The End of Prejudice: An Experimental Study of Intergroup Conflict and Cooperation*

Andrej Tusicisny
Columbia University

Abstract

This paper develops and tests a new model explaining under what conditions people from different ethnic groups cooperate and under what conditions they discriminate against an outgroup. It also uncovers what may be the true causal mechanism underlying the famous contact hypothesis. 402 subjects sampled from the slums of Mumbai, India, participated in a randomized experiment that tested the theory. The experiment showed that (1) people cooperate if they believe that their partner will reciprocate their cooperative behavior; (2) people use their partner's ethnicity as an information shortcut to predict how likely reciprocity is; and, most importantly, (3) observation of individuals' real behavior can change the stereotypical beliefs about groups. Once expectations of reciprocity were successfully manipulated, ethnically heterogeneous groups produced as much – or as little – public goods as the homogenous ones. The experiment demonstrated that it is in fact fairly easy to rationally update deep-rooted stereotypes of outgroups even by a short social interaction. Information updating led not only to more cooperation in public goods games, but also to a radical change in self-reported discriminatory attitudes towards the outgroup as a whole. For example, the number of Hindus who would never accept a Muslim as a neighbor dropped by 56%. Practical implications of the study can guide us in designing better institutions to prevent conflict and increase public goods provision in multiethnic societies.

* This project was supported by a Columbia University GSAS International Traveling Fellowship. It greatly benefited from critical comments and suggestions made by Jack Snyder, Donald Green, Macartan Humphreys, and Neelanjan Sircar. I would like also to thank R.N. Sharma for his guidance during my field research in India and to my two tireless research assistants, Aditya Padwal and Priyesh Gohil, for being willing to work with me 12 hours a day. Other people that I am indebted to are (in an alphabetic order): Rikhil Bhavnani, Banasmita Bora, Rachel Brule, Jennifer Bussell, Basudeb Chaudhuri, Guy Grossman, Francesca R. Jensenius, Paul Kenny, Satendra Kumar, Alexander Lee, Sudha Mohan, Poulami Roychowdhury, Mark Schneider, Abdul Shaban, Vasundhara Sirnate, Divya Vaid, Gilles Verniers, Marie-Hélene Zerah. Finally, I also owe my thanks to the Columbia Global Center South Asia that hosted me during my research stay in Mumbai.

Introduction

In recent years, numerous empirical studies in political science, social psychology, and economics have reached the frontier of our understanding of identity politics. Regardless of their method, researchers all around the world encounter again and again three conspicuous puzzles that cannot be easily explained in the framework of existing theories of intergroup cooperation and conflict: First, ethnic diversity sometimes does and sometimes does not decrease the amount of public goods produced by a society. Second, the same individual usually discriminates against some outgroups, but not against others. Third, people do not always favor members of their group.

Much empirical research indicates that ethnic diversity leads to suboptimal public goods provision, greatly hindering the humanity's fight against poverty. Long ago, development economists pinpointed ethnic fragmentation as a cause of low schooling and inadequate infrastructure in Africa, the most underdeveloped region of the world (Easterly and Levine 1997). For example, Miguel and Gugerty (2005) found that Kenyan communities with an average ethnic diversity generated 20 percent less contributions for their schools than ethnically homogenous communities. Investment in public goods, from education to roads to trash pickup, is inversely related to racial diversity also in U. S. cities (Alesina, Baqir et al. 1999). For a larger sample of countries, Alesina, Glaeser et al. (2001) found a negative correlation between ethnic fractionalization and the government's social spending.

But this malign effect of ethnic diversity is not found in all societies. Edward Miguel compared two nearby and very similar districts, separated only by the Kenyan-Tanzanian border (Miguel 2004). Kenyan communities at an average level of ethnic diversity rose 25 percent less funding for their schools than homogenous communities. Across the border, in Tanzania, heterogeneous communities were equally successful as their homogenous counterparts. If there is any general conclusion, it is that "similar levels of ethnic diversity are associated with very different degrees of conflict and interethnic cooperation" (Alesina and Ferrara 2005: 794).

