

Mälardalen University Press Dissertations  
No. 136

**MODELLING TWO-PERSON INTERACTIONS  
WITHIN AND BETWEEN CULTURAL GROUPS**

**Fredrik Jansson**

**2013**



School of Education, Culture and Communication

Copyright © Fredrik Jansson, 2013

ISBN 978-91-7485-100-7

ISSN 1651-4238

Printed by Mälardalen University, Västerås, Sweden

Mälardalen University Press Dissertations  
No. 136

MODELLING TWO-PERSON INTERACTIONS  
WITHIN AND BETWEEN CULTURAL GROUPS

Fredrik Jansson

Akademisk avhandling

som för avläggande av filosofie doktorsexamen i matematik/tillämpad matematik  
vid Akademin för utbildning, kultur och kommunikation kommer att offentligen  
försvaras fredagen den 22 februari 2013, 13.15 i Lambda, Mälardalens högskola, Västerås.

Fakultetsopponent: Dr Anne Kandler, City University London, Mathematics Centre



Akademin för utbildning, kultur och kommunikation

## Abstract

The groups with which we associate influence our actions. This is often the case even when they are not deliberately organised but rather based on social categories, such as sex and skin colour, or cultural homogeneity, such as common language or customs. Group membership can cause widespread phenomena such as ingroup favouritism, polarisation of opinion and competition. Previous experiments have shown that these effects can be triggered by even completely arbitrary distinctions between groups. This thesis uses mathematical models to investigate under what circumstances these phenomena can arise.

Using a game theoretical approach, the first three papers address the evolution of ingroup favouritism. Previous models have focused on the prisoners' dilemma, interactions where the socially optimal behaviour is to co-operate, but where it is in the individual's self-interest not to. The results presented here suggest that co-ordination problems may have been more important than those of co-operation in the evolution of an ingroup bias. In particular, this applies to common goals that require trust. It is also demonstrated in a behavioural experiment that such trust is most common within groups, but that it can emerge between groups through group reputation.

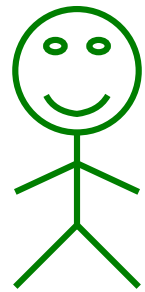
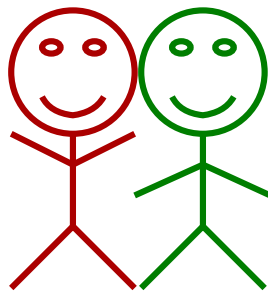
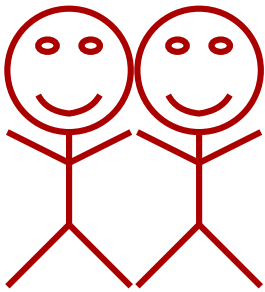
The fourth paper focuses on a model on how cultural groups in contact can develop common norms, rather than polarise into different norm groups, by assuming a confirmation bias. The model is empirically tested on demographic and linguistic data from Mauritian Creole, a natural language developed from the mixing of parent languages.

In the fifth paper, the group is defined by common preferences (e.g. for pop songs), which are transmitted in a random copying model. The competitive success of the groups, with respect to their size, is recorded on a toplist, the turnover rate of which is derived.

In the final paper, people match up in pairs between groups according to their preferences, and all stable matchings are found under a specific assumption of bounded rationality, when people's individual behaviour may be affected by the consequences for fellow group members.

# Modelling Two-Person Interactions Within and Between Cultural Groups

Fredrik Jansson





## Sammanfattning

Vi associeras eller förknippar oss själva med otaliga grupper i samhället. Grupper kan bildas genom att vi exempelvis aktivt väljer vilka vi umgås med eller bildar arbetslag, men kan också vara mindre självvalda och grunda sig på kön och klasstillhörighet eller mer praktiska förutsättningar såsom gemensamt språk. Dessa gruppstillhörigheter kan ofta påverka hur vi bemöter varandra, även när vi inte känner andra gruppmedlemmar bättre än personer utanför gruppen. När vi har grupper runtomkring oss som vi inte ingår i uppstår fenomen som gruppdiskriminering med favorisering av den egna gruppen, polarisering av åsikter och uttryck, med minskat utbyte mellan grupper, samt gruppkonkurrens. Tidigare experiment har visat att den här typen av fenomen kan uppstå från helt godtyckliga gruppindelningar. I den här avhandlingen används matematiska modeller för att undersöka hur olika grundläggande förutsättningar påverkar splittring och samförstånd i parvisa interaktioner inom och mellan grupper.

De första artiklarna undersöker uppkomsten av gruppdiskriminering. I avhandlingens första artikel motiveras varför tidigare modeller är dåligt lämpade för att studera fenomenet. I den andra visas att diskriminering uppstår lättare ur koordinerings- än samarbetsproblem, situationer då det är viktigast att veta vad som förväntas av varandra, snarare än då vi kan bete oss altruistiskt eller själviskt. Tydligast blir detta då gemensamma mål kräver tillit till varandra. Den tredje artikeln visar genom beteendeeperiment att sådan tillit är vanligast inom grupper, men även att den kan uppstå mellan grupper om den ena gruppen har rykte om sig att vara tillitsfull.

I den fjärde artikeln utvecklas en modell för hur grupper med mycket kontakt kan utveckla gemensamma normer istället för att polariseras, genom att människor lägger störst vikt vid sådant som bekräftar deras rådande uppfattningar, ett välkänt fenomen kallat för konfirmeringsbias, samtidigt som de är mottagliga för ny information. Modellen testas empiriskt på demografiska och språkliga data från mauritisk kreol, som utvecklats genom ett behov av att kunna kommunicera mellan människor från olika språkgrupper.

I den femte artikeln ges gruppstillhörigheten av gemensamma preferenser (t.ex. för poplåtar) i en modell där individer tar efter varandra slumpmässigt och gruppernas framgång i konkurrensen uttrycks i en topplista över de största grupperna, vars omsättning härleds.

Avhandlingens sista artikel handlar om parvisa matchningar av individer från två olika grupper (t.ex. arbetsgivare och arbetstagare) utifrån deras preferenser och hur matchningarna påverkas om individerna präglas av illvilja gentemot andra i den egna gruppen.

## Acknowledgements

There are so many people involved in writing a thesis. The brilliant ideas you had usually originated elsewhere, were spread and copied with slight modifications, being refined, until they reached you, sometimes so subtly that you did not notice that you were not the mastermind behind them, or if you did, then you could not easily trace them back to where they came from. To be fair, I am going to claim that several of my numerous bad ideas were not really my doing either, when it comes to it. I could have thanked the whole of humanity for all the knowledge it has provided. However, I have indeed had some transmitters of good ideas round me over the past few years that I can identify as such, and that do deserve special mention here.

This thesis is about groups, and my first thanks go to one that I have been fortunate enough to label as my ingroup since I started my doctoral studies, the *Centre for the Study of Cultural Evolution*. We have all been part of many great discussions, and you have all, in one way or the other, contributed to my knowledge, well-being, and what you are now holding in your hands.

I would never have had the opportunity to come to the centre, were it not for my supervisor *Kimmo Eriksson*, who is full of ideas, and has an almost magic ability to sort out and clarify ideas and key messages when they become messy to me. In this sense, he seems to be unprecedented by no one, except possibly *Pontus Strimling*, who has certainly lived up to his reputation as a man of vivid ideas, some of which have been implemented in this thesis. Pontus became my assistant supervisor after *Jonas Sjöstrand* moved to lower latitudes (one metro stop away). Except for being an exceptional mathematician teaching me the more hardcore mathematics, Jonas has been a true mentor, always there to help, be it with computations, carrying moving carton boxes or providing me with food.

As if three supervisors was not luxurious enough for an ignorant PhD student (well, they all, simultaneously, went on parental leave in my final year), *Johan Lind* has, at times, almost been like one, taking an interest in my research and always willing to discuss almost anything (heatedly), both within and outside science. *Magnus Enquist* has been an inspiration when it comes to cultural evolution in a broader sense, and has arranged several informal Sunday meetings where the grand picture of the discipline has been discussed. And, were I ever to write a thesis on birds in the future, these two people would certainly be the first ones to ask to become my supervisors.

