individual in the population could be a partner. But given that networks will only reach a limited size, the chances that randomly picked pairs of individuals from the population would be good partners is relatively small.

Is there a benefit from cheating a partner if you do not value them? ($b > 0$?) By cheating a partner in this context, I mean receiving help from a partner that you do not help in return. This is beneficial simply because it is getting something for nothing. That this benefit might be marginally less than what could be achieved in another relationship does not matter in this context, since it is still more beneficial than nothing occurring at all.

One possibility is that mechanisms like reputation could regulate this within a society. However, in the case of the !Kung, often individuals want to make relationships that are quite far away, and so it is extremely difficult to obtain any reliable information about their previous relationships.

Is bonding simultaneous? In the case of gift giving, social bonding does not exactly occur as my model assumes. The model assumes that the act of bonding occur at precisely the same time, so that if one partner stops bonding, their partner knows this immediately. In the !Kung, the act of bonding occurs over the period of about a year, where small gifts are sent back and forth. This means that an individual will have to pay an initial cost of sending a gift before receiving any information back from their partner. However, in the grand scheme of things, the cost of creating each individual gift is relatively small, and therefore I believe that I am not too far off the mark by assuming that it is simultaneous.

Is bonding sufficiently costly? ($s > b$?) This is the most difficult question to answer without further quantitative data. However, obtaining data on the relative costs of engaging in a costly activity of gift giving against the relative benefits of receiving help is likely to be very difficult. We know that the costs of creating the gifts can be considerable — individuals may spend a great deal of their time budget creating gifts. We also know that the benefits of receiving help may be considerable, particularly during times of need.

Do the assumptions of the repeated prisoner's dilemma accurately reflect what occurs once a relationship is established? ($b > c$?) As described above, the benefits of entering into a relationship depends primarily on the proximity of the partner. If they are close by, then your partner is likely to include you in the first wave of food sharing. If they are further away, and particularly if they reside in another !nore, then the benefit is a place to go in times of need. In the model I represent this process as a repeated prisoner's dilemma where players play simultaneously. Perhaps a more appropriate model would be one where favours alternate, rather than ones where the moves were simultaneous. However, this does not significantly alter the qualitative results of the model. For example, adopting a model of a prisoner's dilemma does not assume that individuals count the help given and demand something of equal value in return. We could also think of it as a two player version of the *mutual aid game* — a system of insurance — that I investigated in section 4.5.

5.4.3 Pre-State Warfare

Human populations are adept at recruiting participants for war. Small-scale societies are able to muster proportions of their male population for war to a degree that is comparable to industrial state-level societies in the 20th century. While there are some societies who do not engage in warfare at all, it is estimated that between $90 - 95\%$ of all human societies are (or at least have been) involved in some kind of fighting [Keeley, 1996].

The traditional anthropological view of 'primitive' warfare is that it is sporting and ineffectual [Wright, 1942, Turney-High and Rapoport, 1949]. This view mostly derives from observing pitched battles — the largest-scale and most protracted kind of warfare. Most pitched battles are prearranged, typically by issuing a challenge and naming the site. For example, the Dugam Dani send a herald to the enemy border to shout a challenge, which might be accepted and a battle ground chosen [Keeley, 1996]. Normally, when battle takes place, the warriors are dressed and painted in special decorative paraphernalia: war paint, headdresses, armbands are all common. The regalia are not generally chosen for their effectiveness in combat, but rather for their decorative effect. Indeed, this has been true for modern warfare up to and including world war two. For example, the last war fought by the British in their famous bright red coats was the Zulu War of 1879. Likewise, the French army in World War I wore a bright blue uniform, and some of the German troops wore the preposterous *pickelhalbe*.

Other features of battles that are cited as evidence of their sporting nature is the custom of 'counting coup' among North American Plains Natives [Keeley, 1996]. When a warrior counts a coup, he performs an act of daring (the French word *coup* attached itself to the custom). Typical daring acts include stealing a horse, killing an enemy with a hand held weapon and saving a wounded comrade. Again, these attitudes of bravery are also found in modern day warfare. Bravery is rewarded with formal decoration and so on.

When battles commence, many of them consist of little more than two lines of warriors armed with throwing spears or bows, firing at one another at about maximum effective range. They are normally terminated by agreement, typically when each side has suffered a few casualties. These features of prearrangement, elaborate dress, taunting, and low casualty rate give primitive battles their ritualised allure.

These pitched battles tend to be markedly different from other forms of inter-group violence. By far the most common type of warfare is small-scale raiding and ambushes. These normally involve handfuls of men sneaking into enemy territory to kill one or a few persons. Compared to the large-scale warfare, these small raids tend to have very high mortality rates. Raids tend to kill indiscriminately. For example, one technique adopted by the Mae Enga is to quietly surround enemy homes just before dawn and kill the occupants by thrusting spears through the flimsy walls. They therefore also tend to lead to more deaths. For example, among the Dugum Dani, over six months, seven ritual battles had killed only two men, but in that same time period nine raids had killed seven people. The reason for this is that the victims of raids are normally unprepared to fight and attempt to flee. This is evident from the fact that most wounds are inflicted from behind [Keeley, 1996].

5.4.4 Casus Belli

Trying to describe the causes of war is a difficult task. Even formulating an appropriate question is vexing. Do we include all of the motives, or only the publicly declared ones? Should the motives be inferred from the operations, result, or the effect of the wars?

From the ethnographic literature, we know that the declared motives of the individuals are often at odds with those inferred by and external observer. Social conditions may exist for long periods of time while war only occurs sporadically. Similar grievances are sometimes resolved without violence, or by war.

Ethnographer Koch [1974] illustrates the problems inherent in deciphering the causes of war in work on the Jalemo of New Guinea. Village A owed village B a pig for help in a previous war. Meanwhile, a man from village A heard some (untrue) gossip that a man from village B had seduced his wife, and he assaulted the alleged seducer. Village B reacted to this by performing a raid against village B. This then led to a formal battle where both sides took casualties, but no individual was killed.

There was then a truce, but the truce ended when a warrior from village A ambushed a resident of village B. The following day there was another pitched battle. Eventually, this back and forth led to all out war that lasted for several years. The question is, which of these events 'caused' the war?

In the example above, there is no single 'reason' why the two villages went to war. There may be an 'official' reason, but it is quite possible that all of the individuals taking part did so for individual reasons. In the evolutionary literature, we are not very good at describing how a variety of different, individual strategies can culminate in a group taking to the field to fight. We tend to think statistically.

Obviously warfare does achieve things. It can move territorial borders, can result in significant loss of life, and can result in the stealing of valuable items. These examples of war are ends in themselves. But the ritualistic warfare observed in so many human groups does not seem to conform to these materialistic explanations. Territory is not always taken, indeed, is not always at stake. The question is, if not for these things, then what might warfare be for?

5.4.5 Warfare as Bonding

One possibility that seems to have been largely overlooked by the modern literature is that one motive for going to war could be that it represents an act of social bonding between the individuals who take part. As the quote from Henry V in the opening of this chapter indicates, individuals who go to war do seem to experience the subjective feeling of 'brotherhood.'

This theory of warfare focuses on explaining its ceremonial aspects. Were individuals primarily interested in the economic outcome or the acquisition of territory, they would be unlikely to spend so much time and effort creating elaborate decorations that impede them in combat. What is more, the conditions conform to the assumptions made in the model presented in section 5.3.4 — individuals simultaneously take part in an activity that is individually costly.

To highlight the possibility that warfare is often conducted in order to facilitate cooperative relationship, consider the *great ceremonial wars* (*yanda andake* — henceforth 'great wars') of the Mae Enga [Wiessner et al., 1998]. The ceremonial wars were a series of highly ritualised 'tournament' wars fought over several generations. They always involved two tribes, or two pairs of tribes fighting one another in armed combat.

To understand the ceremonial wars, Wiessner et al. [1998] argues that it is important to consider their dual nature. First, warfare is a means of opposing an enemy, second it is a means of making allies. The great wars evolved out the second of these reasons. The Enga conducted these wars at a scale that was hitherto unheard of, involving entire tribes, rather than the smaller clan segments that would normally fight. Emphasis was placed on display of force rather than ousting an opponent, and the unity that such fighting encouraged. Exchange took precedent over fighting, meaning that the ceremonial wars were planned to 'harvest' the economic exchanges that followed.

The battles themselves were highly ritualised. Hundreds of warriors adorned in full ceremonial attire converged on the battlefield while spectators watched from the sidelines. The actual fighting followed the procedure of standard wars [Meggitt, 1977], but death tolls were low in proportion to their size and duration. Violence was not permitted outside of the battle field and so there were no raids.

During the wars, warriors typically stayed in the house of an ally. Every night, they ate together and rehashed the events of the battle. The many days of fighting side by side and the nights together formed bonds between the hosts and the hosted. This is why it is said that

> The Great (Ceremonial) Wars were events for socialising (between the warriors and the hosts, not between opponents). After getting to know each other, they would kill many pigs and hold feasts.[Wiessner et al., 1998]

Is there reason to form alliances with members of other groups and are there many opportunities to do so? (is p small?) There are many reasons why individual would like to establish partners in other groups. Trade is one reason — among the Enga there is a great deal of trade in salt, stone axes, foodstuffs, decorative jewelry and more Wiessner et al. [1998]. One criteria for being a 'big-man' among the Enga is that they have an ample supply of decorative items that they can lend to their fellow clans people.

Another reason for establishing relationships is that they can offer loans to finance a variety of different activities, the most important being to make payments to allies in wars, to make reparations following wars, and to pay the costs of bride wealth. In more desperate times, allies provide a safe place to go during wars.

The Enga are relatively densely populated, and so there are many opportunities to create ties across group. In terms of the equations in section 5.3.4, this means that the chance that any particular partner will be ideal is likely to be relatively small.

Did the great wars achieve anything? (is $s > 0$?) Recall from the equation in section 5.3.4 that the activity that individuals engage in must be net costly. If it achieves anything for its own sake, then it is likely to be accepted as such. Did the great wars achieved anything for the individuals who took part?

I think that it is important to conceptually differentiate two distinct categories of organisation here. These are the individual strategies played by the individuals who take part in the activity, and the structure of the activity itself. To give an example of this that is closer to home for academics, consider the social function that a conference plays in the scientific community. Conferences are used by many as ways of advertising themselves, forming new alliances, engaging in joint activities and so on. However, the conference is also structured around improving the science. Individuals will generally respond to the incentives that they are given, but, as I have argued throughout this dissertation, we are able to create incentives to bring about ends that are good for the entire community.

The same seems to be true of the great wars for the Enga. For the individuals, the great wars were a means of forming alliances. For example:

> They (the great wars) were designed to open up new areas, further existing exchange relations, foster tribal unity, and provide a competitive but structured environment in which young men could strive for leadership. [pp. 267 Wiessner et al., 1998]

But it was equally true that the great wars brought other benefits, the most obvious being a show of unity:

> The great wars were not serious ones, but they were not mock wars either. Warriors received arrow and spear wounds. and some of them died from their wounds. The underlying purpose of these wars was to bring people together — they were formal and ceremonial. They were fought to show the numerical strength and solidarity of a tribe and the physical build and wealth of its members.[pp. 268 Wiessner et al., 1998]

So the wars do seem to achieve something on a higher level of social organisation.

I think that this example highlights what humans are able to do nicely — we can manipulate the motives of individuals in order to bring about certain ends which we want. In this case, the individuals are all motivated to take part in the wars since doing so offers a great opportunity to create new alliance partners, and there is no other way for them to achieve this. but at the same time, this creates a situation where the group as a unit can demonstrate its strength and unity (unity is not a feature of the individuals, but of the group as a whole).

This means that, from the perspective of the individual achieving their own ends, s, the cost of engaging in the bonding activity, is likely to be greater than 0. However, from the perspective of 'group ends,' engaging in the wars are likely to have been beneficial. In terms of what the costs actually were, these would simply be the risk of injury or death through being shot by and arrow during the battles. Although these costs were much less than real wars, they were certainly not negligible.

Do individuals engage in warfare simultaneously? In this example, individuals clearly do. Of course, it is not possible to observe everyone at the same time, but then, engaging in war is not likely to bond an individual to everyone on their side equally. What is more important is who is close to you when the fight was going on. Who made sure you were safe if you got injured, and so on. If your friend left midway during the fight, then it would be quite obvious.

Were the costs comparable to the costs of being cheated? ($s > b$?) Like the last example, this is the most difficult to estimate given the current data. The help that an Enga might receive from an ally could be great, life-saving even. But there is also a chance of dying during the warfare. As it currently stands, this is the most important assumption to test. However, doing so is likely to be extremely difficult, because we are unlikely to be able to calculate the lifetime effect of being cheated by an ally in a way that is statistically significant. Remember that in evolution, the effect only has to be relatively slight over many generations to culminate in a change in gene frequencies. I suggest that the best means that we have might be to simply ask the Enga what they think the costs are. This is no less true in any population though. We ourselves avoid being cheated by our friends, and we know the things that we might do in order to insure against this, but it is unlikely that we would be able to give a precise value of the lifetime fitness cost that such a thing would engender.

Does the model of reciprocal altruism represent the relationships after the relationship is formed? ($b > c$?) As above, there are many different ways that relationships could be formed, and so the repeated prisoner's dilemma should not interpreted to literally here. But it seems likely that reciprocation is the mechanism that individuals use once the relationships have been established. The examples of trade and finance mentioned above are examples of this.

Does warfare provide any information? One criticism of this theory of warfare is that individual within smaller groups do not need to be bonded because they already are bonded. While this is probably true, it does not consider that the purpose of bonding might be with respect to some power-play made within a community. Alliances shift within communities, and it is necessary to keep up with the relationships around you. It is possible that the shared risks of warfare offer such information.

I believe that the most important point here is to realise that there is probably not going to be a single reason for any individual act of war, as discussed in section 5.4.4. In the case of the great wars, which brought individuals together who were not previously bonded, the bonding aspect is likely to be important. In other times it may not be.

5.5 Discussion

Here I have provided two quite different examples of how social phenomena can be explained in terms of social bonding. There are likely to be many other examples — I have already noted the social phenomena of scientists at conferences, but I believe that there are many others. Consider the example of inter-city football hooliganism. It is common for sets of supporters of one team to arrange a place to meet the supporters of another, and, dressed in the decorative shirts of their team, engage in fights that very much resemble pre-state warfare. These incidents do not generally result in fatalities, and like many of the examples of warfare in small-scale communities, there is no territory at stake. In fact, there are no material gains to made at all — only a cost if you are injured. It therefore seems entirely plausible that the evolutionary purpose of these examples of violence is that it is costly, albeit only slightly so.

The example of warfare highlights the difficulty that we have in describing social phenomena that involve many individuals interacting at once. The question 'what is warfare for' seems innocuous enough, but such a question already supposes that there exists a simple answer. The fact that warfare might lead to the acquisition of resources and might allow individuals to bond are not necessarily alternative hypotheses. There may be some circumstances where, if the material motives for war were removed, then the warfare would not occur, but as we have seen, there are many examples of warfare where not a great deal of material is won anyway.

5.6 Conclusion

In this chapter, I present a novel model of social bonding. The logic of this model is inspired by the theories of costly signalling put forwards by Zahavi. The idea is based on three basic premises, which, in one form or another, are found in a variety of different species and contexts. These are firstly that the members of that species wish to engage in cooperative activities with one another. Secondly that there is a chance that they will be cheated. Thirdly that there is a cost to being cheated and that individuals would prefer to avoid these costs.