Another puzzle is why the same person discriminates against some, but not all outgroups. Blalock (1967) predicted that a greater relative size of the minority group to the dominant group would threaten the latter because of increased competition for scarce resources and and the minority's higher potential for political mobilization. In his book on electoral politics in Zambia, Posner (2005) argued that ethno-political cleavages appear between groups, whose size allows them to be part of a minimum winning coalition. However, the relative group size argument cannot explain why discrimination often

targets small or disenfranchised outgroups. If we look at some of the groups discriminated against by the majority according to the most recent (2003) results of the Minorities at Risk project, Haitians represent only three percent of the population of the Dominican Republic, Whites less than one percent of the population of Zimbabwe, and Bahá'ís less than 0.5 percent of the population of Iran. Conversely, governments in Indian states tend to protect the Muslim minority against ethnic riots if it is large enough to give them votes (Wilkinson 2004).

People sometimes do not discriminate against outgroups; they may even favor them. This empirical observation contradicts one of the basic assumptions of the celebrated social identity theory, which predicts that an individual should favor his/her fellow ingroup members over outgroup members. Such positive evaluation of the ingroup in turn maintains one's identification with the group. This is not always the case. According to a Eurobarometer survey, Italians trusted the Swiss, Swedes, and Americans more than their own countrymen. Swedish respondents had in turn more trust in Norwegians than in other Swedes. Social psychologists are aware of this shortcoming of the social identity theory. A widely cited meta-analysis from 1990 concluded that correlation between strength of group identification and ingroup bias was actually close to zero (Hinkle and Brown 1990).

This paper proposes a parsimonious theory that explains why people from different groups sometimes fight and sometimes cooperate. I argue that we can predict cooperation and conflict between groups by looking at the individual's expectations of whether the group's members will reciprocate cooperative behavior or not. Expectations of reciprocity are subject to bounded rationality and can be updated in a way encouraging more cooperation between groups. Although the proposed model is general enough to be applicable to any groups, I focus here on cooperation between ethnic groups.

I tested the theory in the ethnically heterogeneous slums of Mumbai, India. In order to identify and measure the true causal effect on cooperation and discrimination, I conducted a series of laboratory experiments. But unlike most experimental studies, with subjects drawn from Ivy League undergraduate students taking courses in psychology or economics, this research project involved random sampling of 402 subjects from a politically relevant population.

Reciprocity

The idea that reciprocity sustains cooperation is not new. As early as in 1906, social scientists hailed reciprocity as "the vital principle of the society" (Hobhouse 1906: 12). Importance of reciprocity

naturally arises from many evolutionary models of the human society. Early research on the iterated prisoner's dilemma game showed that the most successful strategies are those of conditional cooperators (Axelrod 1984). Evolutionary models also show that a population of indiscriminate altruists will be invaded by free riders. Only if a sufficient number of people refuse to cooperate with free riders, cooperation can be sustained (Nowak and Sigmund 2005). Such discrimination between likely cooperators and defectors is regarded as the society's "immune response" (Nowak and Sigmund 2000). But to make this immune response work, we need to know the reputation of our partner first.

If reciprocity is basis of cooperation, a logical question is how we estimate whether our partner will reciprocate or defect. In repeated direct exchange with the same partner, we can rely on our recollection of his/her previous actions. In direct exchange with new partners, we use their reputation for the same purpose. But if the reputation of another individual is not readily available, what information can we use to estimate the likelihood of cooperation? And what about situations, in which we cooperate with a large number of partners simultaneously, for example while paying our taxes?