I have had several general discussions with and asked so many questions in particular to my fellow postgraduate students and, in the last three cases, (previous) roommates, *Alexander Funcke*, *Micael Ehn*, *Markus Jonsson* and *Hanna Aronsson*. Thank you for your time, help and company. Do not ask me why they are not *current* roommates.

My co-authors have made very tangible contributions to this thesis, by sharing their knowledge on the things I study, and, not least, actually doing part of the work. On the other hand, some co-authors, *Patrik Lindenfors* and



*Mikael Sandberg*, have actually done a pretty good job keeping me away from working on the thesis. Anyway, I am glad you did, as I enjoyed working with you and our efforts paid off in some neat publications on democratic transitions.

Some tangible contributions have also been made by *Torbjörn Lundh*, who were my mock faculty examiner. It is quite a feat just to read all the papers, but he also gave some useful comments on them.

While all of these people have been part of forming the end result, others have been important for getting me to Stockholm in the first place. One of them is the supervisor of my master's thesis, *Einar Steingrímsson*, who believed in me and encouraged me to pursue doctoral studies. The national *Graduate School in Mathematics and Computing* was also important for getting me here, in that, believe it or not, they actually agreed to finance me for all this time.

Last, but most certainly not least, I would like to thank family and friends for brightening the time not spent on coursework and research. Besides, without you, I would probably not have found the energy to write this thesis.

I would also like to thank all of you, friends and random people, who have taken an interest in and criticised the very ideas of my research, thus forcing me to think it through properly and sharpen arguments. When I was doing pure mathematics, before trying to say something about the real world, I never realised there could be such controversies in science. Thank you, postdocs, for inventing Beer o'clock, a nice round off of the week. Thank you, Kalejdoskop in Swedish Radio P2 for providing me with excellent background music to get the ideas flowing.

What the heck, I will thank the whole of humanity, because without you, there would be no culture for me to study.

Stockholm and SJ2000, 23 December 2012

# Contents

## I Introduction

<b>Introduction</b>	<b>3</b>
Overview . . . . .	3
Structure of the Thesis . . . . .	3
Background on Cultural Groups . . . . .	4
What Is a Group? . . . . .	4
Mechanisms with Several Groups . . . . .	4
Ethnocentrism . . . . .	5
Competition . . . . .	6
Polarisation . . . . .	7
Mathematical Modelling . . . . .	8
Neutral Models . . . . .	9
Game Theory . . . . .	10
Classical Game Theory . . . . .	11
Evolutionary Game Theory . . . . .	11
Behavioural Game Theory . . . . .	12
Agent-Based Models . . . . .	13
Previous Models on Intergroup Interactions . . . . .	13
Overview . . . . .	14
Game Theory . . . . .	14
Evolutionary Biology . . . . .	15
Cultural Evolution . . . . .	16
Current Models . . . . .	18
Papers of the Thesis . . . . .	20
Group Membership Affects Behaviour . . . . .	21
Group Membership is Contingent on Previous Interactions . . . . .	21
Matchings Between Groups . . . . .	22
Directions for Future Research . . . . .	23
Bibliography . . . . .	23

## **II Papers**

### **I. Pitfalls in Spatial Modelling of Ethnocentrism**

Fredrik Jansson

*Journal of Artificial Societies and Social Simulations* (in press)

### **II. What Strategic Structures Support the Evolution of Ethnocentrism?**

Fredrik Jansson

### **III. Trusting You Trusting Me: The Importance of Beliefs about Trust in the Stag Hunt/Assurance Game**

Fredrik Jansson & Kimmo Eriksson

### **IV. Modelling the Evolution of Creoles**

Fredrik Jansson, Mikael Parkvall & Pontus Strimling

### **V. Bentley's Conjecture on Popularity Toplist Turnover under Random Copying**

Kimmo Eriksson, Fredrik Jansson & Jonas Sjöstrand

*Ramanujan Journal* 23 (2010): 371–396

### **VI. The Assignment Game with Negative Externalities and Bounded Rationality**

Kimmo Eriksson, Fredrik Jansson & Thomas Vetander

*International Game Theory Review* 13(4) (2011): 443–459

Reprints were made with permission from the publishers.



## **Part I**

# **Introduction**



# Introduction

People organise in groups, whether deliberately, in circles of friends, an organisation or a society; or less consciously, in artificial social categories, perhaps based on social class or skin colour, or cultural groups, originating from common language or customs. The purposes can be of shifting character. Depending on type of aggregation, groups allow us to share knowledge, divide risks as well as specialising in different labour, and perform tasks that involve several people. These effects pertain to dynamics within the group. However, most of the groups we associate with do not entail the grand coalition of everyone in a population, but rather, we cluster into separate groups and maintain boundaries between them. These boundaries may influence us even when they are not deliberately chosen, and may emerge from practical barriers, such as language, or even arbitrary differences, such as colour of shirts.

We can thus ascribe values to group membership, also when we do not know the person we are interacting with, even when we have not chosen our groups. Instead of finding ways to co-operate with everyone across borders, the presence of other groups along with perceivable group markers cause widespread phenomena such as ingroup favouritism (sometimes along with outgroup derogation), competition and polarisation of opinion.

We have arrived at the main questions of this thesis. How are people's behaviours affected by social categories and cultural groups? More specifically, how and why do the aforementioned phenomena emerge in large groups where you do not know everyone?

The questions will be investigated mainly through mathematical models, both with a payoff-based game-theoretical approach and in the form of learning or copying models.

## Structure of the Thesis

The thesis consists of two parts. The first part will give an introduction to the field, the main research questions and methods for investigating them, together with a summary of the six research papers presented in the next part. The second part consists of the papers, each individually contributing to the field by addressing specific questions.

This chapter will start out with a theoretical background to the main research questions to clarify them and to underpin their importance. Following this, we will turn to the methods used: mathematical modelling and behavioural experiments. The subsequent section will present previous models

within the topic of the thesis in general and models addressing the questions presented here in particular. Finally, the contributions of the research papers in the second part will be summarised.

## Background on Cultural Groups

We will start out with explaining the mechanisms that are triggered when people interact with several groups, of which they are a member of one, in further detail, and what evidence there is in the literature. First of all, however, the concept of a group needs to be clarified.

### What Is a Group?<sup>1</sup>

Is it reasonable to name an aggregation of people on the 17.36 train to Gothenburg a group? Students at a lecture? Expatriates from the town of Herrljunga? Football team supporters? The first aggregation is not likely even to recognise each other outside the train, even less to offer each other special treatment, while the last one may trigger cheerful greetings among strangers wearing the team sweatshirt. A definition given by Shaw (1981) excludes the first one, with a group being “two or more people who interact and influence one another”. To clarify that group membership has at least a potential to influence people’s behaviour, we should add that they think of themselves as ‘us’ in contrast to ‘them’ (Turner, 1987).

McGrath (1984) classified aggregations of people into different categories: unorganised (e.g. an audience or a crowd), structured (society), designed (organisation, friendship group), artificial (social category, based on, e.g., sex or skin colour) and patterned (culture, based on common language or customs). The first one falls out from the definition of groups and the second and third one are deliberately designed, often for a purpose. The focus here will be on artificial social categories and cultural groups; those that you do not choose yourself and that can often be recognised. It is not obvious that artificial categories will turn into groups – ethnic identity depends on whether individuals embrace and act on it (Barth, 1969). We will see that even artificial signals may trigger discrimination, and turn artificial aggregations into cultural groups, in that a set of common norms evolves.

### Mechanisms with Several Groups

Arbitrary but perceivable group signals can trigger group sentiments and stereotyping where people become associated with their group and evaluated from the group signal rather than personal characteristics. This in turn

---

<sup>1</sup>A mathematician may answer that it is a set together with a binary operation, satisfying properties of closure and associativity, and containing identity and inverse elements. In this thesis, however, groups are sets of people with the more fuzzy properties defined in the section.



promotes mechanisms such as ethnocentrism, that is, giving preferential treatment to fellow group members; competition; and polarisation, where members of opposing groups abandon neutral or intermediate positions, resulting in a greater distance between groups.