These assumptions are brought together in a model based on the repeated prisoner's dilemma. This model represents the basic idea that individuals can benefit through repeated interactions. The novel aspect of this model is that, before the game begins, there is a random move by nature which determines the payoffs of the players involved. This move determines the 'intentions' of the players.

Signalling is a device that can be utilised by the members of a population in order to discover the intentions of their partner. But to guard themselves against cheats, the signalling must be costly. The idea is that this cost will only be paid by those who believe that they could benefit from the cooperative activities. In the model, the actual method employed in the signal is also significant. This is that both individuals must be willing to signal simultaneously. The reason for this is because then it is possible to see if your partner looses interest immediately, so then you can stop signalling. If this did not occur and the signal were sent in a 'block,' then the costs of signalling would outweigh the benefits.

Social bonding relates to social contract because in order to police the contracts, we must be able to form coalitions of mutual defence. However, larger coalitions are not necessarily better than smaller ones, because the marginal benefit of having one extra partner decreases with the number of partners an individual has. The hundredth partner is worth less than the first. At any given time, it is vital that an individual ensures that they can call on the support of their allies in times of need, and since human communities are normally fluid, this means being able to bond with new partners and to test the bonds of older ones constantly.

I use this model to understand two quite different examples of social bonding. The first is a formal insurance relationship that exists among the !Kung called hxaro. The 'bonding' part of this is represented by a series of formal exchanges of gift that have no economic value. The second example is that of ritualistic warfare, which many small-scale societies engage in, but generally do not seem to conform to the standard explanations of warfare based on taking territory.

This line of research can be expanded in several ways. The first is to ask about bonding between individuals who are already in a relationship. That is, the bonding in my model occurs only once at the beginning, and the subsequent information about the intention of your partner is revealed by their moves in the game. However, many social species continue to bond, albeit often at a reduced intensity, after the bond has been made.

A second extension, and perhaps one more appropriate to the warfare example, is multi-player bonding. This is likely to be important for humans in particular, since we engage in many multi-lateral cooperative ventures, and so knowing the intentions of the individuals around you is likely to be important.

CHAPTER 6

THE EVOLUTION LARGE-SCALE COOPERATION IN HUMANS

6.1 Introduction

Humans act in concert. Acting alone, we are little more inventive or industrious than our ape cousins. Though we are probably more individually intelligent, it is our ability to act concertedly in groups that sets us apart from other social primates. Read [1958] highlights this in a wonderful article where he asks us to consider all of the processes involved in constructing the simplest human artifact — the pencil. He claims that the pencil is not the product of individual human intelligence, but of concerted human effort, arguing that even given an entire lifetime, an individual would be unable to construct a single pencil, because doing so would require detailed knowledge of logging techniques for the wood, graphite manufacture and rubber manipulation. They would also need to construct the tools necessary to extract the raw materials, as well as be able to feed themselves in the mean time. Such a thing would be practically impossible.

Wilson [2012] draws a parallel between human cooperation in groups and ants cooperating within a colony, noting the similarity in the ways that both species divide their labour in order to achieve higher ends. This analogy between ant colonies and human cooperation has long excited evolutionary theorists, because it points to a theory of human behaviour that is not based on the individual, but rather on the collective [Bowles and Gintis, 2011, Fehr et al., 2002, Gintis, 2000, Sober and Wilson, 1999, Gintis et al., 2003].

However, most evolutionary theorists now agree humans are not genetically altruistic in the same way that ants are [Boyd et al., 2010, West et al., 2008, 2011]. When ants cooperate, they do so only with those who are closely genetically related, meaning that it is easy for them to develop adaptations to sacrifice themselves individually for the benefit of the colony. But for humans this is not so. Humans cooperate in groups with those who are genetically unrelated. Bowles [2006] estimates that relatedness in human hunter-gather societies is around seven percent, meaning that an adaptation that benefits another individual picked randomly from the population by an amount b at a cost c, then in order for this to spread, it would have to confer a fitness benefit that is more than fourteen times larger than the cost. Because the theory of natural selection tells us that any behaviour that consistently leads to the lower fitness of the individual who performed it, if that fitness cost is neither paid back to the individual at a later time, nor raises the fitness of that individual's kin, it will be driven out of the population by alternatives that do not exhibit the maladaptive behaviour. It is therefore highly implausible that we have developed adaptations to sacrifice ourselves for the benefit of others in this way. So when an individual works to aid another individual or individuals within a group, that work must, in some way, lead to a benefit to the individual who performed it.

Scholars have put forwards a number of theories explaining how cooperation within groups can be individually beneficial. The basic idea of these models is to extend the consequences of cooperating beyond their immediate effect. We have already seen the examples of costly punishment [Boyd et al., 2010] and indirect reciprocity [Panchanathan and Boyd, 2004] in sections 4.4 and 4.5 of chapter 4. Both theories claim that individual humans are adapted to punish bad behaviour, either through direct acts of corporal punishment, or through social exclusion in indirect networks of reciprocity.

But what exactly is cooperation? I identify at least two distinct uses of the term. The first is simply what we might call the 'natural use' version of the word. This is what we say when observe humans jointly engaging on some task. To separate this from other versions of the word, I henceforth refer to it as 'large-scale cooperation,' which also highlights the fact that, in this chapter, we are interested in explaining cooperation that includes many individuals. But as well as this, there is the theoretical version of the word, which is used to mean 'multi-player prisoner's dilemma.' For example, in the title of Panchanathan and Boyd [2004]'s paper 'Indirect reciprocity can stabilize cooperation without the second-order free rider problem,' 'cooperation' refers to the multi-player prisoner's dilemma. This I call the 'public goods' version of cooperation, since the model is used to represent public goods. The public goods model assumes that cooperators face the dilemma between providing goods for the benefit of the entire group, or for none at all.

My claim in this chapter is that the public goods game is not a suitable model for the evolution of large-scale cooperation. I argue that the model of public goods was a response to the theory that human cooperation rests on the mechanism of dyadic reciprocal altruism. While I agree that dyadic reciprocation is an inappropriate model for large-scale cooperation, I do not believe public goods are the correct replacement.

Instead, I argue that human cooperation often includes excludable goods.

That is not to say that public goods don't exist — clearly they do — but I argue that they are much less important than other forms of cooperation, and in particular, excludable cooperation. I am not the first to notice that there is a mismatch between the assumptions of the public goods model and the reality of large-scale cooperation. For example Trivers [2006] says the following:

> Boyd and Richerson [1988] claim that their 'model closely resembles models of reciprocity in pairs' and from it, one can safely conclude that 'the conditions necessary for the evolution of reciprocity become extremely restrictive as group size increases.' In fact, their model does not model reciprocity within pairs and their pessimism is entirely a result of artificial assumptions at the outset, primarily that 'If an individual switches from defection to cooperation every other member of the group is better off.' ... But why insist that cooperation between two must benefit all? If I groom you (or feed you or protect you), why does everyone else have to benefit? Boyd and Richerson [1988] are pretending to model dyadic relations but have forced on such interactions a large group effect, which inevitably grows in power with group size, they have created a world in which defectors automatically benefit from the trading of favours between reciprocators. This is an unnatural assumption. [pp.75 Trivers, 2006]

But why these assumptions? I believe that it is because they perfectly encapsulate a view held many scholars that humans have evolved moral adaptations for 'the good of the group' [Bowles and Gintis, 2011, Fehr et al., 2002, Gintis, 2000, Sober and Wilson, 1999, Gintis et al., 2003, Wilson, 2012, Wilson and Sober, 1994]. The view expressed by these scholars is that there is a zero-sum relationship between behaviours that are good for the individual and those that are good for the group, and the purpose of this model is to show how humans have evolved to perform the former against the latter. In evolutionary biology, 'good' is measured by fitness, meaning the claim is that humans have become genetically adapted to perform self-sacrificing behaviours for the exclusive good of the group.

While the public goods model remains widely accepted in the theoretical literature, up until very recently, there has been relatively little effort to relate these models to anthropological data. As it stands, there are only two consistently cited cases — that of food-sharing in hunter-gatherers and pre-state warfare [Henrich et al., 2005, Boyd et al., 2011, Boyd and Richerson, 1988, Gintis et al., 2003, Bowles, 2006]. Of these two examples, I have already argued that food-sharing is only a public good if we assume that food is entirely non-excludable, which is not generally what we observe. Warfare *may* represent a public good that is non-excludable, but as it stands there is no definitive evidence that humans took part in large-scale warfare in our evolutionary past [Fry, 2013].

The public goods assumption rules out what I believe to be the most common incentive that humans have found to motivate individuals to provide services to the group — exclusion from the benefits. I argue that the reward for cooperating in large groups often comes in the form of economic goods that are 'paid' to individuals. Payment involves transferring some property that previously belonged to the group into the private possession of its members. While its members ultimately control property owned by the group, it is wrong to suppose that specific individuals own it at any given time. Rather, rules about who can use property and when are determined by some form of institutional structures based on some local arrangement. For example, in our modern economy, we have companies who pay workers. The process of transferring group property into the hands of the individual workers is normally ritualised, with only certain individuals able to do this, at certain times, and only by invoking their institutional power.

I have already argued that this method is employed in food-sharing in chapter 3. I extend these insights by arguing that the institutions that are formed by modern day societies are quite similar. With respect to warfare, my model is not appropriate for hunter-gatherers where there are no excludable goods to steal. However, there is not a great deal of evidence that much warfare took place during this period of our evolutionary history. Here I focus my model on explaining examples of warfare where there is some economic benefit, which is very common [Fry, 2013, Keeley, 1996, LeBlanc, 2003]. To highlight this, I apply the model to the example of the cattle raiding in the Turkana — an empirical example hitherto claimed by the public good theorists — and argue that it is motivated primarily by direct economic reasons. This demonstrates that humans do not need to have been group selected in order to go to war when warfare is used as an economic activity. Together, these two accounts undermine both of the examples previously used to support the public goods theory of large-scale human cooperation.

6.1.1 My Contribution

I review the theoretical literature on large-scale cooperation in humans. I first argue that the dyadic model of reciprocal altruism is inappropriate for large-scale cooperation because it fails to capture the interdependence of actions that is required for large-scale cooperation. Next I review how this model has been extended into the multi-player prisoner's dilemma, also called the public goods game, which attempts to capture the interdependence left out of the dyadic model. I argue that while this model represents a significant step towards understanding how large-scale cooperation evolved, it is ultimately limited by the assumption that the benefits of cooperation are unexcludable.

The model of the multi-player prisoner dilemma rules out what I argue is the most important type of punishment that humans have devised to facilitate large-scale cooperation — exclusion from the benefits that are produced. Building on the assumptions of the models in the previous chapters, I argue that the cultural innovation of property rights is one of the most important social contracts for facilitating human cooperation. To represent this theoretically, I extend the model of the prisoner's dilemma into the multi-player case, introducing a character ε which represents the *excludability* of the good that is produced. Although this is based on the bourgeois model of hawk-dove introduced in section 3, I treat this as an exogenous factor to this model, by abstracting all of its properties into the term ε.

Finally, I investigate the empirical claims of the evolution of cooperation literature — food-sharing norms in hunter-gatherers, and pre-state warfare. I have already argued for the property rights that have been observed in food-sharing practises, which I do not repeat in this chapter, but I extend the discussion to show how this model of large-scale cooperation can be used to help us understand certain aspects of the behaviour.

In this section, I do the following

1. Review the literature on large-scale cooperation.
2. Argue that the assumptions of the prisoner's dilemma are inappropriate for a great deal of human cooperation.
3. Introduce a new model that I call excludable cooperation, based on the assumptions of private property in the previous sections.
4. Investigate the empirical evidence of cooperation in both the examples of pre-state warfare and food-sharing norms.

6.2 Large-Scale Cooperation

Humans are remarkably flexible in the ways we have invented to cooperate; our modern day economy testifies to this fact. From a theoretical perspective, one consequence of this flexibility is that it is difficult to construct a general model of cooperation that describes all examples. As a typical rule, the more general the model, the less it says about any particular example, and the more detailed the model, the narrower its application.

Most of the traditional models of cooperation assume only that dyadic pairs engage in cooperative activities at any given time [Trivers, 1971, Axelrod and Hamilton, 1981, Binmore, 2005b, 1998, 1997]. While these models are excellent for understanding the dynamics of relationships between pairs of individuals, they are not particularly helpful for thinking about large-scale cooperation, which requires many individuals to act simultaneously. This is because large-scale human cooperation is more than a series of dyadic interactions, bilaterally negotiated without regard to others around them. Rather, interactions are structured and aimed towards producing some definite goal. For example, most modern day corporations produce products to sell on the market. All of the workers within these corporations cooperate towards the goal of producing and selling the product. Although dyadic interactions are part of this process, there is clearly more to the phenomena than this. We want to say that the corporation has some goal, and, if it is working well, that each individual's action within the group helps bring about that end.

Some scholars have tackled the problem of large-scale human cooperation head on by modelling it as a game that include n individuals. The most sophisticated of these models adopts assumptions that abstract away the various responsibilities that exists within human groups, and instead assumes a 'group goal' — a project that, presumably through some process of negotiation and consensus building, the members of the group decide they will conduct — and that every individual within the group has 'job' — a responsibility that helps in the completion of the project [Boyd and Richerson, 1988, Panchanathan and Boyd, 2004, Nowak and Sigmund, 2005]. Since this model further assumes that each job is of equal importance, it assumes that each individual is faced by the dilemma of whether to voluntarily pay a fixed personal cost, normally denoted c, to bring a fixed benefit to the group project, normally b.

The advantage of this model derives from the fact that it potentially offers a way of describing all large-scale cooperation in a way that is independent of any specific set of structural relationships. Just as how, in evolutionary terms, we might abstract the action of an individual ant as being beneficial to the nest, without being explicit about what it does, so too can we assume that an action of an individual benefits some group project without knowing the details. But we must remember why we are making this assumption — because we are interested in understanding *why* humans choose to behave cooperatively. The question is about ultimate motivation — exactly what it is that humans do is immaterial to the model.

However, while on the surface, this multi-player model of cooperation seems to accurately represent what we empirically call large-scale cooperation, it includes one assumption that I believe to be indefensible. This is the assumption that large-scale human cooperation is necessarily a *public good*. This assumes that when cooperation is performed, it is performed to the unexcludable benefit of the group. This means that in this model, there is no possible way for the benefit to be excluded from a particular individual or subset of individuals. To continue the example of the modern day corporation, it would mean assuming that if a worker did not turn up to work, that either the corporation would have to keep paying their wages, or that the entire enterprise shut down. These are assumptions that seem to represent the extreme cases of cooperation, and do not chime with what we might expect from a model. But why were they chosen? In the following section I investigate this question.