I argue that humans use group membership as an information shortcut. This idea is in itself quite uncontroversial and underlies many other theories of intergroup relations: it is an accepted wisdom that people tend to reduce the complexity of the social world by transforming innumerable fuzzy differences between individuals into few clear dichotomies (Allport 1954; Tajfel 1969). Henri Tajfel provided a good example more than fifty years ago: although skin color is a continuous variable, many societies have classified people into a "white" and a "black" category (Tajfel 1969). As many other behavioral characteristics have become generalized into stereotypes about groups (Allport 1954), there is no reason why expectations of reciprocity should be immune to this natural tendency.

Stereotypes about outgroups may lead to negative attitudes that we call prejudice. Building on Allport (1954), I define prejudice as antipathy (1) based on group-level generalization and (2) directed either toward a group as a whole or toward an individual because of his/her group membership. Prejudices may manifest themselves in people's behavior in the form of discrimination. If I believe that all immigrants are crooks, I will be less likely to offer them a job, lend them money, or enter into any transactions based on trust with them. As prejudices prevent an individual from obtaining resources by cooperation, competition between groups will become paramount.

We know that prejudices sometimes change in a dramatic way. For example, Schuman, Steeh et al. (1985) described how the American public opinion moved away from the view that whites should have a priority over blacks in the public life between the 1940s and 1970s. However, the accepted

wisdom is that the most salient predispositions – including partisanship, ideology, and ethnic prejudices – are developed early in the life and persist well into the adult life (Sears and Funk 1999).

The most prominent theory on how to change group stereotypes is Gordon Allport's contact hypothesis: Under specific conditions, positive intergroup contact should reduce prejudices (Allport 1954). When Varshney (2002) in his book on ethnic riots in India contrasted peaceful Calicut to conflict-prone Aligarh, he found out that 90 percent of Hindu and Muslim families in Calicut, but only 42 percent in Aligarh, reported that their children played together. Although this illustrative observation says nothing about causation, other, more direct and rigorous tests, do. As Pettigrew and Tropp (2006: 751) unequivocally concluded in their authoritative study, "with 713 independent samples from 515 studies, the meta-analysis finds that intergroup contact typically reduces intergroup prejudice".

Not all contact between groups is beneficial, though. Pettigrew (1971: 275) summarized the optimal conditions of a positive contact as such: "Prejudice is lessened when the two groups (1) possess equal status, (2) seek common goals, (3) are cooperatively dependent upon each other, and (4) interact with the positive support of authorities, laws, or customs." In their meta-analysis, Pettigrew and Tropp (2006) concluded that these conditions were not necessary, but they significantly increased effectiveness of contact. However, the individual causal effects remained elusive.

Allport himself only hinted on the possible causal mechanism behind his brainchild: "Contacts that bring knowledge and acquaintance are likely to engender sounder beliefs concerning minority groups, and for this reason contribute to the reduction of prejudice" (Allport 1954: 268). I argue that not *any* information updating matters. What matters is, specifically, updating of expectations of reciprocity. Molm, Peterson et al. (2003) found that negotiated exchange (such as trade and employment) actually makes conflict between actors more salient and triggers unfavorable perceptions of the other side as unfair. On the other hand, "reciprocal exchange" (in fact indirect exchange) increases perceptions of fairness. According to the affect theory of social exchange, a person-to-group tie is more salient in indirect exchange and positive emotions created by a successful exchange are attributed to the group as a whole (Lawler 2001). What we maybe really need are positive expectations of reciprocity between groups, not necessarily equal status or other conditions mentioned in the contact hypothesis literature.

My causal argument can be summarized by the following three propositions:

1. People cooperate if they believe that their partner will reciprocate their cooperative behavior.

2. People use group membership as an information shortcut to predict reciprocity from others.
3. People rationally update their beliefs about groups' cooperative behavior by observing real interactions with the groups' members.