### **Ethnocentrism**

Already in 1906, Sumner wrote about the division of 'we-groups' and 'other-groups', and coined the term *ethnocentrism*, which referred to "comradeship and peace with the we-group" as well as "hostility and war towards other-groups". While ingroup ('we-group') favouritism and outgroup ('other-group') derogation need not correlate (Brewer, 1999), there is ample evidence for discrimination based on group membership in the form of preferential treatment towards the ingroup, both from field studies and laboratory experiments (Brewer, 1976; Kramer and Brewer, 1984; Yamagishi and Mifune, 2009). Ethnocentrism will here be used interchangeably with the term ingroup bias/favouritism, meaning that people discriminate based on group membership in favour of their own group, whether or not this will entail outgroup derogation.

What does it take then, for ethnocentrism to be triggered? In a seminal work, Tajfel (1970) let participants complete a computer task, and then divided them into two groups based on their scores. Given two lists of anonymised identification numbers together with scores, one with participants scoring similarly to themselves and one with participants scoring similarly to each other, but different from yourself, people were asked to allocate money to people on the lists. Participants then allocated more money to those scoring similarly. This study was followed by one where participants were divided into groups based on aesthetic preference (for paintings), where participants preferred to maximise the money allocated to an ingroup member in a pair of anonymous people from the respective group over maximising their joint profit or dividing equally (Tajfel et al., 1971). It has been shown that people are even willing to maximise relative ingroup gain over absolute ingroup gain in this setting (Brewer and Silver, 1978).

The method for investigating minimal conditions for group discrimination to occur has been termed the *Minimal group paradigm* and has led to discrimination for several other arbitrary distinctions between groups in various settings; for example, showing that people are more willing to take risks of potential mutual advantage towards ingroup members, with group formation based on coin flips (Ahmed, 2007).

The triggering of group sentiments from minimal cues seems to be, at least partly, automatic and subconscious. Experiments show that positive traits are more easily implicitly associated with minimal ingroups (Oftten and Moskowitz, 2000). Furthermore, ethnocentrism is promoted by oxytocin, a hormone and neurotransmitter that makes people more co-operative, generous and trusting of others. These effects were recently found to apply only to ingroup

members (De Dreu et al., 2011). Given oxytocin, participants associated their ingroup with more positive values and were willing to allocate more money to them, while behaviour towards the outgroup was mostly unchanged or slightly more negative.

The case for ethnocentrism is strong and it applies to various situations, from pure sentiments in tasks that do not affect economic gain to strategic situations. What needs to be understood is why people are willing to take on costs for strangers based on crude and symbolic group distinctions and why we would trust someone more only due to shared symbols.

## Competition

In experiments in the minimal group paradigm, participants found it more important to maximise the difference between the payoff of fellow group and outgroup members than maximising ingroup payoff (Brewer and Silver, 1978). Indeed, most people have probably experienced the competitiveness that often arises between groups in some form, from friendly rivalry to large-scale conflicts escalating to wars.

Some authors focus on the outgroup hostility side of ethnocentrism and suggest it may be explained by competition in the meaning of conflicts, with the purpose of exploiting other groups (Kurzban and Leary, 2001). Considering, however, that ingroup love and outgroup hate do not correlate systematically (Brewer, 1999) and that experiments show mostly the former (Otten and Moskowitz, 2000; De Dreu et al., 2011; Yamagishi and Mifune, 2009), it seems that conflicts may not be at the core of ethnocentrism. Nevertheless, people do partake in costly group conflicts, and these processes await further investigation, but conflicts are out of the scope of this thesis.

A less confronting form of competition is group selection, the idea that traits can spread when they are beneficial to the group rather than the individual. This idea has gained little support in biological evolution (see e.g. Dawkins, 1976) – genes leading to self-sacrificial behaviour cannot be favoured by natural selection, except in groups of kin where the ‘altruistic’ gene is shared. However, it has been argued that it may be at work in cultural evolution (Soltis et al., 1995; Henrich, 2004), given cultural differences between groups. Cultural norms that are beneficial to a group may cause its members to become more numerous than those who employ less successful norms, even when a norm may not be beneficial to an individual that holds it, were it not for the enforcement mechanisms of the norm. Group selection has been employed to model group discriminative behaviour (Wilson and Dugatkin, 1997), but these explanations span far beyond the minimal group setting of interest here. Group selection also imposes restrictions on migration and how large within-group selection can be compared to between-group selection.

Some cultural traits only apply to large-scale societies, such as democracy, which is hosted by and spread between entire countries (Lindenfors et al.,

2011). Such groups (in this case, structured), however, are also out of the scope here.

Group selection and competition will here be employed only in the situation where there is no conflict between individual and group interest. When there are several groups, they may arrive at different evolutionarily stable strategies, from which selection will act on the different strategies.

## **Polarisation**

Group polarisation is an enhancement of initial tendencies of a group's members, towards making average tendencies of the members more extreme. While allowing people's attitudes to converge within the group, it leads to further distance between groups, and to segregation. Myers (2008) quotes several studies, giving examples of how initial opinions become more extreme after group discussion, how initially modest gender differences self-segregate into all-male and all-female groups, how neighbourhoods in the US become increasingly conservative or liberal, and how competition and mistrust are worsened when the players in laboratory studies are groups. Polarisation strengthens the 'us' versus 'them' dichotomy and could be one of the processes leading to ethnocentrism.

Polarisation may be triggered by a *confirmation bias*, a tendency to favour information that confirm your preconceptions. Such a term was coined after a study where participants were asked to find a rule for a number series (Wason, 1960). The rule was simple: any sequence of increasing numbers. Subjects were first given the sequence 2, 4, 6 and were to give suggestions for new sequences until they were confident about the rule. Most participants started out with the hypothesis "a sequence of even numbers" and then, instead of trying to falsify it, they only tried even numbers, thus only looking for evidence confirming their initial hypothesis, ultimately leading them to the false conclusion.

Several studies show that the confirmation bias strengthens group polarisation. In one such study, subjects supporting and opposing capital punishment read two purported studies on the deterrent efficacy of the death penalty, one with pro- and another with antideterrent information (Lord et al., 1979). Subjects were more convinced by the study favouring their initial opinion, questioned the opposing evidence more, and reported having shifted their belief more towards their initial inclination.

Are ethnocentrism and polarisation mutually enforcing mechanisms? In the words of Barth (1969), "[t]he identification of another person as a fellow member of an ethnic group implies a sharing of criteria for evaluation and judgement. It thus entails the assumption that the two are fundamentally 'playing the same game'". Does the confirmation bias inevitably lead to polarisation, or can it lead to common values, when the motivation for mutual understanding is sufficiently high?

# Mathematical Modelling

There is compelling empirical evidence for several psychological and behavioural effects in the interplay between groups. Meanwhile, little seems to be known about the underlying mechanisms: What caused the emergence of ethnocentrism? How important are components like population structure and type of interactions? What are the possible outcomes in terms of polarisation and common norms of different learning biases, such as a confirmation bias? These questions will be investigated using mathematical models.

A model is a simplification of reality, for the purpose of making some aspect of reality more manageable and conducive to studying. It is a description representing selected parts of reality, containing essential properties of the phenomenon under study. A well-known example that most people come across is maps. A map is a heavily simplified version of the world concentrating on parts of its geography, but very useful for orientation in it. What should be included depends on the application. A traveller may find a version of a city not drawn to a scale depicting the metro network useful, while orienteers are more interested in natural terrain and contours. Neither of them would like the map to include traffic lights; it would only clutter the image.

By expressing models in the language of mathematics, its assumptions are clearly and unambiguously formulated, and the effects of the assumptions can be studied by manipulating symbols, providing clear, logical consequences. By comparing the implied consequences to observed phenomena, then, we can gain insight into the soundness of the assumptions and their individual significance.

A typical modelling procedure starts out with a theoretical model based on an observed phenomenon. From the model, predictions can be derived that are then compared to empirical data to validate the model. If the model does not make sufficiently satisfying predictions, then it needs to be modified, make new predictions and be revalidated. The model can also give insights into how to test hypotheses on the phenomenon empirically. If it passes the test, then the model can be used for future predictions. (See e.g. Gerlee and Lundh, 2012, for an illustration.)