6.2.1 Altruism, Group Selection and Strong Reciprocity

Why do scholars assume that the benefits from large-scale cooperation must go to all in the group? I believe that the answer to this lies in the long standing debate between the *individual* and the *group selectionists* in biology, and more generally in philosophical debates between individualist and collectivist moral philosophies. This is evident from the fact that the main feature of the public goods model explains how it is possible for 'altruism' to exist in nature. It is generally understood that the first part of this debate was settled by Williams [1966] in response to the theory of Wynne-Edwards [1962]. Wynne-Edwards argues that a member of a species might voluntarily choose to lower the number of offspring that it has during times of environmental degradation, in order for that environment to grow back. The argument is that groups that adopted this strategy, or more precisely, a group whose individual members adopted it, would perform better than the groups who did not. In response to this, Williams pointed out that such an adaptation would likely not occur since the individuals within the groups who continued to reproduce normally would have an immediate and direct advantage, and therefore this strategy would not prevail. Notice that the model for this strategy is the multi-player prisoner's dilemma — every individual can voluntarily choose to produce fewer offspring, thereby creating a 'group good,' which is unexcludable to any members of that group.

Since then, there has been a long back-and-forth between scholars arguing that various different models of group selection might be possible [West et al., 2008, 2011, Dawkins, 1979, Smith, 1964, Wilson, 1975, Wade, 1985, Price, 1972], with the biggest breakthrough coming from Hamilton [1964] who found that biological altruism was possible if the recipient of the altruistic act was sufficiently well related to the donor.

In many scholar's minds, this vindicates 'group selection,' bringing it into the fold of 'selfish gene theory' that was outlined Williams [1966] and elaborated (and named) by Dawkins [1976].

Another way that altruism has been found to be possible is that if, as a consequence of the altruistic act, the cost was somehow reimbursed to the donor. This idea was first outlined by Trivers [1971], who coined the expression 'reciprocal altruism' to explain it. The model was later formalised by Axelrod and Hamilton [1981] as a repeated prisoner's dilemma. The idea is simple. Because interactions are repeated, an individual can punish a cheater by defecting against them in future interactions. When the right conditions hold, this threat of withdrawing future cooperation can suffice to maintain reciprocal altruism at equilibrium. This idea is extended by the theory of indirect reciprocity, which is also aimed at explaining why an individual might evolve to do good to another individual when they do not get any good directly back from them — because *other* individuals do good in return.

However, some scholars are hostile to these theories of cooperation when applied to humans [Henrich et al., 2005, Boyd et al., 2011, Boyd and Richerson, 1988, Gintis et al., 2003, Bowles, 2006, Wilson, 2012, Wilson and Sober, 1994, Herrmann et al., 2008, Bernhard et al., 2006, Fowler et al., 2005, Egas and Riedl, 2008, Marlowe et al., 2010, Chaudhuri, 2011]. This group are united by the idea that humans are in some way 'super' altruistic, and that this altruism is not represented by the theories of relatedness or of reciprocation.This group argues that humans are *strong reciprocators*. A strong reciprocator is an individual who is willing [Fehr et al., 2002]

1. to sacrifice resources to be kind to those who are being kind (= strong positive reciprocity) and

2. to sacrifice resources to punish those who are being unkind (= strong negative reciprocity).

They also argue they have demonstrated this to be true via experimental evidence from a series of economic experiments that have been conducted [Camerer and Fehr, 2006, Fehr and Schmidt, 1999, Fehr and Gächter, 2002, Gintis et al., 2003, Henrich et al., 2005].

Here I make two claims about this position. Firstly, that the experimental evidence of human altruism is inconclusive. Secondly, that the theoretical foundations of strong reciprocity are inconsistent, and that in fact, it is no different from 'weak reciprocity,' that is the name that this group has given to traditional theories of cooperation [Gintis et al., 2008].

Experimental evidence for strong reciprocity has been criticised from a number of sources. Binmore [2010] criticises it for not being able to distinguish between a 'means' and an 'end.' For example, strong reciprocity theorists argue that:

> Our most important finding... is that most individuals treat moral values as ends in themselves, not merely means toward maintaining a valuable social reputation or otherwise advancing their self-interested goals. [p. 6 Gintis et al., 2008]

Since I have discussed at length the difference between a 'means' and an 'end,' or what I call a 'taste' and a 'value' in section 2.4.1 I will not repeat the full argument here. In short, Binmore (and I) argue that individual humans are conditioned by social norms to respond in certain ways. We call these our values. When individuals engage in laboratory experiments, they do not leave their values at the door, but remain conditioned to use them, at least initially. During the game, when the strategy is seen not to pay off, individuals tend to head towards an evolutionary stable strategy, which is normally set up to be defection in these games [Binmore and Shaked, 2010, Binmore, 2010, 2007, 1998, 1997]. Marsh and Kacelnik [2002] demonstrates this point by showing that calling the game by a different name in an experimental setting significantly alters their behaviour.

The relationship between tastes and values are that tastes are what an individual would want if there were no social interactions constraining behaviour. For example, we might observe a slave working (their value), but we would not conclude from this that the slave preferred to work as a slave (their taste), but that there were social pressures causing them to adopt this strategy. Were we to measure a slave's behaviour under laboratory conditions, we would expect that, at least initially, they would continue to play according to their values. What is more, because there are so many different values, we are never sure which one we have triggered in any given case.

Burton-Chellew and West [2013] produce a number of experiments that undermine the basic conclusion that human behaviour under experimental conditions has anything to do with pro-social preferences and strong reciprocity. In this paper, they mimic the experimental design of the behavioural economists to produce three results that undermine the conclusions of strong reciprocity theory. Their first experiment was to test whether humans behaved differently in experiments when they knew they were against other humans or playing against a computer. They found that knowing that they played against a human rather than a computer made little difference to the strategy they played. In a second experiment, they demonstrate that increasing the participant's awareness of the benefits to others from their cooperation leads to a decrease, rather than an increase, in the levels of cooperation. Lastly, in another game, they made cooperation not only best for their partner, but also best for the individual, and they still found that individuals reached nowhere near one hundred percent cooperation, but rather, tended to remain at between fifty and seventy five percent. They conclude that they believe that individuals who play this game generally are trying to maximise their financial payoffs, but that they are doing so based on uncertainty or bad beliefs. I argue that the 'bad beliefs' are likely to be the values of the individuals who play the game, shaped by their internalised understanding of the social contract

Guala [2012a] expresses doubt that the choice of behavioural experiments reflects real social dilemmas that individuals face, or ever did face in our evolutionary history. In particular, he argues that there is virtually no evidence of any third-party corporal punishment in the anthropological literature. In response to this, Gintis and Fehr [2012] and Gächter [2012] argue that, at equilibrium, we would expect to see little evidence of punishment because it is so effective. Guala [2012b] replies that it is unfair to argue both an empirical demonstration *and* the lack of empirical evidence can simultaneously be given as evidence of its existence.

Guala [2012b] argues that rather than relying on decentralised acts of violence to ensure cooperation, that human populations enact systems of coordinated punishment. Here he draws on the work of Ostrom [1990], Ostrom et al. [1992], who argue that populations set up punishing institutions that do not rely on individual 'willingness' to punish. He offers the example of the *Carte di Regola*, or 'charters:' ancient written codes used by communities in the Trentino region in the northeast of Italy to regulate the exploitation of common pastures. They were progressively introduced from 1200 until 1800, when they were eventually abolished by Napoleon. The charters were adopted by single villages rather than imposed from above, and were aimed at preventing the over-exploitation of communal fields — a specific instance of a prisoner's dilemma. A function of charters was to set up and regulate the monitoring inside and outside users of the fields. The monitoring system was organised by the community and involved designated guards who could impose fines on free-riders. The guards did not inflict physical punishment, and were incentivised by retaining a third of the fine. Reports of transgressions by community members were also incentivised in a similar way. Instead of letting the punishers bear the cost of monitoring, the carte thus introduced mechanisms that alleviated the costs, and even made sanctioning a lucrative activity.

This is an example of an application of the social contract of property rights discussed in chapters 3 and 4. Recall that in the previous section, the 'guard' is not a specific individual within the population, but is the victim of the transgression. But here we see an innovation on this idea — rather than the victim of a transgression who must punish the perpetrator, it is a specific individuals who take the fine. This individual has the same incentive as the victim in section 4, in that they personally profit from policing the transgression. When this norm is working, no player has the incentive to deviate — the farmer does not want to fight the guard since the guard is defending his new property, but the guard does not want to take the goods when the farmer has not transgressed for the same reason.

Throughout this dissertation, I have argued that it is misleading to base theories of human cooperation on 'spontaneous order' based on our preferences, and that instead we should focus on how particular social contracts solve specific cultural problems. Therefore like Guala, I see no reason to assume that there is one universal mode of punishment written into our genes. Is it true that costly punishment has been adopted from time to time. Undoubtedly. For example in section 4.6, I argue that it is the basic model of how a medieval king retains his power — through the implied threat of force from all of the other lords if an individual disobeys him. Likewise, many military units operate this way — individuals that desert are shot. Leaders who fail to execute deserters are also shot. But these are far from universal behaviours. An individual who does not turn up to work is not flogged or killed for their transgression — they are simply not paid. If a friend stops doing favours for us then we stop doing favours for them in the future and so on.

6.2.2 Is Strong Reciprocity Different to Weak Reciprocity?

Regardless of the question of empirical support, the theoretical position of strong reciprocity is inconsistent. Here I show that it is no different from weak reciprocity. The two theories follow precisely the same logic. In defining the theory of *strong reciprocity*, the authors state the following:

> Strong reciprocity (is the) propensity, in the context of a shared social task, to cooperate with others similarly disposed, even at personal cost, and a willingness to punish those who violate cooperative norms, even when punishing is personally costly. We deem this 'reciprocity' because it embraces an ethic of treating others as they treat us, bestowing favours on those who cooperate with us, and punishing those who take advantage of our largesse. We call this reciprocity 'strong' to distinguish it from forms of reciprocity, such as tit-for-tat and reciprocal altruism, that are the forms of long-run enlightened self-interest. [pp.3 Gintis et al., 2008]

What they call a 'shared social task' I have called 'concerted action,' but they both amount to the same thing.

What is unclear about this passage is what meant by 'long-run self interest.' Since most of these scholars seem to agree that relatedness in human groups is not high enough to foster human altruistic behaviour [Bowles, 2006, Boyd et al., 2010], then it seems there is no other choice. If a behaviour is individually costly, then it must logically lead to some fitness advantage somewhere down the line, that is the basis of all adaptive theory. What is odd is that strong reciprocity theorists agree with this sentiment:

> (W)e are often interpreted as rejecting the 'gene-centered' approach to modeling human behavior. In fact, our results in no way contradict the standard population biology approach to genetic and cultural change. A gene that promotes self-sacrifice on behalf of others will die out unless those helped carry the mutant gene or otherwise promote its spread. In a population without structured social interactions of agents, behaviors of the type found in our experiments and depicted in our models could not have evolved.

But here is the crux — if 'a gene that promotes self-sacrifice on behalf of others will die out unless those helped carry the mutant gene or otherwise promote its spread,' then how is strong reciprocity any different from weak reciprocity? There are only two ways that a behaviour can be selected. Either, it benefits the individual who performed the action, or it benefits others that have the gene. If relatedness in humans is too low for human cooperation to consistently benefit others with the gene, then it must ultimately benefit the performer, and so it is exactly the same as weak reciprocity.

I must therefore conclude that strong reciprocity is not theoretically distinct from weak reciprocity at all, but this leaves the question of why it is promoted. I believe that a clue to the reasoning can be found in the following passage:

> We wish to avoid three common misunderstandings of our argument. First, many contemporary researchers reject our critique of Dawkins, Alexander, and others in the 'selfish gene' school by asserting that their pronouncements should not be taken at face value. Rather, they say, references to phenotypic behavior as 'selfish' should be understood as asserting that the underlying genetic structures are subject to Darwinian evolutionary forces. Yet, these authors understood that their assertions were likely to be taken at face value, rather than being dramatic circumlocutions expressing completely unexceptionable propositions. It is only plausible, then, to suggest that they meant them, that it was plausible at the time to make such statements, but that they are now seen to be incorrect.

But this claim is indefensible. Scholars such as Dawkins and others go to great length to explain that selfish gene theory does not imply selfish phenotypic behaviour in the moral sense, only that genes must benefit themselves. There were no 'dramatic circumlocutions' at work — Dawkins was quite clear that his position was supposed to represent an unexceptional proposition, an invariant of natural selection. We are left to conclude that the only thing that this discussion leads to is the conclusion that Dawkins perhaps should not have used the word 'selfish' in the title of his book lest people get the wrong idea[1].

[1] Dawkins' book [Dawkins, 1999] explains this.

To sum, weak reciprocity is the theory that population have evolved to punish anti-social behaviour by excluding individuals from cooperation. Strong reciprocity is the theory that, during shared social tasks, humans have evolved to punish anti-social behaviour by acts of corporal punishment. The idea that one is more 'selfish' than the other is incorrect — both theories lead to a fitness advantage to the individual who performs the action, albeit through slightly different routes.

6.2.3 Large-Scale Cooperation and Indirect Reciprocity

There are models of human cooperation other than dyadic reciprocation and public goods. One such theory, *indirect reciprocity*, was first introduced by Alexander [1987]. Alexander argued that human morality has not evolved to be strictly dyadic reciprocation, but rather, that reciprocation is indirect. The basic idea is that an individual does not do a favour because they believe that the same individual will return the favour, but because some other individual will return the favour. This has sparked a great deal of theoretical work to discover whether these conditions could hold in an evolutionary system [Boyd and Richerson, 1989, Nowak and Sigmund, 1998, 2005].

The idea of indirect reciprocity is simple, and since I have already dealt with it previously in this dissertation in section 4.5, I only briefly recount the salient details here. Indirect reciprocity assumes that individuals within a population meet randomly with one another to play a donation game [Boyd and Richerson, 1989, Nowak and Sigmund, 1998, 2005], which consists of one individual being offered the chance to pay a cost c to bring a benefit of b to that player. In order to protect themselves from invaders, the population labels players with either a 'good' or 'bad' reputation, and adopts the following rule *cooperate with good players, defect against bad ones.* The rule for whether a player gains a good or bad reputation is as follows: *give a good reputation to anyone who cooperates with a good individual, give a bad reputation to anyone who defects against a good reputation or cooperates with a bad reputation.* This effectively means that an individual is 'buying' a good reputation with their cooperation, which benefits them greater than the cost of doing so.

Does indirect reciprocity explain large-scale cooperation? It does seem to capture some of the properties that we want. Most importantly, it is a model containing $n > 2$ players, where the relationships are not bilaterally negotiated. But this model still fails to capture any of the structural relationships that exist between individuals in cooperative groups.

To give an example, take the Turkana raiding party that I will return to in greater detail later on. The Turkana are a population who find it relatively easy to create war-bands with the purpose of rustling cattle from their neighbours. The group splits into two, the older men usually form a skirmish line in order to block the defenders, while the younger men rush in and herd the cattle away. They then meet up at a pre-arranged spot to divide up the cattle, of which the older men usually get more. It is quite plain that this behaviour cannot be modelled as a series of direct, bilaterally negotiated arrangements between pairs of individuals, but neither does it follow the logic of indirect reciprocity. Individuals are not doing 'favours' for other individuals at all, but are doing 'services to the group.'

That is not to say that indirect reciprocity never occurs. But it is better thought of as a model of insurance, rather than as a model of large-scale cooperation, because it loses the idea that human cooperation can be planned.