The idea that people care about expectations of reciprocity and not necessarily about the ingroup/outgroup dichotomy leads to several testable implications:

Table 1: Predicted Individual Behavior Depending on Expected Reciprocity

		Outgroup	
		Cooperates	Defects
Ingroup	Cooperates	No bias, cooperate	Ingroup bias
	Defects	Outgroup bias	No bias, defect

The first cell of Table 1 corresponds to the case of generalized reciprocity. People have expectations that others will cooperate regardless of group membership. I expect no bias towards one or the other group:

Hypothesis 1: If expectations of reciprocity in social interactions are generalized (i.e. both ingroup and outgroup members are expected to reciprocate cooperative behavior), an individual will be equally likely to cooperate with ingroup and outgroup members. No ingroup bias will be observed.

Unlike the social identity theory, my model predicts that we should observe ingroup bias only in one special case: if ingroup members are expected to reciprocate cooperation, while outgroup members are not. This case is captured by the second cell of the two-by-two matrix above.

Hypothesis 2: If ingroup members are expected to cooperate and outgroup members to defect, an individual will tend to cooperate with ingroup members and defect in relations with outgroup members (ingroup bias).

The third cell of Table 2 predicts the ever-puzzling outgroup bias. Here it naturally arises from calculations of a rational individual.

Hypothesis 3: If outgroup members are expected to cooperate and ingroup members to defect, an individual will tend to cooperate with outgroup members and defect in relations with ingroup members (outgroup bias).

The fourth cell of our two-by-two matrix captures a situation of general mistrust and is thus the polar opposite of generalized reciprocity:

Hypothesis 4: If both groups are expected to defect and no punishment of defectors is possible (anonymous setting), an individual will be equally likely to defect in relations with ingroup and outgroup members (no ingroup bias).

Research Methodology

Following Harry Eckstein's case-selection criteria, I chose the *least-likely crucial case* to test my argument (Eckstein 1975). A least-likely case is one that "on all dimensions except the dimension of theoretical interest, is predicted not to achieve a certain outcome, and yet does so" (Gerring 2007: 115). In other words, the least-likely case offers the toughest test for my argument. I found such least-likely crucial case in the slums of Mumbai, India.

Communal violence between Hindus and Muslims is a regular feature of Indian politics. But even in the Indian context, Mumbai gained notoriety as the country's historically most riot-prone city, with 1,137 recorded deaths in 1950-95 (Varshney 2002). The single worst episode of the perennial ethnic conflict in Mumbai followed the destruction of the Babri Mosque in Ayodhya by a mob of Hindu extremists in December 1992. The demolition sparked an increasingly violent series of Muslim protests that escalated into full-blown street fighting between Muslim and Hindu mobs. The police intervened, killing disproportionately more Muslims than Hindus (Srikrishna 1998). The second, more serious round of rioting began in January 1993 and was to a great degree orchestrated by the local fascist party Shiv Sena and occurred with complicity of the police (Engineer 1993; Srikrishna 1998; Hansen 2001). Two moths of rioting left 900 dead, among whom 575 were identified as Muslims and 275 Hindus (Srikrishna 1998). Rioters razed or burnt down nearly 10,000 houses and the estimated 200,000 people fled the city (Engineer 1993). Large numbers of Muslims previously living among the Hindu majority left their little enclaves and moved to few overcrowded, but more defensible ghettos (Hansen 2001; Shaban 2010). My own field work in Mumbai in 2011 showed that the riots of 1993 are still the single most important item in the city's collective memory. Every conversation about intergroup relations in Mumbai inevitably drifted to the topic of riots.