The complexity of the model is usually a trade-off between predictive power and complexity. The more complex the model, the more realistic and usually more accurate it is. The more simple the model, the easier it is to understand and analyse. A key principle, Occam's razor, is that, when faced with two models of equal predictive power, the simplest one of them is preferred. A sound modelling approach could thus be to start with a simple analytically tractable model and add assumptions until its predictions are sufficiently accurate. With analytical solutions, we have full control over the parameters involved and how they affect the process, and the model can be generalised to other situations. If necessary, the model can be extended further and be evaluated using simulations. By resorting to simulations, we

can no longer have an exhaustive evaluation of the whole parameter space and we lose the full control over their individual significance. However, a simulation model is easier to handle if it can be an extension of, and thus based on, a simpler analytical one. This approach is taken in Papers II and IV. Sometimes simulations can be used to show that there is little difference between a simulation-based model and an analytically tractable counterpart, as in Paper V.

This thesis is concerned with modelling behaviour contingent on group membership in human encounters between individuals. When all I know about you is your group membership, to me you represent the whole group. Since my behaviour is not contingent on the number of people I meet, the models here can be restricted to two-person interactions. For such models, there are two possible update rules for the result of an interaction.

- Our group memberships may influence my behaviour.
- Your behaviour may influence my future group membership.

Mathematical models occur in many forms, and what kind of models are suitable depends on which of the update rules is applicable to the situation in question. This thesis contains models based on neutral theory, a basic null model that sometimes provides good predictions in itself; game theory, adding utilities to agents; and agent-based models, keeping track of individual agents.

## Neutral Models

Even though adaptive mutations are important in evolution, neutral theory states that random drift of selectively neutral mutants may also be of major importance, observing that neutral mutations do take place at very high rates in molecular evolution (Kimura, 1968). Models without selection can explain much of polymorphism within and divergence between species. Neutral theory has also found its way into the realm of cultural evolution, where traits like names, pottery designs and cited patents follow power laws that can be explained by a simple model of random drift (Bentley et al., 2004). These models are based on random copying, where individuals choose cultural parents at random to imitate (possibly subject to copying errors), where the imitated trait does not affect the success of the individuals in any other respect. Neutral models in their own right may serve as good predictors of observed phenomena, and when this is not the case, they can function as null models to build upon and with which we can compare predictive power to measure the significance of further assumptions.

Two of the most common neutral models are the Wright–Fisher and Moran models, and variations upon them (see e.g. Ewens, 2004, for an elaborate account). In both models, the population size is fixed at  $N$  individuals. In the Wright–Fisher model, at each new generation of offspring, the whole

population is replaced by  $N$  new individuals where each individual in the offspring generation picks a parent at random from the previous generation, and inherits the traits of the parent. Everyone has the same fitness, meaning that all individuals are chosen with equal probability. From this model we can easily derive the probability to obtain  $k$  copies of a trait that has frequency  $p$  in the next generation to be

$$\binom{N}{k} p^k (1-p)^{N-k},$$

from which can derive expected value and variance to predict the process, which can then be tested against experiments. A reasonable extension is to add copying errors, mutations. This can be done by changing the last rule such that with probability  $1 - \mu$ , an individual inherits the trait of the parent, and with probability  $\mu$ , it will bring a new trait into the population.

The Moran model instead assumes overlapping generations, such that only one individual is replaced at a time, choosing its parent from the previous time step. This means that in each time step, the number of copies of a trait will stay the same, or it can increase or decrease by one, making it easier to find analytical solutions than in the Wright–Fisher model (while the latter model has a lower complexity in simulations, since the former model requires  $N$  time steps to replace the same number of individuals). It has been shown through simulations that the models often provide similar results (Strimling et al., 2009; Paper V), something which has proven useful. Bentley et al. (2007) showed, through simulations, that the Wright–Fisher model is consistent with an empirically verified equation for the turnover rate on toplist of pop songs, baby names and dog breeds. In Paper V, we use the Moran model to prove the equation (with a slight modification) analytically.

Since the research questions targeted here deal with groups without intrinsic selective advantages, neutral models are suitable for interactions affecting your group membership, and are used in Papers V and IV. In the latter, we start out with a neutral model as a null model, show that it cannot explain the observed phenomenon, and make the updating contingent on preferences by introducing a confirmation bias.

## Game Theory

Game theory is the study of strategic decisions. A game is a set of strategies to choose from and where the payoff to a player depends on what all the players choose. Contrasting to neutral models, different outcomes are of different values to individuals. Game theory is concerned with finding (Nash) equilibria in strategic structures, that is, outcomes where no individual would gain from deviating from its chosen strategy, expected to be stable solutions to how people will behave in the modelled situation.

Given its nature of predicting how people will act when accounting for how others act, game theory is more apt for the first update rule in this

thesis, using people's group membership as an indicator of how they will act, and thus affecting my behaviour.

Starting out as mathematical theory of economic behaviour (von Neumann and Morgenstern, 1944), the discipline has become increasingly useful in other fields, such as biological and cultural evolution, and political science, and has come to incorporate concepts such as bounded rationality and evolutionary stability, and theories are tested in behavioural experiments.

### **Classical Game Theory**

Game theory initially only dealt with rational agents optimising their utility. The rational agent generally has full information on the game it is playing and not only its own, but also other players' utilities, and it considers strategic analyses of the other players when determining its own choice. The rational agent is a highly idealised assumption that often does not correspond well with how people actually behave. Paper III constitutes one example, where people commonly do not choose the socially optimal strategy that gives both players the highest possible payoff, which means that the players either are not rational, or they assume other players not to be. See, for example, Camerer (2003) for more examples.

Simon (1972) suggested how risk, uncertainty and incomplete information about alternatives could be included in game theoretic models, founding the concept of bounded rationality, and that social preferences could be incorporated into players' utility functions, thus opening for formal models that could be designed to better reflect human psychology, and also, in the face of complex problems, enabling for solutions to be found analytically.

In Paper VI, bounded rationality is introduced as a consequence of social preferences. The paper deals with matchings, game theoretic problems where what is of main concern is not what set of strategies are stable, but rather what kind of coalitions of people will form and be stable, such that no two people would rather match up with each other than remain in their current state. See Roth and Sotomayor (1992) for a comprehensive introduction to two-sided matchings. In Paper VI, which builds on one of the games presented there, people do not only care about with whom they are matched, but they may also be spiteful to other players in their own group and act accordingly. This leads to unforeseeable consequences that are not only hard to derive analytically, but puts so high demands on computational powers that people are not expected to perform the full calculations. The concept of bounded rationality here serves to make the model both more realistic and analytically solvable.

### **Evolutionary Game Theory**

On the other end of the spectre, we find evolutionary game theory, which assumes no rationality at all. Instead of utilities, agents have a fitness that

determines how many replicas of themselves they are expected to make. Agents have an inherent strategy and its success is determined by how many offspring the agent produces. With time, agents are usually replaced, so an evolutionary game is a competition between strategies rather than individuals.

With its gene's eye view and the focus on evolutionary stability, evolutionary game theory was first launched to make the game theoretic approach more apt for biology (Maynard Smith, 1982), but it is also useful in a cultural evolutionary framework and has become of increasing interest in other disciplines, also economics (Weibull, 1995).

An evolutionarily stable strategy (ESS) is a strategy which, if adopted by a population of players, cannot be invaded by an alternative strategy that is initially rare. More specifically, an ESS is a Nash equilibrium, that is, one where no other strategy performs better against the ESS than the ESS itself, with the added condition that the ESS performs better against all other strategies than the strategies do against themselves.

Since agents are not rational, it is not evident how they arrive at an ESS, or which one will attract the population in the event of several ESSs. While evolutionary stability focuses on the role of mutations, replicator dynamics rests on selection, represented mathematically by a system of ordinary differential equations. The replicators are the pure strategies, whose success is determined by the current (and changing) population state. Weibull (1995) gives an introduction to replicator dynamics, and it is used in Paper II, providing an example of how it can be used and giving a derivation of the replicator equation, which in its general form can be expressed as

$$p_i'(t) = (\pi_{ip}(t) - \pi_{pp}(t))p_i(t),$$

where  $p_i(t)$  is the proportion share of strategy  $i$  in the population at time  $t$ , and  $\pi_{ip}(t)$  is the fitness payoff given the population state vector  $p(t)$  and  $\pi_{pp}(t)$  the average payoff in the population.