6.2.4 Large-Scale Cooperation and Property Rights

Now that I have reviewed the literature on large-scale cooperation and pointed out what I believe to be its problems, I want to offer an alternative explanation. Before doing so, I briefly recap what it is that we want to explain. Humans cooperate in groups. Within these groups, we divide our labour and responsibilities. The roles are generally selected because they bring about some 'group benefit.' For example, with the cattle rustling in the Turkana, the two different jobs brought about the beneficial outcome. From an evolutionary perspective, the question that we wish to answer is *why do humans cooperate in their allotted roles?* Notice that this question is independent of the actual details of the roles and responsibilities they will take. That is, in real life, we can apply this model to any example of cooperation no matter how complex the actual set of interactions are. Of course ultimately we would like a theory that also includes how these roles are negotiated in the first place, but this is beyond the scope of this dissertation, though I do outline some basic ideas in chapter 7.

Therefore I make the following assumptions about large-scale cooperation. First, there is a task that will produce some benefit, that I call a 'good.' Each individual has a choice to unilaterally cooperate, or defect. I suppose that cooperating costs them c, while defecting costs them nothing. Here we can see that, by focusing on the question of motivation, we do not need to know the exact details of what each individual is expected to do. It may be that some individuals are expected to do more, meaning that the cost is higher. However, putting such details aside, I assume that the cost is equal for everyone. I also assume that each individual who cooperates creates a benefit of b/n, where n is the number of indiviudals within a group. Again, there is no reason to assume that this is strictly speaking true. It may be that the activities of some individuals are vastly more important than of others, but again, it suffices as first order approximation.

The standard model now assumes that the benefit is divided equally among all members of the group equally, regardless of their current or past behaviour. This is the assumption that what is produced is a public good. I challenge this for large-scale cooperation. Why must the good that is produced go to every individual within the population? Why not just to the individuals who produced it? In our modern-day examples of cooperation, such a thing seems obvious — those who do not turn up to work do not get paid. How about in the smaller-scale societies? Those who do not hunt do not get property rights over food. Those who do not go on raids do not get property rights over the cattle that are stolen.

So I offer an alternative to this model. Humans cooperate in large groups because they devise a set of social arrangements such that they expect to receive a portion of the goods produced. That is, they expect to receive property rights over a portion of the goods produced, just as if they had produced it themselves in the previous models.

There are some important issues that I need to clear up before we go on to model this idea formally. Firstly, just because I have modelled property rights as a 'bourgeois' strategy in the hawk-dove game does not mean that we consider that the person 'standing near to' a good is the owner as the model was originally conceived. Rather, the owner of the good is determined by the social norm that prevails within a given culture. Just as how food was modelled as the property of all the individuals who took part in hunting, so are the goods that a group produces in other types of cooperative activity owned by those who helped to produce them.

A second and important part of this model is that in human cooperation, the good that is produced does not pass into the private property of individuals immediately after it is produced. The zebra that is killed is not divided on the spot, but taken back to the village to divide up formally. One n*th* of the money taken in a shop is not immediately the property of the individual who produced it. Rather, the good is normally considered the property of 'the group' until such a time as it has been distributed. The distribution process is normally ritualised, with only specific individuals being able to 'create' the property at a specific time, and only in response to specific actions that have occurred previously. For example, in the Ache, the distribution process is handled by an older and respected individual. Among the Turkana, the property is only made when the group is safely away from their enemies. To say that the good is the property of the 'group' before this point does not make it a 'public good,' it merely means that no single individual controls it. It may be handled by an individual, meaning that they are responsible that no one steals it before hand. However, it would be impossible to devise a model of cooperation that included all of the various ways that individuals might divide up all of the responsibilities, and so instead, we simply subsume all of this under the 'cost' of having a responsibility.

The property rights themselves are held in place by the implicit threat of coalitional violence, as described in chapter 4. These individuals may or may not be those who have take part in the cooperative activity itself. Any given activity may only have involved a subset of the population, or perhaps none-at-all of your allies. While this is relatively unlikely, it is possible.

Notice that this is a model about motivation, and not about the actual structure of cooperation. It is possible to make the model more complex to introduce some of the details of a specific cooperative arrangement, but since our model is about 'general cooperation,' rather than any specific task.

6.3 A Model of Large-Scale Cooperation

In this section I introduce a theoretical model of the evolution of large-scale cooperation. I first review the existing public goods model, exploring the assumptions as I go along, before introducing my alternative.

The assumptions of the model of large-scale cooperation is that a group of individuals are trying to achieve some task. I assume that the task is divided into various 'jobs' that are distributed among the members. As discussed in the previous section, although the structural relationships between the actual jobs are probably quite complicated in real life, I assume this complexity away and suppose that each job the benefits the group project by the same amount. I further assume that when the project complete, it brings about some sort of benefit that is shared by members of the group. For reasons that I justify later, I also assume a parameter ε that can take any value from zero to one inclusive that represents the degree to which a good can be excluded from any given individual. An excludability $\varepsilon = 0$ indicates that a good is entirely private, so that individuals can be fully excluded from the benefits, and an excludability $\varepsilon = 1$ indicates that a good is entirely public. Intermediate values of ε represent partially private goods. We consider the excludability of a good a proportion of the good that can be costlessly withheld from another individual. In the current model we treat ε as an exogenous character, but later we show how it emerges as an equilibrium of a hawk-dove game.

In order to compare my theory with the current state-of-the-art models, I make the assumptions in the model as similar as possible to that of the seminal work of Boyd and Richerson [1988].

1. Form a migrant pool of infinite size as the population.

2. Selected at random from this migrant pool, individuals form groups of size $n+1$ (from the perspective of one of those individuals, we want there to be n others in the group for ease of calculation).

3. Each individual now selects who in the group they would like to exclude from their cooperation. On the first round in their groups they exclude nobody.

4. Each individual now chooses whether they will cooperate or defect. Cooperating means paying a cost c to give a benefit of b to every individual in the group they

decided not to exclude, and $(1-\varepsilon)b$ to every individual that they did choose to exclude. Defecting means that the focal individual pays no cost and no other player receives any benefit.

5. This game is repeated with probability δ (for a total expected period of $\frac{1}{1-\delta}$).
6. After all of the games have stopped, each individual adds up their payoffs to form their total payoff w. This represents their fitness.
7. New agents form a migrant pool and the process continues.

6.3.1 Public Goods

I now use this model to analyse public goods. A public good is a good that by necessity must be provisioned to every other individual in the group, or to none at all. To represent this in our model, we say that $\varepsilon = 0$. If we define σ_c as the type that always cooperates, and σ_d as the type that always defects, then we see that

$$\begin{aligned} w_{\sigma_c}(i) &= \frac{i}{n}b - c \\ w_{\sigma_d}(i) &= \frac{i}{n}b \end{aligned} \tag{6.1}$$

where i is the proportion of other individuals who play strategy σ_c in the group. To avoid confusion, we have assumed that the players do not provide public goods to themselves. It should be clear from the equations above that individuals who provide a public good do worse than those who do not provide the public good. This leads to the following proposition.

Proposition 23. *If $S = \{\sigma_d, \sigma_c\}$ and $\varepsilon = 0$, then σ_d is the only ESS.*

Proof. This follows trivially from the fact that $w_{\sigma_d}(n) = b > b - c = w_{\sigma_c}(n)$, and $w_{\sigma_d}(0) = 0 > c = w_{\sigma_c}(0)$. □

Boyd and Richerson [1988] show that if the public goods game was repeated sufficiently many times, then it is possible that certain reciprocal strategies could be stable in the population, or part of a polymorphic equilibrium with defection. Part of the problem here is in determining exactly what a reciprocal strategy should be. Boyd and Richerson choose to investigate a series of strategies they call *contingent cooperation* σ_{cc_a} which cooperates on the first round, and then continues to cooperate if a other individuals in the group also cooperated, otherwise they defect. a can be thought of as the 'tolerance' of the reciprocal strategy — how many defectors they tolerate until they switch to the defect strategy. They show that there are polymorphic equilibria containing σ_{cc_a} for all $a < n-1$ that is also resistant from invasion from any other σ_{cc_b} where $b \neq a$, and also that $\sigma_{cc_{n-1}}$, the least forgiving strategy, are evolutionarily stable if the game is repeated enough times. The problem with all of these reciprocal strategies is that when the groups get larger, the effect of reciprocation becomes weaker, meaning that in order to invade from rarity the relatedness needs to be unrealistically high.

This model extends reciprocation into the multi-lateral case, but in doing so it introduces the assumption that what is produced in an unexcludable public good. Consider what this model implies for cooperation — that in response to the presence of any single defector, each individual has to unilaterally decide whether to continue helping everyone including the defector, or stop working entirely.

6.3.2 Multi-lateral Cooperation

Because of the complexity of human cooperation, all models of it will be reductive. The public goods model represents a significant step in understanding how large-scale cooperation evolved in humans, because it offers a way to theoretically represent the phenomena, without having to explicitly model the details. However, to judge the usefulness of a model, we need to ask whether it sufficiently captures the salient interactions appropriate level of abstraction. I argue that the public goods model fails with respect to how it deals with defectors, and that the assumption that all individuals must necessarily continue to benefit from cooperation even when they continue to defect does not represent what we observe.

In the model of excludable cooperation that I introduce in this section, I assume that human populations are able to form social contracts that connect work to the benefits. The social contract that I focus on is precisely the ones that that I investigated in section 3.4.7. I argue that goods are produced and remain the property of the group. Then, at an arranged time, the contract converts group property into private property of those who have cooperated in the task. We do not model this process explicitly in the model, but exogenise as a variable that we call *excludability*. I assume that the excludability of a good is the amount of the good that can be excluded to defectors by the terms of the contract. Of course, the contract would not work if there was no reason for the individuals within the community to police it, but as we have seen in chapters 3 and 4, these contracts can be self-policing, meaning that when the property becomes an individual good, then it is no longer in any individual's interest to steal the property, since the cost of fighting is higher than the value of the good.

For simplicity, I assume that the strategy *cooperation* (σ_c) pays a cost c and gains a benefit of ib/n to where i is the number of other cooperators in the group. I assume further that the *defector* σ_d does not pays the cost to cooperate, and gains a reduced benefit of $i(1-\varepsilon)b/n$. Notice that if we reduce ε to 0, the model is of public goods as above.

The payoff functions that the different types receive when they are in groups containing i individuals playing σ_c are

$$\begin{aligned} w_{\sigma_c}(i) &= \frac{i}{n}b - c + \frac{\delta}{1-\delta}\frac{i}{n}(b-c) \\ w_{\sigma_d}(i) &= \frac{i}{n}b + \frac{i}{n}(1-\varepsilon)\frac{\delta b}{1-\delta} \end{aligned} \tag{6.2}$$

Proposition 24. *If $S = \{\sigma_d, \sigma_c\}$, then σ_d is an ESS, and if*

$$\varepsilon > \frac{c}{\delta b} \tag{6.3}$$

σ_c is also an ESS.

Proof. Let i be the number of individuals in the group playing strategy σ_c. To prove that σ_d is an ESS, we show that $w_{\sigma_d}(0) = 0 > -c = w_{\sigma_c}$. To prove that σ_c is also an ESS, we have to show that $w_{\sigma_c}(n) > w_{\sigma_d}(n)$, which follows from stipulation of equation 6.3 □

Thus, a set of conditions can exist whereby individuals in the group would individually do better by playing σ_c than by playing σ_d, so excludable cooperation is stable within the group.

Both the cases — where $\varepsilon = 0$ and where $\varepsilon = 1$ — could be considered extensions of the prisoner's dilemma in the multi-player context, each one based on a different set of assumptions. With this extension, we see that it is easy to explain cooperation by simply assuming that the cooperators exclude defectors from the benefits of cooperation. In a sense, this model is the more natural extension of reciprocal altruism, since both show that punishment by exclusion can solve the problem.

Where does ε come from? ε represents a population's ability to create social norms that link benefit to work. The primary method for achieving this is by constructing norms similar to those seen in chapters 3 and 4, where property rights are given in return for cooperation. In this model, the punisher is the social contract. The social contract is the arrangement that tells individuals who owns what. In modelling terms, it tells them how to behave in contests over resources by pointing them to an equilibrium. As we saw in chapter 3, there are many such equilibria in games of property. So in effect, it is the decision of the other players to follow the social contract that causes the punishment to occur. For a more complete version of this argument, refer back to the discussion in section 2.4.2.

Another important assumption is that the cost of cooperation is lowered after the first round. The reason for this is that I assume that individuals scale back the operation in line with the number of individuals who are willing to cooperate. For example, if the cooperative activity is cattle raiding, then there is no reason to assume that the population continue to raid in the same way if very few of the population turn up.

Rather, I assume they engage in some other less risky venture with lower costs and smaller benefits. Just as how in the first round of a repeated prisoner's dilemma is the opportunity to discover whether you partner is a cooperator, this same assumption can be carried over to the multi-player case. Of course in real life it may take several rounds to discover exactly who is cheating, but this does not significantly alter the model.

6.3.3 Invasion

The standard model of invasion assumes that there is a conditional probability of an =individual picked at random from the group as being the same type as the focal individual as being

$$m(\sigma|\sigma) = r + (1-r)p \tag{6.4}$$

where r is the measure of genetic relatedness and p is the proportion of σ's in the population. For invasion of σ, we assume that $\omega \approx 0$, so that $m(\sigma|\sigma) \approx r$. Therefore we assume that the distribution of individuals of the same type as the focal individual is equal to

$$M(\sigma) = \binom{n}{j} r^j (1-r)^{n-j} \tag{6.5}$$

We can therefore calculate the payoff for playing σ as being

$$W_\sigma = M(\sigma) w_\sigma(j) \tag{6.6}$$

If we take σ to be strategy of exclusion, then its payoff is equal to

$$\begin{aligned} W_{\sigma_e} &= \sum_{j=0}^{n} \binom{n}{j} r^j (1-r)^{n-j} \left(\frac{i}{n} b - c + \frac{\delta}{1-\delta} \frac{i}{n} (b-c) \right) \\ &= rb - c + \frac{\delta}{1-\delta} r (b-c) \end{aligned} \tag{6.7}$$

The fitness of the average defector is $W_{\sigma_d} = 0$ (since on average they are in groups with no cooperators). Therefore multi-lateral cooperation will invade from rarity whenever

$$r > \frac{1-\delta}{\delta} \frac{c}{b-c} \tag{6.8}$$

This result demonstrates the remarkable synergy between repeated interactions and associative group formation. When both are seen together, it is possible for strategies to invade that would not do so otherwise. In the example above, if meetings are entirely random ($r = 0$), then the payoff for cooperating is equal to $-c$. If, on the other hand, cooperation were not repeated ($\delta = 0$), then we need a very high relatedness between those who interact. Another important property of this model is that invasion is entirely independent of group size. Invasion for the public goods model was difficult when group size increased, but this is not true of our model.