## Behavioural Game Theory

Behavioural game theory is the study of how people actually behave in strategic situations where the individual's outcome depends on the choices of all players, which can often contrast considerably to theoretical predictions (Camerer, 2003). Differences may be due to, say, limited recursive reasoning (what do I think that she thinks that I think that she thinks that I think ...) or social preferences, and even a rational actor needs to account for the possibility that other players are not perfectly rational.

Theories predicting behaviour can be verified, modified and new ones created through experimental testing, most commonly in a laboratory environment, where people are invited to make decisions in strategic situations and receive proper economic payment accordingly. In Paper III, we investigate a



game where risk aversiveness makes people perform worse than had all players been (known to be) rational and test experimentally whether particular group signals will affect the behaviour.

## **Agent-Based Models**

Many games are too complicated to allow for analytical solutions and sometimes the simplifications needed for finding solutions deviate too much from a realistic setting by removing essential parts. In these cases, we may revert to simulations, that is, running computer programs that follow the rules of the model.

Agent- (or individual-) based models assume autonomous strategic actors of the game that interact in some interaction topology with other actors, each having its internal state of decision-making heuristics and possibly learning rules, and through simulations we can learn how these microscale assumptions can give rise to population level phenomena. The model can be turned into an evolutionary model by adding a reproduction phase, where agents produce offspring, which inherit their behaviour, in proportion to their success.

Agent-based models can be designed to include an arbitrary number of real-world premises. Agents can represent real human beings or groups using behavioural heuristics known from human psychology and living in a landscape following some real topological conditions. Meanwhile, we may be faced with an insurmountable amount of design choices for which it is hard to determine parameter values and that do not allow for an exhaustive search through the parameter space. And even if we manage to design a model with high predictive power, it is hard to pinpoint what are the driving processes, and thus to learn from it and draw the right conclusions. Paper I illustrates how, in a well-cited agent-based model, what drives the results has little to do with the phenomenon that the model was intended to explain, leading to unreasonable conclusions.

Combined with analytical results, agent-based models are used in this thesis to provide more detailed predictions. In Paper II, simulations indicate more specifically which games form the breeding ground for ethnocentrism and verifies the robustness of the model against random drift and mutations. In Paper IV, agent-based simulations reflect and incorporate demographic data to verify a model empirically.

## **Previous Models on Intergroup Interactions**

There is quite an extensive literature on empirical research on intergroup interactions, not least within experimental psychology, and references are given in the Background section. Meanwhile, modelling intergroup phenomena seems to be in the bud, with some first steps taken that will be outlined

here. First, we will have a look at models aiming at providing a general framework and then we will delve further into specific models addressing the phenomena laid out here.

## Overview

Groups may be seen as a coalition of players, living creatures recognising each other from common phenotypes or people sharing cultural norms. Not all approaches cover the “us interacting with and influencing each other versus them” definition, but previous modelling approaches from several disciplines may be useful for developing new models, including game theory, biology and cultural evolution. Note that approaches are often interdisciplinary and that labellings intercept, so the following sectioning are but crude distinctions.

## Game Theory

Game theoretical models on groups are often concerned with the formation of a coalition of players and their mutual co-operation. Levin (2009) has edited a recent volume of chapters by several of the most renowned researchers in the field. These, however, mainly deal with co-operation in  $n$ -player games, such as the tragedy of the commons, which is out of the scope here. Young (1998) similarly concentrates on co-operation within the group, but starts from the microlevel with two-player games of boundedly rational players and shows how this can lead to complex institutions. The book treats concepts like risk dominance and local interactions, which are relevant here, but in another setting.

Skyrms (2004) summarises much of his work and draws an outline for future modelling focusing on the stag hunt game, a game with one payoff dominant and another risk dominant equilibrium. His example of such a game is rowing a boat (with one oar each): if we both row, then we will get forward, while if no one rows, then we end up getting nowhere, which is still better than if only I do it, because then my effort will be wasted on going round in circles. Contrasting to the prisoners’ dilemma, there is a co-operative equilibrium, making such an outcome rationally possible. The work focuses on location, signals and association. The first part on local interaction models gives an alternative to the replicator dynamics enabling gradual shift towards the optimal equilibrium, an assumption that can, however, lead to undesirable results, as discussed in Paper I. The next part on signals sketches out how they can transform the basins of attraction and that cheap talk signals become correlated with behaviour in the game, something which is elaborated upon in the group sense in Paper II. Finally, a model is given where agents update interaction probabilities with other agents based on the success of previous interactions, which, however, leads back to group formation and into the realm of designed groups.

Instead of local interactions, Tarnita et al. (2009) let the groups form the structure of the population. In their model, agents have several set memberships and play a prisoners' dilemma in two-person interactions with people belonging to the same set (communicating more often with individuals with whom they share several sets), finding conditions for when co-operators are more abundant than defectors, in terms of how large the benefit-to-cost (benefit from you co-operating, cost from me doing it) ratio needs to be compared to how many sets and strategies two randomly chosen individuals have in common on average. The model is also generalised such that agents may co-operate if they have a certain minimum number of sets in common.

## **Evolutionary Biology**

While models of group behaviour among animals generally deal with the emergence of sociality – why do animals live in groups at all? what are the trade-offs between increased protection and sharing of resources? – and co-ordination problems like bird flight in flock, some modellers have turned their attention to the green-beard effect. The idea of a green beard was suggested by Hamilton (1964*a,b*) and so named by Dawkins (1976). A green-beard is a perceptible trait that can help individuals identify other individuals with a common gene and give preferential treatment to these. Using the green-beard as a signal for group membership, models may be useful also in the study of human group discrimination.

A well-cited model was presented by Riolo et al. (2001), where agents have a visible marker on a continuum and co-operate with sufficiently similar others. The number of offspring is determined by the success of the interactions and offspring inherit marker and tolerance level, subject to mutations. The result is that co-operation is maintained within small tolerance levels, but as tolerance levels increase due to drift, mutants with lower tolerance levels invade and form new co-operative clusters consisting of their offspring. The model in its original form seems to have little bearing on human ethnocentrism, with signals being but proxies for kin recognition, and it has been shown that co-operation relies on the fact that agents are not given the possibility of co-operating with no one (tolerance levels cannot be negative) (Roberts and Sherratt, 2002), but similar models have been developed in its wake, directed towards finding conditions when tag-based signals may induce the emergence of co-operation in a prisoners' dilemma.

Firstly, a simplified version of the model has been analysed, with two tags and binary tolerance levels in a well-mixed population (Traulsen and Schuster, 2003). The authors focused on explaining the observed oscillations, and showed that adding drift towards more tolerance to the replicator dynamics gives rise to an alternating dominance of both groups.

Jansen and van Baalen (2006) showed that tag-based co-operation may withstand agents defecting towards everyone in a spatial setting. In the model, agents can co-operate or defect with their own colour in a spatial

setting. Co-operation can emerge also without tags in a viscous environment, but colours allow the population to be less viscous than in the case of blind kin selection.

Traulsen and Nowak (2007) showed that an alternative to spatially structured populations is sufficiently small well-mixed populations. The result is again oscillations, with single co-operators with a different tag first multiplying due to random drift (as the strategy is neutral in relation to the current population) and then invading and dominating, until same-coloured defectors emerge and the process is set anew. This contrasts to spatial systems, where persistent co-operation based on tags is possible. Another possibility is that mutation rates differ between tags and strategies, such that mutations between tags are frequent and mutations in behaviour are rare. Under such circumstances, co-operators are more abundant than defectors in a well-mixed population also for relatively low benefit-to-cost ratios (slightly larger than two) (Antal et al., 2009).

Masuda and Ohtsuki (2007) brought the model of Riolo et al., with the added ability of complete defection, closer into the realm of culture by adding a direct reputation system, implemented as public tolerance levels, which may be perceived incompletely. Agents may be willing to donate only to individuals that are expected to donate to other individuals similar to themselves. If the tolerance level is completely known, then everyone will co-operate, and if it is incomplete, then tag-based co-operation may emerge.

## **Cultural Evolution**

While biology mainly deals with groups and signals on the level of genes and classical game theory with coalitions that are deliberately formed for reasons of self-interest, many of the groups we adhere to are not formed on genetic grounds, nor did we actively choose to be part of them. For reasons such as isolation and segregation different aggregates of people may develop different sets of norms, be it from reasons of intentional design or simply drift, and, in parallel to green-beards in biological evolution, these groups may come to be symbolically marked, identified and maintained by fads, dialects and other markers (Boyd and Richerson, 2005; Lindenfors, 2011). Group discrimination may then emerge as a result of shared beliefs about norms or knowledge about what to expect from each other, or as a means of exploiting less fortunate groups. Previous models include both these driving forces. However, the latter assumption seems not to be necessary for discrimination to emerge, as laid out in the Background section.

Bowles et al. (2003) elaborate upon the Price equation to quantify a trade-off between within- and between-group selection. In a prisoners' dilemma, within-group selection acts against co-operative behaviour towards fellow group members, while between-group selection promotes it. The latter requires small within- and large between-group variance, which, it is suggested, can be maintained through institutions at the group level acting to reduce

the former. Through agent-based simulations, the authors then show how cooperative behaviour may evolve when individual behaviours are transmitted genetically to offspring, while losing groups in group competition adopt the winning group's (the one with the highest total payoff) institutions through cultural transmission. Choi and Bowles (2007) proceed with group competition, introducing intergroup hostility as a driving force for preferential treatment towards the ingroup, a line of thought paving the way for various models that will not be elaborated upon here.

Boyd and Richerson (2005) concentrate on the formation of ethnic tags and how behavioural differences can become ethnically marked. The idea is that "[n]onrandom interaction increases correlation between arbitrary markers and locally adaptive behaviors. This, in turn, makes markers more useful, setting up a positive feedback process that can amplify small differences". There needs to be some initial weak correlation between marker and group membership, with the groups being partly isolated. McElreath et al. (2003) elaborate on this by modelling interactions as co-ordination games, where it is in the interest of both parties to do the same as the other. Individuals have a marker and a behaviour in the game, and those who do well are imitated. People may interact more with those who share their marker. With segregated groups (and some migration in between), common combinations of markers and behaviours will spread in the respective groups, and migration will increase the co-variance between marker and behaviour within groups.

Another approach to the spread of group markers is to have a macroscopic view on the population level. Kandler and Steele (2008) use a reaction-diffusion system to study the dynamics of language competition. Language is an efficient group marker in that only those exhibiting the highly acquired trait are able to communicate with each other. In physics, reaction-diffusion systems model the concentration of particles in space with respect to time under the influence of reactions and diffusions. Diffusion is the spread of particles over a surface (from regions of higher to those of lower concentration), which can be used in cultural evolution to represent the spread of new ideas in spatially structured populations. In reactions, particles are transformed into each other, in our case leading to growth of the cultural trait under study. For languages, the system models a competition process in which the number of speakers of a language depends on birth-death processes and migration, and language shift, where shift is often directed from a locally used language to a more prestigious wide-spread one. Kandler and Steele found that co-existence of languages in such a competitive situation is possible with spatial heterogeneity. Kandler et al. (2010) then studied a real-world example of Gaelic-English competition in Scotland and found that co-existence is possible also with diglossia, where bilingual speakers use different languages in different social domains. In this thesis, shifting processes are instead studied from the individual-based level, with dyadic interactions,

and applied to the very domain of language shift and emergence in Paper IV. In Paper V, group markers are subject to purely random copying, leading to continual shifts of dominant groups.

## Current Models

Some previous models occur in this thesis to be scrutinised, analysed or extended, or they bear resemblance to models constructed here for the study of specific phenomena. The following are models more directly related to the papers of the thesis. The origins of the models span over the various disciplines presented here, relating to green-beard effects, symmetric two-player games, linguistic interaction models, neutral theory and matchings.

**Green-beard effects** A model similar to that of Riolo et al. (2001) and other models presented here in that context, but that explicitly claims to deal with ethnocentrism, was presented by Hammond and Axelrod (2006*b*). The model allows for tag-based discrimination in either direction by equipping agents with a group marker, one strategy, co-operate or defect, towards the ingroup, and one strategy towards the outgroup, in a prisoners' dilemma. The outcome of the game affects the agents' potential to reproduce. The only stable strategy in a prisoners' dilemma is to always defect, so the model requires further assumptions for any discriminative behaviour to establish. Here, agents populate a toroidal lattice where they communicate only with their immediate neighbours. Agents are born with a marker and their strategies, and reproduce onto neighbouring patches.

The model is successful in the means of providing a breeding ground for ethnocentric agents, which come to dominate the population. As was acknowledged already in a previous publication, however, the strict spatial structure is both a necessary and sufficient condition for cooperation to emerge in the model (Hammond and Axelrod, 2006*a*) and it causes such high degrees of relatedness in agents' neighbourhoods that the model was first launched as an illustration of the "armpit effect" (distinguishing strangers from unfamiliar kin) (Axelrod et al., 2004). Thus, the model relates to the green-beard models presented here. Paper I scrutinises the model to find what it actually says, and Paper II instead assumes well-mixed populations to investigate effects beyond kin selection, and instead introduces other games into the strategic interactions.

**Symmetric two-player games** With models of ethnocentric behaviour concentrating on the prisoners' dilemma as the underlying strategic structure of interactions, Paper II investigates what other games may lead to. The analysis takes into account all symmetric two-player games with two strategies,

that is, all games that can be described by the matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where  $a_{ij}$  is the payoff I will get from choosing strategy  $i$  when you choose  $j$ . With respect to evolutionary stability, there are only four such games: prisoners' dilemmas, harmony games, and co-ordination and anti-co-ordination games (Weibull, 1995). If you choose strategy  $j$ , then what is interesting to me is which is largest of  $a_{1j}$  and  $a_{2j}$ .

With  $a_{11} < a_{21}$  and  $a_{12} < a_{22}$ , we have a prisoners' dilemma, where the rational choice is always to choose strategy 2, commonly termed defect. This is actually a nonstandard definition that entails the standard definition. In order to turn the game into a dilemma, we need also assume that  $a_{11} > a_{22}$ , such that the result is that we will be worse off than had we both chosen strategy 1, co-operate. The examples of prisoners' dilemmas are numerous, such as competitors on a market, to name one. I will attract more customers by lowering my prices below yours if you sell identical products, so we will race down towards no profit margins, while we would have been better off having the same high prices and half of the customer pool.

The very opposite of these games are harmony games, where  $a_{11} > a_{21}$  and  $a_{12} > a_{22}$ , and we will both co-operate. We would both like to go to the pictures, either alone or, preferably, in company. Of course, in the end, we will both be very happy finding each other there.

If  $a_{11} > a_{21}$  and  $a_{12} < a_{22}$ , then I will prefer to do whatever you do, such as driving on the right-hand side of the road. In the opposite case,  $a_{11} < a_{21}$  and  $a_{12} > a_{22}$ , I want to do the opposite. If you play the guitar, then I would like to sing, and vice versa, in order to produce the most interesting performance.

Even if these classes of games cover all ESSs, in reality, some subclasses of these games may be more interesting than others. Changing the pay-offs alters the basins of attraction when there are multiple equilibria, and can induce competition between payoff and risk dominance. A strategy is risk dominant if it maximises my expected payoff assuming you randomise among strategies, that is, strategy 2 is risk dominant if  $a_{21} + a_{22} > a_{11} + a_{12}$ . A payoff dominant strategy is one where no other equilibrium offers any of us a better payoff.

In Paper II, it turns out that the subclasses of stag hunt and hawk-dove games are especially interesting with respect to group discrimination. The stag hunt is a co-operation game where the payoff dominant strategy differs from the risk dominant one. The hawk-dove game is an anti-co-ordination game where  $a_{12} > a_{22} > a_{12} > a_{11}$ . I always prefer meeting a dove, and being one if and only if you are a hawk. The former game has been exemplified as rowing a boat with one oar each. In a similar manner, the latter game may be compared to cycling on a tandem bicycle. As long as you pedal, I prefer not to, but those who pedal in concert will put less effort into it than a single pedaller, while pairs who do not pedal at all will remain at a standstill.