6.3.4 Dynamics

We have seen the conditions under which cooperation will invade. But we can determine whether it has an internal unstable equilibria by equating the fitnesses of the two types. If we continue to assume that $m(\sigma|\sigma) = r+(1-r)p$ is the chance that another individual in the group is of type σ, given that the focal individual is of type σ and r is the degree of positive assortment, then we can calculate the fitness of strategy σ as

$$W_\sigma = \sum_{j=0}^{n} \binom{n}{j} (r + (1-r)\omega)^j (1 - (r + (1-r)\omega)^{n-j} w_\sigma(i) \tag{6.9}$$

Since both w_c and w_d are linear functions of i, we can write these fitness functions as

$$\begin{aligned} W_{\sigma_c} &= \frac{x}{n} b - c + \frac{\delta}{1-\delta} \frac{x}{n} (b - c) \\ W_{\sigma_d} &= \frac{x}{n} b + \frac{\delta}{1-\delta} (1-\varepsilon) \frac{x}{n} b \end{aligned} \tag{6.10}$$

where $x = r + (1-r)p$. Equating these we get $W_{\sigma_c} = W_{\sigma_d}$ when

$$p^* = \frac{cn(1-\delta)}{(1-r)(\varepsilon b - c)\delta} - \frac{r}{1-r} \tag{6.11}$$

p^* is the unstable equilibria, which, depending on the values of the other variables, may or may not fall between 0 and 1. We can also describe the change in frequency of type σ_c by the following equation

$$\bar{p}\Delta p = p(1-p) \left(-c + \frac{r + (1-r)p}{n} \frac{\delta}{1-\delta} (\varepsilon b - c) \right) \tag{6.12}$$

where $\bar{p} = p w_{\sigma_c} + (1-p) w_{\sigma_d}$. It is easy to see that p^* increases with c and decreases with b. It also decreases when δ gets larger and when r gets larger as we would expect.

6.4 Empirical Validation

In the previous section, I discuss two models of large-scale cooperation, both of which extend the model reciprocal altruism. The public goods model (when $\varepsilon = 0$) assumes that humans produce non-excludable public goods. My alternative model of excludable cooperation ($\varepsilon > 0$) assumed that part of the good is excludable. The question that I address in this section is which of these better represents large-scale cooperation in humans?

The first thing that I point out is that there are many examples of cooperation where the benefits are clearly excludable. Most of our modern day economy runs on the understanding that wages are paid, and wages are clearly an excludable good. Therefore I see little point in investigating these examples in detail. Rather, In this section, I investigate some examples that scholars positively claim must be represented by public goods.

In the literature, there are surprisingly few links to anthropological data. However, there are two paradigmatic cases of human cooperation that are claimed to be represented by the public goods model. These are food-sharing norms and pre-state warfare [Boyd and Richerson, 1988, Fehr and Fischbacher, 2003, Boyd et al., 2010, Boyd and Richerson, 2005, Gintis et al., 2003, Fehr et al., 2002, Boyd et al., 2003, Gintis, 2000, Gat, 2011, Bowles, 2009, Boyd and Richerson, 2005, Soltis et al., 1995]. I have already conducted a review of the food-sharing literature and so will not not repeat it here. However, I do demonstrate how the outstanding assumptions made in this section apply to the example of the Ache. With respect to pre-state warfare, I review the empirical literature and argue that, while there is a great deal of evidence that small-scale societies conduct warfare and did so in our history, there is a great deal of doubt over whether our hunter-gatherer ancestors did. My model is therefore more likely to represent later warfare such as large-scale raiding where there are some excludable benefits.

6.4.1 Pre-State Warfare

There is currently little consensus on the scope and importance of warfare in human evolutionary history [Ferguson, 2000, Fry, 2013, Ferguson, 2013, Pinker, 2011, Keeley, 1996, Chagnon, 1968, 1988, LeBlanc, 1999, 2007, Bowles, 2009, Choi and Bowles, 2007, LeBlanc, 2003]. Among the empirical challenges is the difficulty in making inferences from ethnographic accounts of of hunter-gatherers about conditions before the domestication of plants and animals, and the fact that hunter-gatherers made little or no use of fortifications, and used the same weapons to fight as they did to hunt, leaving few archaeological traces except skeletal remains.

There are two opposing sides to this debate. The first argue that war was ubiquitous in our species during human evolutionary history, and that it is a natural expression of evolved tendencies toward deadly violence against those outside the social group [Pinker, 2011, Bowles, 2009, Choi and Bowles, 2007, LeBlanc, 2003, Wrangham and Peterson, 1996]. This group argues that throughout evolutionary history, war casualties were sufficiently high to select for behavioural tendencies conferring reproductive advantage in inter-group competition. A second group argue against this position [Ferguson, 2000, Fry, 2013, Ferguson, 2013, Parker Pearson, 2005], claiming that there is little evidence that war was ubiquitous among our species in evolutionary pre-history.

With his book, Keeley [1996] sparked a great deal of interest in warfare in small-scale societies, asserting that war was ubiquitous in pre-history, and complaining of many scholars 'pacifying the past.' Although he draws most from the ethnographic literature, he does present nine archaeological cases, concluding that prehistoric war accounts for "from about 7 percent to as much as 40 percent of all deaths." Another foundational book is LeBlanc [2003]'s *Constant Battles.* Like Keeley, these scholars concludes that "every one had warfare in all time periods" LeBlanc [pp.8 2003], attributing war to the Malthusian tendency of population growth. Yet another important set of scholars are Bowles and his colleagues [Bowles and Gintis, 2011, Bowles, 2009, Choi and Bowles, 2007], who also argue that there was sufficient war in our evolutionary pre-history to explain how humans have evolved altruistic tendencies. In their book, Bowles and Gintis [2011] compile a list of a adult mortality due to warfare, finding deaths from warfare range from zero to forty six percent. Finally there is Pinker [2011]'s book *The Better Angels of Our Nature,* which opens with archaeological illustrations of violence in human past. He discusses the evolutionary logic behind the deadly competition, and claims that the archaeological evidence supports the claim humans naturally tend toward violence. Pinker's archaeological evidence combines the evidence given by Keeley [1996] and Bowles [2009], producing 21 prehistorical cases to calculate an average pre-historic death from warfare rate as fifteen percent [pp.8 Pinker, 2011]. As far as I know, this is the most comprehensive list of archaeological data of casualties warfare throughout human pre-history.

In response to these claim, Ferguson [2013], who is one of the scholars accused of 'pacifying the past,' responds that there is in fact very little evidence of widespread warfare in our evolutionary past. He argues that the examples that are listed by Pinker, and which are made up by the previous lists, are hand-picked and are by far the most violent examples. Of this list, he first excludes eight of the original twenty-one — five are eliminated because there was only zero or one instance of violent deaths, and three were dropped because they were duplicated. The fourteen that remain represent the most extreme examples, and are in no way representative of pre-historical warfare.

Ferguson argues that there is no evolved predisposition to go to war, but instead that humans have plastic behaviour (much as I have been arguing throughout this dissertation). Broadly speaking, he argues that war starts when those who start it believe that it is in their own self-interest. He also offers many examples of where the preconditions for war were all present and yet no warfare occurred.

Personally, I am dubious about theories of human behaviour like the 'general tendency to go to war,' and even more dubious about the linear relationship between the 'innate desire' to do something and the social contract being setup to achieve this. Like Ferguson, I believe that there is far more evidence that humans go to war when it serves their interests, and that it is too early to conclude that humans desire to go to war for war's sake. My hypothesis is that war is way of obtaining things. An individual may win resources, groups may win territory and so on. It is an example of large scale cooperation. So I argue that where necessary, humans have evolved to develop social contracts that, among other things, encourage individuals to fight. For example, the models of costly punishment and social ostracism reviewed in section 4 are possible means that humans have found to encourage warfare where the good is non-excludable, and my own model of exclusion can explain situations where the goods are excludable.

6.4.2 Warfare and Large-Scale Cooperation

Given that I have established that human cooperation can be explained by exclusion and repeated interactions, it is now our task to discover empirical observations of cooperation and see if they conform to the assumptions in the model. To do this, I once again draw attention to the assumptions of the model of exclusion presented in section 6.3.2.

1. **The provision is personally costly ($c > 0$)** For cooperation to be modelled as a public good, provisioning the good must be personally costly. This is not to say that the provision *involves some costs*, but that the direct personal costs of provisioning behaviour (i.e. going hunting/going to war) outweigh the direct personal benefits. If this is not the case, then direct selection is sufficient to explain the emergence of the behaviour and we do not require reciprocal altruism or group-wide cooperation to explain them.

2. **The goods produced are beneficial to the group ($b > 0$)** For cooperation to be modelled as a public good, the benefit that goes to other members of the group must be greater than 0, that is, they must be producing something that benefits the rest of the population.

3. **The good is not sufficiently excludable ($\varepsilon > c/\delta b$)** If the excludability of the good is lower than $c/\delta b$, then the portion of the activity that provides a private good is insufficient to explain the activity. Otherwise exclusion is a possible explanation.

Although warfare is one of only two examples of altruism, there is relatively little reference to the anthropological literature on warfare. For example, Choi and Bowles [2007] state that

> The ethnographic and archaeological record suggests that warfare was a frequent cause of death among some hunters-gatherer groups and early tribal societies. [p. 637 Choi and Bowles, 2007]

where they cite Keeley. As we have seen in the previous section, Keeley does find that a great many deaths are caused by warfare. However, this fact alone does not imply that warfare is necessarily a public good. As far as I know, the only study to question the assumptions explicitly is reported by Mathew and Boyd [2011], and is a detailed quantitative study of warfare among the Turkana — contemporary nomadic pastoralists who dwell in East Africa. Although Mathew and Boyd [2011] do not state explicitly that warfare is a public good among the Turkana, they refer to it as an example of the theory of costly punishment, discussed in this dissertation in section 4.4. By drawing attention to each assumption of the model, I argue that there is little evidence to support the assumptions of the public goods model, and far more evidence to support my model of excludable cooperation.

Before I go through the assumptions explicitly, I first give a general description of warfare among the Turkana. The Turkana are a large ethno-linguistic group, organised as a small-scale society. The population regularly forms warbands to perform incursions into eighbouring groups. These raids can be large, reaching around three hundred men. While there are some positive externalities to this raiding such as territorial expansion, the primary objective is to capture loot. This typically involves the division of labour — when the raiders surround a camp, some of them , typically the older and more experienced men, engage the residents in combat, while others, typically the less experienced younger men, drive away cattle and some engage in delaying actions, allowing their allies to flee to safety. If all goes according to plan, then once they reach safety, the cattle is distributed among the members depending on age.

Is warfare personally costly ($c > 0$)? There are high personal costs for engaging in warfare, with 32% of male-reproductive period mortality due to offensive warfare and a 1.3% chance of death per single force raid. However Mathew and Boyd [2011] also report that raids produce both direct and indirect benefits through personal loot (the explicit goal of raids) and social capital. They are unable to quantify whether engaging in warfare is overall net personally beneficial or net personally costly and, more importantly, whether the marginal effects of a small amount of recruitment avoidance/desertion/battlefield reluctance are individually beneficial or costly. Mathew and Boyd [2011] present field data on the individual returns from engaging in raids, force and stealth, and notes that the personally riskier force raids result in higher personal returns with a mean of 11 cattle looted per head in force raids and 3 in stealth raids. Dyson-Hudson et al. [1998] note that the Turkana are exceptionally dependent on their cattle. 92% of their food comes directly or indirectly from their cattle and they have an unusually high bride price compared with other East African ethnic groups with a mean of 43 cattle. The average age of male sexual maturity is 18, but the average age of male-female cohabitation is 32. The delay between initiating marriage negotiations and being in a position to hand over the bride price can be as long as 20 years. In that context it is entirely possible that the individual lifetime fitness cost of participating in force raids is smaller than the individual lifetime fitness benefits.

Does warfare produce public benefits ($b > 0$)? In contrast with the detailed data collected on individual returns, the data on collective returns rests is speculative: retaliatory raids may produce deterrence, large scale raids may affect population movement and grazing land availability. One informant states that he had been on a retaliatory raid on which lack of loot was not an issue, though it is unclear how common such raids are. Loot is not a public good — in fact, even loot-sharing distribution norms among the raid party are quoted as failing in 56% of raids. Since there is only speculative evidence that loot produces a public good, and a great deal of evidence that loot produces a private good, we believe it is a mistake to model it as a prisoner's dilemma.

How excludable are the benefits of warfare ($\varepsilon < c/\delta b$)? As noted above, the loot is not shared out within the ethnic group, it is not even shared out evenly among the warriors. Notice that there is an assumption here that once an individual takes ownership over cattle that they are able to defend them, therefore meaning that cattle do not become public goods. Furthermore, within the troop who raided, it is members of the local age group of a raider who take the primary role in assessing whether an act of shirking demands a response. In such situations, they are the ones who will administer any corporal punishment and they are the ones who will share and consume any 'fine' applied to the offender. These facts imply that the excludability of the goods is high within both the entire ethno-linguistic group, since they do not directly benefit, and also high among individuals who defected during the raids, since those that do not fight may have loot taken away from them.

In conclusion, I propose that this is an example of large-scale cooperation producing private, excludable goods. Warriors band together, rustle cattle from their neighbours, and divide it up along prearranged lines. If a public good is produced, then it produced incidentally — the primary reason these individuals go to war is to capture cattle. One final point that is worth considering is that, in this example, to work out whether a public good was produced, we had to couch the explanation in terms of other social norms. In this case, I assume that the Turkana are able to 'connect work to property' in much the same way that I assume in the model of property rights in section 3.4.7.

6.4.3 Food Sharing

Food sharing is the second widely cited example of a public good in the theoretical literature [Boyd and Richerson, 1988, Fehr and Fischbacher, 2003, Boyd et al., 2010, Boyd and Richerson, 2005, Gintis et al., 2003, Fehr et al., 2002, Boyd et al., 2003, Gintis, 2000]. However, explicit details of *how* food-sharing is a public good is scarce.

Gurven [2004a] provides an excellent review of the food sharing literature where the question of the degree to which food sharing is a public good is a central issue. For example, he states that

> Observing the extent of producer control is confounded by a lack of understanding of how distribution decisions are made in the context of the conflicting push and pull of interested parties. It is also confounded by the implicit assumptions that lack of control is signified by a hunter's receiving $1/n$, and that complete control is viewed as an ability to hoard 100% of a resource. However, keeping $1/n$ does not signify a lack of control if the acquirer decides that $1/n$ is the optimal portion to keep, given the expected payoffs from sharing. Even when hunters relinquish complete control of game, as among the Ache, such abandonment may be voluntary, as Ache do not relinquish control when at the reservation. [p. 548 Gurven, 2004a]

There are several common patterns of producer control in the ethnographic literature of food sharing. These include biased distributions, preferential shares to acquirers and their families, or more frequent sharing to close kin outside the nuclear family at the expense of more distant kin and unrelated individuals (see [Gurven, 2004a] for details). As Gurven points out, expectation of sharing is highest within camp, which opens up the opportunity for hunters to consume portions of their catch at the site where it was killed. For example, Ache hunters could bring their family members directly to the kill site to cook and consume meat. However, food is always transported voluntarily to the camp, indicating an intention to share with others voluntarily.

To apply our model of cooperation directly to a well studied population, I focus once again on the detailed studies on the Ache of Paraguay [Hill, 2002, Kaplan et al., 2009]. Recall that the Ache are a population where meat sharing is particularly wide, making it an excellent test for reciprocation theory. Game makes up 60–80% of the hunter's diet, with honey, fruit and other less calorie-dense food making up the rest. On average an individual will share 74% of the total calories that they collect, with meat being the most widely shared single food item, with an average of around 85% shared. Most importantly, in studies of the Ache, there is significant evidence that there is no direct contingency between giving and receiving between individuals within the same group, meaning that, within a hunting band, amount of food given at one time does not predict the likelihood of food being given in return.