**Language models** Paper IV models the process of merging groups with different norms, with the outcome being a common set of norms, a typical situation for language. The model presented there resembles an earlier model called *the Naming Game* (Steels, 1996, 1997). In that model, agents develop their own vocabulary to map words to meanings. The agents then communicate in pairwise interactions taking on the roles of one speaker and one hearer. The speaker randomly selects a topic and encodes it with the word that has been most successful in previous interactions of the speaker concerning the present topic. Should the speaker lack words for encoding the topic, she will invent one. If the hearer does not understand the word, then he might include it in his inventory for future reference. Simulations have shown that a globally shared vocabulary can emerge under such circumstances.

In our paper, we employ a similar model that is analytically tractable, where individuals instead have preferences represented as probabilities for using a set of words depicting a certain meaning which is updated with respect to whether communication was successful.

**Neutral models** The Wright–Fisher and Moran models are presented in the section on Neutral Models in the Mathematical Modelling section, and are compared through simulations in Paper V. The paper provides an in-depth analysis of the latter model with respect to how often popular traits are replaced.

**Matchings** The *stable marriage problem* is one where people are divided into two equally sized groups, commonly stated as monogamous heterosexual men and women. People of either sex have ranked all members of the opposite sex in order of preference, and the problem is to find pairings of members of the two groups such that no two people of the opposite sex would rather have each other than their current partners, in which case the pairings are *stable*. There is an algorithm that guarantees that everyone gets paired up in a stable marriage. The algorithm assumes that one group proposes and the other group accepts or rejects, and the resulting stable matching is the best possible one for the proposers, while it is the worst possible one for the acceptors. Paper VI deals with a related game called the *assignment game* (Shapley and Shubik, 1972), where there is also money involved. Associated with each possible partnership is a nonnegative real number, and the outcome is a matching along with a payoff vector. See Roth and Sotomayor (1992) for a more detailed account.

## Papers of the Thesis

The thesis deals with both types of interaction models presented in the Mathematical Modelling section: agents either engage in group-contingent behaviour, or behaviour affects their group membership. The former apply to



group discrimination and ethnocentrism, while the latter relate to polarisation and group competition (without intergroup hostility). The final paper reverses the perspectives, with some agents being hostile towards certain individuals within the group when they form pairwise relationships between groups.

## **Group Membership Affects Behaviour**

In the first type of models, used to investigate the evolution of ethnocentrism, agents consider the group membership of themselves and their partner when deciding how to act.

**Paper I** Previous models of ethnocentrism have mainly focused on the prisoners' dilemma, with the aim of finding conditions leading to co-operation towards fellow group members and defection towards others. However, since it is always rational to defect in such a dilemma, irrespective of group membership, further assumptions need to be made. One such assumption has been to introduce spatial viscosity and local interactions, as did Hammond and Axelrod (2006*b*). Paper I illustrates the pitfalls of such assumptions by investigating their model in detail through simulations, and concludes that the model is one of clone recognition, rather than ethnocentrism.

**Paper II** Instead of resorting to kin selection, Paper II investigates whether other games than the prisoners' dilemma are more conducive to group discrimination, and finds that co-ordination and anti-co-ordination games often lead to such, and mostly in favour of the ingroup. In particular, this applies most often to the subclasses of stag hunt and hawk-dove games. Group discrimination has been observed in various settings, and it is difficult for people to identify the strategic structure of interactions, so discrimination may stem from these games.

**Paper III** Except for theoretical modelling predictions, group discrimination has been shown to be triggered in behavioural experiments on stag hunt situations. Paper III takes these experiments further, and shows that negative group discrimination can be overcome by the establishment of a group reputation of trust.

## **Group Membership is Contingent on Previous Interactions**

Groups are not generally fixed at birth, and the result of an interaction may also be that you approach, or distance yourself from, the people you meet. This is true for phenomena such as polarisation of opinions and attitudes – the arguments of someone may make you align with the beliefs of that person, and the size of competing groups vary as people influence each other.

**Paper IV** As to the polarisation, Paper IV focuses on a model on how cultural groups in contact can develop common norms, rather than polarise into different norm groups, by assuming a confirmation bias. It has been assumed that such a bias leads to polarisation. In our model, we rather prove that it is a prerequisite for developing common norms, so it seems that the bias can work both ways, depending on population structure and transmission of information.

A situation where common norms is a highly desired end product is when people speaking different languages meet and are in the need to communicate. A situation that is sufficiently controlled to allow for collecting data is the evolution of a creole – a natural language developed from the mixing of parent languages – which has occurred in former slave territories. The model is empirically tested on demographic and linguistic data from Mauritian Creole, giving a good fit to the language as we know it today.

**Paper V** Groups may compete for members under circumstances that are not very hostile, such as those formed by abiding by trends. It has been demonstrated that for cultural traits such as pop albums, baby names and dog breeds, the turnover rate on a toplist of the most popular traits follows a remarkably simple expression consistent with that produced by a model of random copying. In Paper V, the turnover rate is derived analytically, given a neutral model where, in each time step, one individual leaves its current group and joins that of a randomly chosen individual.

## Matchings Between Groups

Intergroup interactions are not all about competitive encounters; as depicted in Paper II, (anti-)co-ordination and specialisation, with use of the expertise of each other, are important parts of human sociality. Under some circumstances, people would rather match up between groups. Group membership then affects whom to match up with, and possible matchings have been studied in marriage and assignment games.

**Paper VI** Contrasting to the previous models, with anonymous but group-contingent encounters, in Paper VI, people are acquainted with the individuals of the whole population. In such cases, people would commonly know those of their own groups better than those of other groups, and have more preferences (for or against, as friends or competitors, say) concerning their success. In the paper, people match up between groups, but their behaviour may be affected by hostility towards fellow group members, and all the stable matchings are found under a specific assumption of bounded rationality.

## Directions for Future Research

Some of the main results are that, in a well-mixed population, ethnocentrism may stem from other games than those under previous focus, that reputation may counteract negative discrimination, and that a confirmation bias may actually lead to consensus rather than polarisation.

The latter model can be taken further in at least two directions. The first one is to elaborate further on polarisation. There is a vast literature on such models, but none seem to incorporate a confirmation bias. This model may thus be able to give new insights on polarisation by involving a well-known psychological phenomenon. The other direction is to apply it to language change and develop it further in that relation. Firstly, it would be useful to test it in other creole settings, but it may also be relevant in more general settings. Similar models (such as the Naming Game in the paragraph on Language models) have been used to study the emergence of a vocabulary with unique and commonly acknowledged names for different meanings, but the present model has the advantage of being analytically tractable. The Naming Game has been studied only through simulations, making it difficult to understand basic underlying mechanisms.

Models of group discrimination are still in the bud, but with a growing interest. The results presented here suggest that a successful path may be to look into a broader range of games than previously, and that successful interactions may be subject to maintaining positive signals. It thus seems to be a promising line of research to elaborate further on the currently simple models of co-ordination and anti-co-ordination to get a fuller understanding of the possible mechanisms underlying ethnocentrism, and possible ways to overcome abysses between groups.