Is food sharing personally costly ($c > 0$)? It is often assumed that the act of sharing food is net costly, since food not shared can be consumed by the hunter and his nuclear family. However there is also wide agreement that the marginal returns for consuming a catch decreases with the size of the catch. This fact is significant when compared to the highly variable returns from meat. For example, if a peccary (a type of pig) is caught, the number of calories provided can be around $50 - 60,000$, but the hunter would require only around of $10 - 20\%$ of these to feed himself and his family. Therefore, the marginal costs of giving away excess food, though positive, might be relatively low.

Does food sharing provide a benefit to the group ($b > 0$)? There is strong evidence that the benefit to food sharing is large. To see this, we once again think about the decreasing marginal returns of the fitness value of food. Receiving food every day is more valuable than receiving a large package of food, followed by a long period of no food at all. In this, food sharing is seen as a way of reducing the variance of the returns of food. There seems to no reason to doubt that receiving food when hungry provides large direct benefits, so we expect b to be relatively large.

Is food excludable from non-sharers ($c/b < \varepsilon$)? Food sharing occurs after the day's hunting. The food is distributed by a respected individual, usually an older male, within the company. This method of distribution immediately rules out food sharing as a pure public good; the distributer is capable of withholding food to a family or individual if they decide to. What is more, excluding a certain family in no way precludes giving food to any other family in the group. Indeed, we see that when food sharing is measured, food is *not* distributed evenly among all individuals present. A single food item is likely to be shared with around only 41% of the rest of the band. On 2 of the 9 trips recorded, at least one of the nuclear families would receive less than 5% of a catch. This means that, under these conditions, we would expect excludability to be high.

We are given a clear example that not only *can* food be withheld from another family or individual, it often *is*. Unfortunately the researchers never observed an adult individual refuse to hunt and still expect a fair share of the food, but they did observe teenagers refusing to hunt.

> Teenage boys who don't hunt are not guaranteed a share, but those who hunt seriously (i.e., all day long) receive an adult share. [Hill, 2002, p. 112]

Note that this is despite the fact that hunting success is highly contingent on experience and there is little difference between teenagers that do hunt and those that do not. Hunters are aware of who refused to hunt, and respond in kind by withdrawing cooperation from them. Since all individuals are together when the food is distributed, it is relatively hard to hold some back without it becoming common knowledge.[2]

[2] One interesting point to note is what occurred when the environment of the Ache changed. Gurven

To conclude, I have established that the costs of food sharing is likely to be relatively small in relation to the benefits of sharing ($b >> c$), and also established that it is possible that the excludability of food sharing is high. Taken together, this means that it is entirely plausible that $c/b << \varepsilon$. Thus here again we show that multi-lateral reciprocation can explain cooperation.

6.5 Discussion

In the opening chapter of this dissertation, I describe the evolution of cooperation literature as being primarily theoretically driven. What I mean by this is that the assumptions of many of the theoretical models are selected as a response to other theorists and not for how well they represent any particular example of cooperation. When we investigate these assumptions in detail, as I have done in this section, we find that they are difficult to justify, and that they paint a false picture of human cooperation. This is primarily why, throughout this dissertation, I have been diligent to justify each assumption in the models.

It is particularly important that the researcher highlights the assumptions that they are *least* confident with, since this helps focus empirical work. I believe that most work has hitherto simply pointed out the assumptions that concur with the model, and has not listed its weaknesses. For example, when modelling food sharing as a public good, the positive aspect is that food is shared widely. However, the weakest part of this model is the assumption that it *must* be shared equally.

One ideal assumption of this model is that cooperation brings an equal benefit to all members of the group. While this assumption makes the model theoretically tractable, it is the most difficult to defend empirically. It is not completely indefensible because, as we have seen, in our hunter-gatherer past, there is every indication that we were relatively egalitarian in our social organisation. However, we also know that subsequently, we have become much less so. This fact requires explanation, but it cannot be explained using the theory introduced in this chapter, since this theory already assumes the distribution to be equal.

The most important innovation of my model of excludable cooperation is that I allow other social norms to directly affect what it means to cooperate or defect. This allows the model to cover a broader range of human activities. For example, in Turkana

[2004a] studies a group of Ache that settled in Arroyo Bandera, which was founded in 1980. In 1998, there were 117 permanent residents comprising 23 nuclear family-based households living in Arroyo Bandera. One aspect of settled life is that the behaviour with respect to food sharing on the reserve differs dramatically from the behaviour in the forest. While on a hunting trip, direct contingency is *not* a significant predictor of the sharing of meat (though it *is* a significant predictor of sharing of non-meat item in the forest). However, on the reservation, contingency is found to be the *strongest* predictor of a family giving food to another. Our hypothesis for this change in behaviour is that while it is relatively easy to exclude cheats in the forest, it is relatively hard in the settlement, since there are too many behaviours to keep track of.

warfare, the jobs are either forming a skirmish line, or stealing cattle. If we simply focus on the raid, it seems as if the Turkana are producing a public good — each works for the good of the group project. However, by isolating part of the picture, this interpretation has ignored the effects of other social norms, in this case, the social norm that connects individual effort to the benefit of gaining property over cattle.

This philosophy of isolation and measuring one part of the social contract from the rest is one of the reasons why have I been resistant to the popular trend of using economic games as a means of quantitatively measuring behaviour. If we take the Turkana example, it is reasonable to suppose that, were we to play a public goods game with them, that they would interpret the game as warfare, and would therefore act as they do in real life. What is missing is the fact that, later on, the 'public good' is divided among the warriors. But whether or not such a system exists must be discovered by observing that particular culture.

6.6 Conclusion

I agree with the project of the evolution of cooperation scholars. I, like they, believe that simple models like dyadic reciprocation fail to capture the dynamics of how cooperation operates in large groups. Individuals do not interact in dyadic pairs and try to maximise their payoff relative to other pairs. I also agree with the method of investigating the motivations for cooperation from the perspective of genes trying to maximise their fitness. However, I disagree with the model of public goods and indirect reciprocity that these scholars propose as an alternative.

Instead, I argue that humans build social contracts in order to encourage pro-social behaviour. The type of social contract that I argue has been by far the most important, but is by no means the only one, are ones that connect 'work' to 'economic benefit,' and is generally achieved through turning 'group property' into 'individual property.' The model that I build to describe such cooperation I call 'excludable cooperation.' In this model, I assume that individuals are able to build social contracts that allow members to exclude those who did not take part from the benefits of cooperation, thereby creating an environment where cooperation is individually beneficial.

In this section, I argue that the assumptions of the public goods model were selected to contrast with that of models like dyadic reciprocation as much as they could, rather than being selected because they represent any particular set of empirically observed examples of cooperation. In trying to maintain the contrast between these public goods models, and models of dyadic reciprocation, the theory has become increasingly complex, and often the models are impenetrable to any newcomers. But I argue that the theoretical literature need not be complex in this way, and that we serve the scientific community better by producing simple models.

However, while I argue that models should be simple, I also argue that empirical work should be more sophisticated. The philosophy of economic games is to go out and test the theory directly by assuming that 'payoff' is equal to 'monetary rewards.' The problem is that the theory of cooperation is so abstract, that it is impossible to know what is being measured. So, as with all of the models in this dissertation, I have opted to build the assumptions of the models directly from observation and justify each one.

CHAPTER 7

DISCUSSION AND CONCLUSION

Each chapter of this dissertation includes its own conclusion and discussion. In this chapter, I summarise the main conclusions of each chapter, highlight some of the important themes that run through this work, and suggest some places for further research.

7.1 Chapter Summaries

In chapter 2, I describe the phenomena I wish to explain — the evolution of human cooperation — and outline my method of investigation. I argue that game theory can be used to help scholars understand the evolutionary motivations behind human actions. Humans act to achieve definite ends. But sometimes, individuals cannot all achieve their individual ends, and so plans must change in order to accommodate for others. Game theory is a way of describing how this system of partially conflicting desires can lead to stability, rather than chaos.

I also outline how I believe an evolutionary science of human social contracts ought to be done. I argue that it is important to recognise the fact that human social interactions are very often complex and indirect. Very often, when we observe humans acting, we are not so much interested in communicating the objective facts, but rather, what the individual meant by the action. Therefore, I advocate empirical work in natural settings, rather than in laboratories. In laboratories, we can observe closely how people act, but we cannot observe what they meant by those actions.

In chapter 3, I outline how I believe the first social contract was formed. I argue that this was to solve the problem of food-sharing. The chief problem of food-sharing is the question of who has power of the distribution of food. Previous to my work, there were two major positions — that the food is owned primarily by the hunter, or that it is a public good. I argue that, at least in some cases, neither of these positions seem to hold. Instead, the food is often owned by 'the group' rather than a specific individual within the group. But it is wrong to suppose that just because it is owned by the the group, that necessarily means that food it is a public good. Rather, I argue that being owned by the group means that its distribution is dependent on a set of social contracts. I argue that humans have evolved to create social contracts that motivate individuals to perform pro-social behaviours.

In chapter 4, I extend the model in chapter 3. I am primarily interested in explaining how groups design their contract to encourage some behaviours and prohibit others. I do this by building a number of models focusing on punishment. In particular, I theorise about the types of social contracts that our ancestors used to solve early public goods problems — namely, to defend large catches of meat against those who did not take part in the hunting. I also explain how a system of fines are commonly used to prohibit bad behaviour, and I show how this can be applied to understanding particular social contracts by using it to analyse a system of law among the Nuer.

In chapter 5, I investigate some strategies that humans and other social species have found to protect themselves against exploitation. In this model, I assume that individuals engage in social bonding — costly activity that is otherwise evolutionarily pointless. I argue that the purpose of bonding is to obtain information about the likely strategy of a particular partner in future interactions. I argue that this is important in humans because our social systems are constantly changing.

In chapter 6, I show how small-scale cooperation can evolve into large scale cooperation. The purpose of this model is to provide an empirically feasible alternative to the 'public goods' model of human cooperation, which I argue is inappropriate to explaining most examples of human cooperation. Instead, I argue that through systems similar to the ones outlined in chapters 3 and 4, that we have learnt to create social contracts that link actions that work for the 'good of the group' to some economic benefit. When cooperation becomes larger and more complex, such as with modern day large corporations or even states, this process is likely to become obscured. The central point here is that just because interactions are not reciprocal does not mean that they are necessarily public goods.

7.2 General Themes

In this section I outline some of the general themes of this dissertation.

7.2.1 Cooperation and Punishment are Often Features of Groups and Not Individuals?

Throughout this dissertation, I have been careful to delineate between individual and group strategies. The basic method is to explain all action as a function of the individual maximising their inclusive fitness, or, at the very least, maximising some property that closely relates to this value. However, in doing this, I believe that there is some danger of me being misunderstood for advocating a biological group selection framework, which I am not. Therefore, in this section, I explain clearly what I mean by cooperation and punishment being features of groups, and not of individuals.

All Darwinian models require three things [Williams, 1966, Dawkins, 1976, Hull, 2001, 1988]:

1. *Replicator* — A replicator is an entity that passes its structure on in successive generations.

2. *Interactor* — An interactor is an entity that interacts as a cohesive whole with its environment so that it causes differential replication.

3. *Lineage* — A lineage is an entity that persists indefinitely either in the same or an altered state as a result of replication.

In biology, the paradigmatic replicator is the gene, interactor the phenotype, and lineage the species. One important feature of this setup is that the 'evolutionary gene' does not need to be a 'gene' in the biological sense of the word. Rather, more generally, it simply needs to be information that is passed along to the following generation through some evolutionary process.

With respect to culture, the paradigmatic replicator is the 'meme' [Dawkins, 1976, Hull, 2001, Blackmore, 2000, El Mouden et al., 2014] — theoretical units of culture that exist in the heads of members of the population. Like the evolutionary gene, we do not so much care about the physical properties of the meme, just that it contains enough information to pass onto other individuals in such a way that it can affect their behaviour. This is the position that I have taken throughout this thesis — individuals have strategies that can evolve through the process of copying other successful strategies.

Some scholars argue that although evolution is unlikely to act on genetic groups, it does act on cultural ones [Henrich, 2004, Soltis et al., 1995, Boyd et al., 2003]. These models mimic the models of genetic evolution in group structured populations. In these models, the replicator is the individual meme, and the interactor is both the individual human, and the group. That is, there is interaction at two different levels, and the success of a meme is a trade off between these two levels. Henrich [2004], Boyd and Richerson [2010] argue that humans posses adaptations that reduce individual level variation within the group to zero, meaning that we can effectively discuss an entire culture as being a single type. For example, one of these mechanisms is *conformism bias*, which supposes that individuals are adapted to copying the most common strategies within a population.

But in order for cultural group selection to occur, there needs to be variation at the group level. Boyd and Richerson [2010] argue that there are three ways that this can occur. Firstly, it could be sampling variation within a population that partially randomly determines who gets copied. Secondly, environments may vary, meaning that populations might get shifted from an equilibria. Finally, the frequency of cultural traits is affected by learning, and chance variation in cues from the environment will lead to drift-like shifts in trait frequencies.

This vision of cultural evolution is at odds with how I have described large-scale cooperation. In particular, it assumes that there is no variation in the behaviour at the individual level. That would seem to be the case in the model in section 6.3.2, since all individuals choose to 'cooperate,' but recall that cooperation does *not* mean that all individuals are performing the same duty. Consider the example of the guards who, through the act of creating a social contract, can fine members of the population for cheating in public good games. We might be tempted to model their meme for 'guarding' as a feature of the individual that can be copied according to the model of cultural evolution first introduced in section 2.5.4. However, there are two problems with this view. Firstly, there is consistent and stable variation at the individual level within the group. Not everyone is a guard and not everyone is a farmer. What is more, the culture of 'guard' is not spreading through the population by means of social learning. No non-guard is copying their behaviour randomly in the hope that it will pay off. Rather, the group negotiates and appoints the guard.

Where I depart from the description provided in this popular paradigm is that the 'guard' is not a feature of the individuals at all, rather, it is a feature of the group. Of course the *motive* to fulfil the role of the guard is a feature of the individual. That is why, in chapter 6, I highlight that the model is about motivation. But cultural evolution is not concerned by motives, but by the structure of the interactions.

If we go back to the three parts that we require of an evolutionary model, I claim that the *replicator* can be the structural arrangements of the group. For example, a group that is structured around having a leader, or having a method for choosing a leader, or having a particular form of punishment and so on, can be selected in the sense that other groups could adopt those strategies. But like the evolutionary gene, this theory does not mind if there is variation at the individual level within the group. In fact, the variation at the individual level *is* exactly the information that makes up the replicator — it is the thing that is passed on. Of course, this does not mean that individual motivation can be ignored altogether. Far from it. The space of possible structures are likely to be highly constrained by the evolutionary motives of the individuals. All this means is that it is not the individual behaviour that is passed on, but the structure of the the group.

To give an example, take a modern corporation. Within a corporation, there is likely to be a large degree of the division of labour. Individuals are likely to perform a variety of different task in service of the 'group end' — which we can assume is to produce a product to turn a profit. Why does each individual perform their role? Because they get paid to do so, or equivalently, they get fined if they do not. Notice that this is not based on dyadic reciprocation or 'spontaneous cooperation,' there are many institutions that have devised to ensure that individual workers are motivated to work. Here we have stable variation at the individual level of cooperation — we do not need to invoke evolutionary mechanisms like conformism bias or prestige bias to explain how this is stable. It is stable because these are the arrangements that have been made and no individual benefits from deviating from them. Now suppose that this business model is in direct competition to another business model from a different corporation, and that this second company is more successful. For example, this second community has a different and more economical way of dividing its labour. In response to this, the first corporation restructures its arrangements along the lines of the more successful one.

Notice that the 'restructuring' has nothing to do with gradual evolution at the level of memes. It is not a product of individuals behaviourally copying random other individuals within the group. The group is copying the other group. Of course, this raises some questions about how evolutionary change occurs. In particular, how does the group 'decide' to enact a new norm, and how does it negotiate and so on. This is a different, and, I believe, very important question, but is beyond the scope of this thesis to answer. Regardless of the precise mechanism, we can say with some surety that culture does not need to *spread* in a gradual process that the level of the individual meme, it can 'jump' at the level of group, and this jump is part of the evolutionary process at the group level.

The only attempt that I know of to answer the question of how humans have evolved to decide social norms is the work of Binmore [2005b, 1998, 1997]. Binmore argues that humans use a system of the *veil of ignorance*. Briefly, this is a system based on the Rawlsian suggestion of how we *ought* to decide our culture. He begins with the fact that in the game of life — the set of social interactions for which we have to find equilibria — there are many different solutions. He then argues that humans play a *game of morality* over the game of life — a bargaining game where we maximise our utility over the outcome, where we don't know what outcome we get. For example, if we imagine a society where there is a slaver and a slave, then the slaver would not be able to justify his own position in the game of morality, because he would dislike being a slave more than he would like being a slaver. In this view, humans make increasingly better social contracts by negotiating fairer and fairer deals.

7.2.2 Orthodox Economic and Evolutionary Theory Does Not Predict that People are Selfish in the Narrow Sense of the Word

The evolution of cooperation literature has settled on a simple narrative, often using it to justify a particular line of research. This narrative is that it is orthodox in the social sciences to assume that individuals consistently act to maximise their material well being. Hopefully it is now clear that, at least orthodox economics, anthropology and evolutionary game theory does not make these assumptions.

I have made this point consistently throughout this dissertation, and I do not wish to belabour it here. I hope that if we can agree that there is no fundamental disagreement, the we can begin thinking about how we approach experimental work from a different perspective. In particular, there is no need to show that people are willing to give things away, since we have long known that individuals within small-scale societies are willing to do this. Instead, we should be trying to work out why, in their different social contexts, they choose to be so kind.

7.2.3 The Altruistic/Egoist Dichotomy is too Simplistic

A great deal of the debate around human behaviour is centred around the question of whether humans are fundamentally egoistic or altruistic. Indeed, it is impossible to understand the current theoretical and mathematical work on human cooperation without appreciating how they address this basic question. Models assume a zero-sum relationship between altruism and egoism, and are aimed at showing how the former can be evolutionarily adaptive. Economic games are aimed at demonstrating that even in the most controlled of settings, individuals are willing to give selflessly.

But I argue that much like the overly simplistic left-right distinction in politics, the structure of this debate is too simplistic, and that the community ought to move beyond this question as an organising principle for future studies. The problem with the dichotomy is that it does not sufficiently highlight the complexities of studying interactions that are indefinitely repeated. Because interactions repeat, and groups of individuals form webs of responsibilities based on this repetition, and each web is unique, human social life is far more complex than what our theoretical games allow for. So given that the complex repeated games only partially reflect reality, it is hopeless to think that we can accurately describe them in terms of binary 'egoistic' and 'altruistic,' much less that we can empirically measure 'altruisticness.'

7.2.4 Game Theory is a Useful Theoretical Tool for Studying Human Culture, but we must be Careful to use it as an Empirical Tool

Game theory is chiefly an theoretical tool for explaining how two or more individuals with differing goals can interact to form stable outcomes. Its primary use is in distinguishing a *goal* from a *equilibrium*. However, while I argue it can be useful for guiding empirical work, we must be careful with relying on information from economic games too much [Cronk, 2007].

When humans interact, we form extremely complex webs of responsibilities, each one with its own unique characteristics. When groups of humans come together like this, we tend to create equally complex systems of communication. A raised eyebrow might imply a threat. Laughing at a joke might imply undermining another individuals position in the group and so on. The problem is that these are all quite difficult things to measure, especially if we try to take a cross-cultural view. The recent application of game theory as an empirical tool has to be aware of these problems. Many scholars who use economic games are aware of this criticism, but do believe that it is problematic. For example:

> A particularly unsatisfactory maladaption story is the argument that experimental subjects are not capable of distinguishing between repeated interactions and one-shot interactions. As a consequence, so the story goes, subjects tend to inappropriately apply heuristics and habits in experimental one-shot interactions (i.e., they take revenge or they reward helping behavior), that are only adaptive in a repeated interaction context but not in the one-shot context. Fehr et al. [2002]

Nevertheless, I think the criticism still stands. I believe that by investigating behaviour in games, we are far more likely to observe some aspect of the social contract, rather than an underlying preference. I suggest that rather than taking the games to the societies and testing them 'out-of-the-box,' that if we want to use economic games, we should try to invoke some part of the social contract. The current methodology tries to obscure all possibe normative behaviour, which I believe is simply impossible.

7.3 Future Work

There are two primary areas where I would like to extend the work in this dissertation. The first is to devise a theory of how humans have evolved to form social contracts. As mentioned previously, the only work that I currently know of that attempts to do this was by Economist and Philosopher Ken Binmore. Binmore begins with Rawls' theory of the 'veil of ignorance.' The veil of ignorance is a mechanism in moral philosophy that can be used to judge. The idea is that in order to create a moral social contract, we must argue as if we did not know what position we would have in a society. The idea is that if we did this, then we would make a fair society.

Binmore goes further by suggesting that this is not only how humans ought to devise our social contracts, but that this what we actually do. He represents this by thinking about how humans choose an equilibrium in the game of life. He argues that what we do is play the game of morality. The game of morality is a bargaining game, played over the game of life where we do not know what player we will be.

While I believe Binmore's theory is a significant step forwards, and is certainly more sophisticated than my own assumption that contracts are simply perfectly egalitarian, I have several doubts about how accurately they reflect reality. The reason for this that if this was the primary method that humans use, then we would find that humans would distribute things out more evenly. Rather, I argue that the 'veil of ignorance' itself is part of the social contract — it is a part of the culture that some individuals have chosen to adopt.

Where I primarily disagree with this model is that it assumes that social dominance does not occur in humans. In my framework, dominance is simply a social contract — it is a cultural arrangement that solves problems in the game of life. However, by definition, dominance is not 'fair.' I believe that we have a reasonably solid understanding of why dominance is stable — the simple model of hawk-dove explains this. However, I argue that we have a relatively bad understanding of why it is accepted in the first place.

The second way that I would like to develop this work is to better extend the framework for describing cultural evolution from the perspective of individuals within groups dividing their labour. Currently, the 'memetic' theory of change focuses on individual strategies spreading through a population because they are individually beneficial. However, I argue that this model is inappropriate for many examples of cultural evolution.

The basic idea that I have tried to communicate in this dissertation is that, in the same way that we can think of individuals acting to achieve things, so can we think about group acting to achieve ends. To achieve these ends, we divide our labour and responsibilities. The current models of cultural evolution from the perspective of groups denies this fact, and instead assumes that individuals all perform the same behaviour. But these assumptions are selected for their mathematical convenience, rather than because they reflect reality.

BIBLIOGRAPHY

Richard D Alexander. *The biology of moral systems.* Transaction Publishers, 1987.

Robert Axelrod and William D Hamilton. The evolution of cooperation. *Science*, 211 (4489):1390–1396, 1981.

Helen Bernhard, Ernst Fehr, and Urs Fischbacher. Group affiliation and altruistic norm enforcement. *The American Economic Review*, pages 217–221, 2006.

Laura Betzig and Paul Turke. Fatherhood by rank on ifaluk. *Father-child relations: cultural and biosocial contexts. Foundations of human behavior*, pages 111–129, 1992.

Laura L Betzig and Paul W Turke. Parental investment by sex on ifaluk. *Ethology and Sociobiology*, 7(1):29–37, 1986.

Eran Binenbaum. *Knight versus Herskovits: A Methodologically Charged Debate in the 1940s.* School of Economics, University of Adelaide, 2005.

Ken Binmore. *Game theory and the social contract: Just Playing.* MIT Press, 1997.

Ken Binmore. *Game theory and the social contract: Playing fair*, volume 2. The MIT Press, 1998.

Ken Binmore. Economic man–or straw man? *Behavioral and Brain Sciences*, 28(6): 817–818, 2005a.

Ken Binmore. *Natural justice.* Oxford University Press, 2005b.

Ken Binmore. Does game theory work? the bargaining challenge. 2007.

Ken Binmore. Social norms or social preferences? *Mind & Society*, 9(2):139–157, 2010.

Ken Binmore and Avner Shaked. Experimental economics: Where next? *Journal of Economic Behavior & Organization*, 73(1):87–100, 2010.

Douglas W Bird. Delayed reciprocity and tolerated theft: The behavioral ecology of food-sharing strategies. *Current Anthropology*, 38(1):49–78, 1997.

Rebecca Bliege Bird, Eric Smith, and Douglas W Bird. The hunting handicap: costly signaling in human foraging strategies. *Behavioral Ecology and Sociobiology*, 50(1): 9–19, 2001.

Susan Blackmore. *The meme machine*. Oxford University Press, 2000.

Nicholas G Blurton Jones. A selfish origin for human food sharing: tolerated theft. *Ethology and sociobiology*, 5(1):1–3, 1984.

Christopher Boehm. *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press, 2009.

Samuel Bowles. Group competition, reproductive leveling, and the evolution of human altruism. *science*, 314(5805):1569–1572, 2006.

Samuel Bowles. Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science*, 324(5932):1293–1298, 2009.

Samuel Bowles and Herbert Gintis. *A cooperative species: Human reciprocity and its evolution*. Princeton University Press, 2011.

Robert Boyd. *Culture and the evolutionary process*. University of Chicago Press, 1988.

Robert Boyd and Peter J Richerson. The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132(3):337–356, 1988.

Robert Boyd and Peter J Richerson. The evolution of indirect reciprocity. *Social Networks*, 11(3):213–236, 1989.

Robert Boyd and Peter J Richerson. Solving the puzzle of human cooperation. 2003.

Robert Boyd and Peter J Richerson. Solving the puzzle of human cooperation. *Evolution and culture*, pages 105–132, 2005.

Robert Boyd and Peter J Richerson. Transmission coupling mechanisms: cultural group selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365 (1559):3787–3795, 2010.

Robert Boyd, Herbert Gintis, Samuel Bowles, and Peter J Richerson. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, 100(6): 3531–3535, 2003.

Robert Boyd, Herbert Gintis, and Samuel Bowles. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978):617–620, 2010.

Robert Boyd, Peter J Richerson, and Joseph Henrich. Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral ecology and sociobiology*, 65(3):431–444, 2011.

Nicholas F Britton. *Essential mathematical biology.* Springer, 2003.

Maxwell N Burton-Chellew and Stuart A West. Prosocial preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences*, 110(1):216–221, 2013.

Colin F Camerer and Ernst Fehr. When does "economic man" dominate social behavior? *Science*, 311(5757):47–52, 2006.

Lewis Carroll. *Sylvie and Bruno concluded.* Macmillan, 1893.

Napoleon A Chagnon. *Yanomamö: The fierce people.* Holt, Rinehart and Winston New York, 1968.

Napoleon A Chagnon. Life histories, blood revenge, and warfare in a tribal population. *Science*, 239(4843):985–992, 1988.

Ananish Chaudhuri. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, 14(1):47–83, 2011.

Jung-Kyoo Choi and Samuel Bowles. The coevolution of parochial altruism and war. *science*, 318(5850):636–640, 2007.

Scott Cook. The obsolete anti-market mentality: A critique of the substantive approach to economic anthropology. *American Anthropologist*, 68(2):323–345, 1966.

Lee Cronk. The influence of cultural framing on play in the trust game: a maasai example. *Evolution and Human Behavior*, 28(5):352–358, 2007.

Charles Darwin. On the origin of the species by natural selection. 1859.

Marian Stamp Dawkins and Tim Guilford. The corruption of honest signalling. *Animal Behaviour*, 41(5):865–873, 1991.

Richard Dawkins. *The selfish gene.* Oxford University Press Oxford, 1976.

Richard Dawkins. Twelve misunderstandings of kin selection. *Zeitschrift für Tierpsychologie*, 51(2):184–200, 1979.

Richard Dawkins. *The extended phenotype: The long reach of the gene*. Oxford University Press, 1999.

Richard Dawkins and John R Krebs. Animal signals: information or manipulation. *Behavioural ecology: An evolutionary approach*, 2:282–309, 1978.

Frans De Waal. *Chimpanzee politics: Power and sex among apes*. JHU Press, 1982.

Frans De Waal. The integration of dominance and social bonding in primates. *Quarterly Review of Biology*, pages 459–479, 1986.

Frans De Waal. *Primates and Philosophers: How Morality Evolved*. Princeton University Press, 2009.

Martin Dufwenberg, Simon Gächter, and Heike Hennig-Schmidt. The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478, 2011.

Robin Dunbar. *Grooming, gossip, and the evolution of language*. Harvard University Press, 1998.

Emile Durkheim. The rules of the sociological method. *John H. Mueller and Sarah A. Solovay, translators. Glencoe*, 111, 1895.

Rada Dyson-Hudson, Dominique Meekers, and Neville Dyson-Hudson. Children of the dancing ground, children of the house: Costs and benefits of marriage rules (south turkana, kenya). *Journal of anthropological research*, pages 19–47, 1998.

Martijn Egas and Arno Riedl. The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 275 (1637):871–878, 2008.

C El Mouden, J-B André, O Morin, and D Nettle. Cultural transmission and the evolution of human behaviour: a general approach based on the price equation. *Journal of evolutionary biology*, 27(2):231–241, 2014.

David Erdal, Andrew Whiten, Christopher Boehm, and Bruce Knauft. On human egalitarianism: an evolutionary product of machiavellian status escalation?, 1994.

Ilan Eshel and Luigi Luca Cavalli-Sforza. Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335, 1982.

Edward Evan Evans-Pritchard. *The Nuer*, volume 940. Oxford Clarendon, 1940.

Ernst Fehr and Urs Fischbacher. Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *The economic journal*, 112(478):C1–C33, 2002.

Ernst Fehr and Urs Fischbacher. The nature of human altruism. *Nature*, 425(6960): 785–791, 2003.

Ernst Fehr and Simon Gächter. Altruistic punishment in humans. *Nature*, 415(6868): 137–140, 2002.

Ernst Fehr and Klaus M Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

Ernst Fehr, Urs Fischbacher, and Simon Gächter. Strong reciprocity, human cooperation, and the enforcement of social norms. *Human nature*, 13(1):1–25, 2002.

Brian Ferguson. The causes and origins of "primitive warfare": on evolved motivations for war. *Anthropological Quarterly*, 73(3):159–164, 2000.

Brian Ferguson. The prehistory of war and peace in europe and the near east. *War, Peace, and Human Nature: The Convergence of Evolutionary and Cultural Views*, page 191, 2013.

Ronald A. Fisher. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1930.

James H Fowler, Tim Johnson, and Oleg Smirnov. Egalitarian motive and altruistic punishment. *Nature*, 433(10.1038), 2005.

Douglas Fry. War, peace, and human nature. *War, Peace, and Human Nature: The Convergence of Evolutionary and Cultural Views*, page 1, 2013.