## Bibliography

- Ahmed, Ali M. (2007): Group identity, social distance and intergroup bias. *Journal of Economic Psychology* 28: 324–337
- Antal, Tibor, Hisashi Ohtsuki, John Wakeley, Peter D. Taylor and Martin A. Nowak (2009): Evolution of cooperation by phenotypic similarity. *Proceedings of the National Academy of Sciences* 106: 8597–8600
- Axelrod, Robert, Ross A. Hammond and Alan Grafen (2004): Altruism via Kin-Selection Strategies that Rely on Arbitrary Tags with which They Coevolve. *Evolution* 58: 1833–1838
- Barth, Fredrik (1969): Introduction. In Fredrik Barth (ed.): *Ethnic groups and boundaries*. Little, Brown
- Bentley, R. Alexander, Matthew W. Hahn and Stephen J. Shennan (2004):

- Random drift and culture change. *Proceedings of the Royal Society B* 271: 1443–1450
- Bentley, R. Alexander, Carl P. Lipo, Harold A. Herzog and Matthew W. Hahn (2007): Regular rates of popular culture change reflect random copying. *Evolution and Human Behavior* 28: 151–158
- Bowles, Samuel, Jung-Kyoo Choi and Astrid Hopfensitz (2003): The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology* 223: 135–147
- Boyd, Robert and Peter J. Richerson (2005): *The Origin and Evolution of Cultures*. Oxford University Press, New York, New York
- Brewer, Marilynn B. (1976): *Ethnocentrism and intergroup attitudes: East African evidence*. Halstead Press, New York, New York
- (1999): The Psychology of Prejudice: Ingroup Love or Outgroup Hate? *Journal of Social Issues* 55: 429–444
- Brewer, Marilynn B. and Madelyn Silver (1978): Ingroup bias as a function of task characteristics. *European Journal of Social Psychology* 8: 393–400
- Camerer, Colin F. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, Princeton, New Jersey
- Choi, Jung-Kyoo and Samuel Bowles (2007): The Coevolution of Parochial Altruism and War. *Science* 318(5850): 636–640
- Dawkins, Richard (1976): *The Selfish Gene*. Oxford University Press, Oxford, United Kingdom
- De Dreu, Carsten K. W., Lindred L. Greer, Gerben A. Van Kleef, Shaul Salvi and Michel J. J. Handgraaf (2011): Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences of the United States of America* 108: 1262–1266
- Ewens, Warren J. (2004): *Mathematical Population Genetics. Second edition*. Springer, New York, New York
- Gerlee, Philip and Torbjörn Lundh (2012): *Vetenskapliga modeller*. Studentlitteratur, Lund, Sweden
- Hamilton, Willam D. (1964a): The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7: 1–16
- (1964b): The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* 7: 17–52

- Hammond, Ross A. and Robert Axelrod (2006a): Evolution of contingent altruism when cooperation is expensive. *Theoretical Population Biology* 69: 333–338
- (2006b): The Evolution of Ethnocentrism. *Journal of Conflict Resolution* 50: 1–11
- Henrich, Joseph (2004): Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization* 53(1): 3–35
- Jansen, Vincent A. A. and Minus van Baalen (2006): Altruism through beard chromodynamics. *Nature* 440: 663–666
- Kandler, Anne and James Steele (2008): Ecological Models of Language Competition. *Biological Theory* 3(2): 164–173
- Kandler, Anne, Roman Unger and James Steele (2010): Language shift, bilingualism and the future of Britain’s Celtic languages. *Philosophical Transactions of the Royal Society* 365: 3855–3864
- Kimura, Motoo (1968): Evolutionary Rate at the Molecular Level. *Nature* 217: 624–626
- Kramer, Roderick M. and Marilynn B. Brewer (1984): Effects of Group Identity on Resource Use in a Simulated Commons Dilemma. *Journal of Personality and Social Psychology* 46(5): 1044–1057
- Kurzban, Robert O. and Mark R. Leary (2001): Evolutionary origins of stigmatization: the functions of social exclusion. *Psychological Bulletin* 127(2): 187–208
- Levin, Simon A. (ed.) (2009): *Games, Groups, and the Global Good*. Springer Physica-Verlag, Berlin, Germany
- Lindfors, Patrik (2011): *Samarbete*. Fri tanke förlag, Stockholm, Sweden
- Lindfors, Patrik, Fredrik Jansson and Mikael Sandberg (2011): The Cultural Evolution of Democracy: Saltational Changes in a Political Regime Landscape. *PLoS ONE* 6(11): e28270
- Lord, Charles G., Lee Ross and Mark R. Lepper (1979): Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37(11): 2098–2109
- Masuda, Naoki and Hisashi Ohtsuki (2007): Tag-based indirect reciprocity by incomplete social information. *Proceedings of the Royal Society B* 274(1610): 689–695

- Maynard Smith, John (1982): *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, United Kingdom
- McElreath, Richard, Robert Boyd and Peter J. Richerson (2003): Shared norms can lead to the evolution of ethnic markers. *Current Anthropology* 44: 122–130
- McGrath, Joseph E. (1984): *Groups: Interaction and Performance*. Prentice Hall, Englewood Cliffs, New Jersey
- Myers, David G. (2008): *Social Psychology. Ninth Edition*. McGraw-Hill, New York, New York
- Otten, Sabine and Gordon B. Moskowitz (2000): Evidence for Implicit Evaluative In-Group Bias: Affect-Biased Spontaneous Trait Inference in a Minimal Group Paradigm. *Journal of Experimental Social Psychology* 36(1): 77–89
- Riolo, Rick L., Michael D. Cohen and Robert Axelrod (2001): Evolution of cooperation without reciprocity. *Nature* 414: 441–443
- Roberts, Gilbert and Thomas N. Sherratt (2002): Does similarity breed cooperation? *Nature* 418: 499–500
- Roth, Alvin E. and Marilda A. Oliveira Sotomayor (1992): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge University Press, Cambridge, United Kingdom
- Shapley, Lloyd S. and Martin Shubik (1972): The assignment game I: The core. *International Journal of Game Theory* 1: 111–130
- Shaw, Marvin E. (1981): *Group dynamics: The psychology of small group behavior*. McGraw-Hill, New York, New York
- Simon, Herbert A. (1972): Theories of Bounded Rationality. In C. B. McGuire and Roy Radner (eds.): *Decision and Organization* (pp. 161–176). North-Holland Publishing Company, Amsterdam, the Netherlands
- Skyrms, Brian (2004): *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press, Cambridge, United Kingdom
- Soltis, Joseph, Robert Boyd and Peter J. Richerson (1995): Can group-functional behaviours evolve by cultural group selection? An empirical test. *Current Anthropology* 36(3): 473–494
- Steels, Luc (1996): A self-organizing spatial vocabulary. *Artificial Life* 2: 319–332
- (1997): Language Learning and Language Contact. In Walter Daelemans, Antal van den Bosch and Ton Weijters (eds.): *Proceedings of the workshop on Empirical Approaches to Language* (pp. 11–24). Prague, Czech Republic

- Strimling, Pontus, Jonas Sjöstrand, Magnus Enquist and Kimmo Eriksson (2009): Accumulation of independent cultural traits. *Theoretical Population Biology* 76(2): 77–83
- Sumner, William G. (1906): *Folkways. A Study of the Sociological Importance of Usages, Manners, Customs, Mores, and Morals*. Ginn and Company, Boston, Massachusetts
- Tajfel, Henri (1970): Experiments in intergroup discrimination. *Scientific American* 223: 96–102
- Tajfel, Henri, Michael Billig, Robert Bundy and Claude Flament (1971): Social categorization and intergroup behaviour. *European Journal of Social Psychology* 1: 149–178
- Tarnita, Corina E., Tibor Antal, Hisashi Ohtsuki and Martin A. Nowak (2009): Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences* 106(21): 8601–8604
- Traulsen, Arne and Martin A. Nowak (2007): Chromodynamics of Cooperation in Finite Populations. *PLoS ONE* 2(3): e270
- Traulsen, Arne and Heinz Georg Schuster (2003): Minimal model for tag-based cooperation. *Physical Review E* 68: 046129
- Turner, John C. (1987): *Rediscovering the social group: A self-categorizing theory*. Basil Blackwell, New York, New York
- von Neumann, John and Oskar Morgenstern (1944): *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey
- Wason, Peter C. (1960): Group Selection and Assortative Interactions. *The American Naturalist* 12(3): 129–140
- Weibull, Jörgen W. (1995): *Evolutionary Game Theory*. The MIT Press, Cambridge, Massachusetts
- Wilson, David Sloan and Lee A. Dugatkin (1997): Group Selection and Assortative Interactions. *The American Naturalist* 149: 336–351
- Yamagishi, Toshio and Nobuhiro Mifune (2009): Social exchange and solidarity: in-group love or out-group hate? *Evolution and Human Behavior* 30: 229–237
- Young, H. Peyton (1998): *Individual Strategy and Social Structure*. Princeton University Press, Princeton, New Jersey