Simon Gächter. In the lab and the field: Punishment is rare in equilibrium. *Behavioral and Brain Sciences*, 35(01):26–28, 2012.

Andy Gardner and Stuart A West. Inclusive fitness: 50 years on. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1642):20130356, 2014.

Azar Gat. The changing character of war. *The Changing Character of War*, 2011.

Herbert Gintis. Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206(2):169–179, 2000.

Herbert Gintis and Ernst Fehr. The social structure of cooperation and punishment. *Behavioral and Brain Sciences*, 35(01):28–29, 2012.

Herbert Gintis, Samuel Bowles, Robert Boyd, and Ernst Fehr. Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24(3):153–172, 2003.

Herbert Gintis, Joseph Henrich, Samuel Bowles, Robert Boyd, and Ernst Fehr. Strong reciprocity and the roots of human morality. *Social Justice Research*, 21(2):241–253, 2008.

Maurice Godelier. *The making of great men: male domination and power among the New Guinea Baruya*. Number 56. Cambridge University Press, 1986.

Cristina M Gomes and Christophe Boesch. Wild chimpanzees exchange meat for sex on a long-term basis. *PLoS One*, 4(4):e5116, 2009.

Alan Grafen. Biological signals as handicaps. *Journal of theoretical biology*, 144(4): 517–546, 1990.

Francesco Guala. Reciprocity: Weak or strong? what punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences*, 35(01):1–15, 2012a.

Francesco Guala. Strong reciprocity is real, but there is no evidence that uncoordinated costly punishment sustains cooperation in the wild. *Behavioral and Brain Sciences*, 35(01):45–59, 2012b.

Michael Gurven. To give and to give not: the behavioral ecology of human food transfers. *Behavioral and Brain Sciences*, 27(4):543–559, 2004a.

Michael Gurven. Reciprocal altruism and food sharing decisions among hiwi and ache hunter–gatherers. *Behavioral Ecology and Sociobiology*, 56(4):366–380, 2004b.

Michael Gurven, Kim Hill, and Hillard Kaplan. From forest to reservation: Transitions in food-sharing behavior among the ache of paraguay. *Journal of Anthropological Research*, pages 93–120, 2002.

William D Hamilton. The genetical evolution of social behaviour. ii. *Journal of theoretical biology*, 7(1):17–52, 1964.

Peter Hammerstein. The role of asymmetries in animal contests. *Animal Behaviour*, 29(1):193–205, 1981.

Christopher M Hann. *Property relations: renewing the anthropological tradition*. Cambridge University Press, 1998.

Charles William Merton Hart, Arnold R Pilling, and Jane C Goodale. *The Tiwi of North Australia*. Holt, Rinehart and Winston, 1966.

John Hartung. The group selection chimera. pages 22–27, 1999.

Kristen Hawkes. Sharing and collective action. *Evolutionary ecology and human behavior*, pages 269–300, 1992.

Kristen Hawkes and Rebecca Bliege Bird. Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology: Issues, News, and Reviews*, 11 (2):58–67, 2002.

Friedrich A Hayek. The fatal conceit: The errors of socialism. *The collected works of Friedrich August Hayek*, 1, 1988.

Joseph Henrich. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53(1):3–35, 2004.

Joseph Henrich, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, Michael Alvard, Abigail Barr, Jean Ensminger, et al. "economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and brain sciences*, 28(6):795–814, 2005.

Benedikt Herrmann, Christian Thöni, and Simon Gächter. Antisocial punishment across societies. *Science*, 319(5868):1362–1367, 2008.

Kim Hill. Hunting and human evolution. *Journal of Human Evolution*, 11(6):521–544, 1982.

Kim Hill. Altruistic cooperation during foraging by the ache, and the evolved human predisposition to cooperate. *Human Nature*, 13(1):105–128, 2002.

Kim Hill, Hillard Kaplan, and Kristen Hawkes. On why male foragers hunt and share food, 1993.

David L Hull. Interactors versus vehicles. *The role of behavior in evolution*, pages 19–50, 1988.

David L Hull. Science and selection. *Cambridge Un. Press, Cambridge*, 2001.

David Hume. *A treatise of human nature*. Courier Dover Publications, 1738.

Sharon Hutchinson. *Nuer dilemmas: coping with money, war, and the state*. University of California Pr, 1996.

Glynn Isaac. The food-sharing behavior of protohuman hominids. *Scientific American*, 1978.

Nicholas G Blurton Jones. Tolerated theft suggestions about the ecology and evolution of sharing, hoarding, and scrounging. *Primate politics*, page 170, 1991.

Hillard Kaplan, Kim Hill, Kristen Hawkes, and Ana Hurtado. Food sharing among ache hunter-gatherers of eastern paraguay. *Current Anthropology*, 25(1):113–115, 1984.

Hillard Kaplan, Kim Hill, Rowe V Cadeliña, Brian Hayden, David Charles Hyndman, Richard J Preston, Eric Alden Smith, David E Stuart, and David R Yesner. Food sharing among ache foragers: Tests of explanatory hypotheses [and comments and reply]. *Current anthropology*, pages 223–246, 1985.

Hillard Kaplan, Kim Hill, and A Magdalena Hurtado. Risk, foraging, and food sharing among the ache. *Risk and uncertainty in tribal and peasant economies*, pages 107–143, 1990.

Hillard Kaplan, Kim Hill, Jane Lancaster, and A Magdalena Hurtado. A theory of human life history evolution: diet, intelligence, and longevity. *Evolutionary Anthropology Issues News and Reviews*, 9(4):156–185, 2000.

Hillard Kaplan, Michael Gurven, Kim Hill, and A Magdalena Hurtado. The natural history of human food sharing and cooperation: a review and a new multi-individual approach to the negotiation of norms. *Moral sentiments and material interests: The foundations of cooperation in economic life*, pages 75–113, 2005.

Hillard S Kaplan, Paul L Hooper, and Michael Gurven. The evolutionary and ecological roots of human social organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3289–3299, 2009.

Lawrence H Keeley. *War before civilization*. Oxford University Press, 1996.

Frank H Knight. Anthropology and economics. *The Journal of Political Economy*, pages 247–268, 1941.

Klaus-Friedrich Koch. *War and peace in Jalemo: The management of conflict in Highland New Guinea*. Harvard University Press Cambridge, 1974.

John R Krebs, Richard Dawkins, et al. Animal signals: mind-reading and manipulation. *Behavioural Ecology: an evolutionary approach*, 2:380–402, 1984.

Steven A LeBlanc. *Prehistoric warfare in the American Southwest*. University of Utah Press Salt Lake City, 1999.

Steven A LeBlanc. *Constant battles: The myth of the peaceful, noble savage*. Macmillan, 2003.

Steven A LeBlanc. Why warfare? lessons from the past. *Daedalus*, 136(1):13–21, 2007.

Frank W Marlowe. Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, 14(2):54–67, 2005.

Frank W Marlowe, J Colette Berbesque, Clark Barrett, Alexander Bolyanatz, Michael Gurven, and David Tracer. The 'spiteful' origins of human cooperation. *Proceedings of the Royal Society B: Biological Sciences*, 2010.

Barnaby Marsh and Alex Kacelnik. Framing effects and risky decisions in starlings. *Proceedings of the National Academy of Sciences*, 99(5):3352–3355, 2002.

Karl Marx and Friedrich Engels. *Wage-labor and capital*. Andrades, 1847.

Sarah Mathew and Robert Boyd. Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, 108(28):11375–11380, 2011.

John Maynard Smith. The theory of games and the evolution of animal conflicts. *Journal of theoretical biology*, 47(1):209–221, 1974.

Richard McElreath and Robert Boyd. *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press, 2008.

Kelly R McGuire and William R Hildebrandt. Re-thinking great basin foragers: prestige hunting and costly signaling during the middle archaic period. *American Antiquity*, pages 695–712, 2005.

Mervyn J Meggitt. *Blood is their argument: Warfare among the Mae Enga tribesmen of the New Guinea highlands*. Mayfield Palo Alto, CA, 1977.

Carl Menger. *Principles of economics*. Ludwig von Mises Institute, 1871.

John F Nash et al. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

Ludwig Johann Neumann and Oskar Morgenstern. *Theory of games and economic behavior*, volume 60. Princeton university press Princeton, NJ, 1947.

Friedrich Wilhelm Nietzsche. *The birth of tragedy and the genealogy of morals*, volume 81. Random House LLC, 1887.

Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573–577, 1998.

Martin A Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437 (7063):1291–1298, 2005.

Mancur Olson and Mancur Olson. *The logic of collective action: public goods and the theory of groups.*, volume 124. Harvard University Press, 2009.

Elinor Ostrom. *Governing the commons: The evolution of institutions for collective action*. Cambridge university press, 1990.

Elinor Ostrom, James Walker, and Roy Gardner. Covenants with and without a sword: Self-governance is possible. *American Political Science Review*, 86(02):404–417, 1992.

Karthik Panchanathan and Robert Boyd. Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 432(7016):499–502, 2004.

Michael Parker Pearson. Warfare, violence and slavery in later prehistory: an introduction. *Warfare, Violence and Slavery in Later Prehistory*, pages 19–33, 2005.

Steven Pinker. *The better angels of our nature: The decline of violence in history and its causes*. Penguin UK, 2011.

Karl Polanyi. *The great transformation: The political and economic origins of our time*. Beacon Press, 1944.

Karl Polanyi. The economy as instituted process. *Trade and market in the early empires*, 243, 1957.

Leopold J Pospisil. *The Kapauku Papuans of West New Guinea*. Holt, Rinehart and Winston, 1963.

Simon T Powers, Daniel J Taylor, and Joanna J Bryson. Punishment can promote defection in group-structured populations. *Journal of theoretical biology*, 311:107–116, 2012.

George R Price. Extension of covariance selection mathematics. *Annals of human genetics*, 35(4):485–490, 1972.

John Rawls. *A theory of justice*. Harvard university press, 1971.

Dwight Read. Culture: The missing piece in theories of weak and strong reciprocity. *Behavioral and Brain Sciences*, 35(01):35–36, 2012.

Leonard E Read. I, pencil. *Irvington-on-Hudson, NY: Foundation for Economic Education*, 1958.

Matt Ridley. *The origins of virtue*. Penguin UK, 1997.

Arthur J Robson. Efficiency in evolutionary games: Darwin, nash and the secret handshake. *Journal of Theoretical Biology*, 144(3):379–396, 1990.

Paul A Samuelson. Consumption theory in terms of revealed preference. *Economica*, pages 243–253, 1948.

Amartya Sen. Behaviour and the concept of preference. *Economica*, pages 241–259, 1973.

Brian Skyrms. *Evolution of the social contract*. Cambridge University Press, 1996.

Eric A Smith and R Bliege Bird. Costly signaling and cooperative behavior. *Moral sentiments and material interests: On the foundations of cooperation in economic life*, pages 115–148, 2005.

Eric Alden Smith and Rebecca L Bliege Bird. Turtle hunting and tombstone opening: public generosity as costly signaling. *Evolution and Human Behavior*, 21(4):245–261, 2000.

John Maynard Smith. Group selection and kin selection. *Nature*, 201:1145–1147, 1964.

John Maynard Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.

John Maynard Smith and David Harper. *Animal signals*. Oxford University Press, 2003.

John Maynard Smith and Geoffery A Parker. The logic of asymmetric contests. *Animal Behaviour*, 24(1):159–175, 1976.

John Maynard Smith and George R Price. lhe logic of animal conflict. *Nature*, 246:15, 1973.

Maynard J Smith and David GC Harper. Animal signals: models and terminology. *Journal of theoretical biology*, 177(3):305–311, 1995.

Elliott Sober and David Sloan Wilson. *Unto others: The evolution and psychology of unselfish behavior*. Number 218. Harvard University Press, 1999.

Joseph Soltis, Robert Boyd, and Peter J Richerson. Can group-functional behaviors evolve by cultural group selection?: An empirical test. *Current Anthropology*, pages 473–494, 1995.

Herbert Spencer. *The principles of sociology*, volume 1. Appleton, 1895.

Niko Tinbergen. On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20 (4):410–433, 1963.

Robert Trivers. The evolution of reciprocal altruism. *Quarterly review of biology*, pages 35–57, 1971.

Robert Trivers. Reciprocal altruism: 30 years later. In *Cooperation in primates and humans*, pages 67–83. Springer, 2006.

Paul W Turke. Evolution and the demand for children. *Population and Development Review*, pages 61–90, 1989.

Paul W Turke and Laura L Betzig. Those who can do: Wealth, status, and reproductive success on ifaluk. *Ethology and Sociobiology*, 6(2):79–87, 1985.

Harry Holbert Turney-High and David C Rapoport. *Primitive war: Its practice and concepts*. University of South Carolina Press Columbia, South Carolina, 1949.

Peter Vanderschraaf. The informal game theory in hume's account of convention. *Economics and Philosophy*, 14:215–248, 1998.

Michael J Wade. Soft selection, hard selection, kin selection, and group selection. *American naturalist*, pages 61–73, 1985.

Stuart A West, Ashleigh S Griffin, and Andy Gardner. Social semantics: how useful has group selection been? *Journal of Evolutionary Biology*, 21(1):374–385, 2008.

Stuart A West, Claire El Mouden, and Andy Gardner. Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32 (4):231–262, 2011.

Polly Wiessner. Risk, reciprocity and social influences on !kung san economics. *Politics and history in band societies*, 61:84, 1982.

Polly Wiessner. Hunting, healing, and hxaro exchange: A long-term perspective on !kung large-game hunting. *Evolution and Human Behavior*, 23(6):407–436, 2002.

Polly Wiessner. Experimental games and games of life among the ju/'hoan bushmen. *Current Anthropology*, 50(1):133–138, 2009.

Polly Wiessner, Akii Tumu, and Nitze Pupu. *Historical vines: Enga networks of exchange, ritual, and warfare in Papua New Guinea*. Smithsonian Institution Press Washington, DC, 1998.

George C Williams. Adaptation and natural selection: A critique of some current evolutionary thoughtprinceton university press. *Princeton, NJ*, 1966.

David Sloan Wilson. A theory of group selection. *Proceedings of the national academy of sciences*, 72(1):143–146, 1975.

David Sloan Wilson and Elliott Sober. Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, 17(04):585–608, 1994.

Edward O Wilson. *The social conquest of earth*. WW Norton & Company, 2012.

Bruce Winterhalder. A marginal model of tolerated theft. *Ethology and Sociobiology*, 17(1):37–53, 1996.

James Woodburn. Sharing is not a form of exchange: An analysis of property-sharing in immediate return hunter-gatherer societies. *Property relations: renewing the anthropological tradition*, pages 48–63, 1998.

Richard W Wrangham and Dale Peterson. *Demonic males: Apes and the origins of human violence*. Houghton Mifflin Harcourt, 1996.

Quincy Wright. *The Study of War*. The University of Chicago press, 1942.

Vero Copner Wynne-Edwards. Animal dispersion in relation to social behaviour. 1962.

Amotz Zahavi and Avishag Zahavi. The handicap principle: a missing piece of darwin's puzzle. 1997.

Kevin JS Zollman, Carl T Bergstrom, and Simon M Huttegger. Between cheap and costly signals: the evolution of partially honest communication. *Proceedings of the Royal Society B: Biological Sciences*, page 20121878, 2012.