Although ethnic riots have become more of an abstract threat plaguing the city's subconsciousness, frequent terrorist attacks attributed to Indian Muslim terrorist organizations with ties to Pakistan regularly rekindle the simmering ethnic conflict. The first large-scale terrorist attack came as a direct response of the Muslim underworld to the riots. Ten powerful bombs exploded at Mumbai's

landmarks, including Bombay Stock Exchange, and killed 251 people. Since then, acts of terrorism periodically stir the ethnic conflict between Hindus and Muslims living in Mumbai. A terrorist attack that truly mesmerized media worldwide came on November 26, 2008. A band of ten Pakistani terrorists landed with assistance of an unknown number of locals and attacked multiple locations, including Mumbai's main train station, the famous Taj Mahal Palace Hotel, and a center run by an international Jewish movement. The drama full of indiscriminate shooting, explosions, and hostage murdering went on for 60 hours followed by millions of TV viewers. The most recent terrorist attack in Mumbai incidentally happened just after I finished my field research there. In the evening of July 13, 2011, bombs exploded at several crowded locations across the city (Bajaj and Goodman 2011). They killed 26 and injured 141 people.

Mumbai would be an interesting case even without its ongoing ethnic conflict. RAND Corporation appropriately called Mumbai "India's Wall Street, its Hollywood, its Milan" (Rabasa, Blackwill et al. 2009: 1). Already being the biggest city in India, Mumbai is poised to become the world's third biggest urban agglomeration after Tokyo and Mexico City (Mumbai 2010). According to the last census with complete data (2001), 44 percent of people in Mumbai were immigrants (Mumbai 2010). Whereas Gujaratis (10% of immigrants) are considered to be part of the city's traditional commercial bourgeoisie (Hansen 2001), more recent immigrants from North India are often frowned upon because of their high number (24% from Uttar Pradesh) or their dramatically increasing immigration rate (from Bihar).

Mumbai's population density is the highest among all cities on Earth. There are only 20 square meters per person in the city of Mumbai – and this number includes not only apartments, but also all the roads, parks, office buildings, factories, and landfills. In comparison, there are 94 square meters per person in New York City. Consequently, housing and employment are a scarce resource and a frequent point of contention between ethnically-based political parties. The labor market in Bombay has traditionally been organized along caste lines, with castes (*jatis*) controlling particular occupations for generations (Munshi and Rosenzweig 2006). The two dominant nationalist political parties in Mumbai, Shiv Sena and MNS, have marketed themselves from the beginning as defenders of the Marathi people against South Indians (Shiv Sena) and North Indians (MNS) in competition for jobs. At the same time, the inevitable spatial proximity of different groups creates a quirky mix of competition and cooperation. This is true especially in slums, in which lower-caste Hindus (from the Scheduled Castes) mostly live next to and around Muslim ghettos (Shaban 2010).

Among all ethnic groups living in Mumbai, Marathas have the most salient ingroup identity. They are the only ethnic group that strictly insists on the use of their language (Marathi) instead of the usual *Bambaiya Hindi* pidgin of Hindi, English, and Marathi. They are the largest supporter of the nationalist political parties of Shiv Sena and MNS (Palshikar and Deshpande 1999; Palshikar 2004; Shaban 2010). The Shiv Sena (SHS) movement was founded in Bombay in 1966 as a vehicle for the interests of middle-class Marathas in the city against skilled South Indians (Hansen 2001). Just like fascist parties in the interwar Europe, this anti-establishment social movement employed marginalized from lower classes to brutally terrorize its opponents. Over the years, Shiv Sena's targets changed from South Indians to Communists to the city elite to Indian Muslims (Hansen 2001). Having such a distinct identity, defined in contrast to other ethnic groups, Marathas are the least-likely to support my argument that it is possible to reduce prejudices by updating of expectations of reciprocity. That is why I chose them as my population of interest.

I define a Maratha as a Marathi-speaking Hindu, born in Maharashtra. Strictly speaking, Marathas are not a single caste (*jati*). The ancient society of the Deccan Plateau was organized around clans that shared the same Marathi linguistic identity and a relatively homogenous culture (Hansen 2001). Unlike in the case of more conventional castes, the boundaries of the Marathi identity were traditionally more fluid. However, this common identity was later stabilized "by references to *savarna jati*, the caste communities of pure blood, or the *hindu samaj*, the larger Hindu community as opposed to Muslims" (Hansen 2001: 24). Marathas became the politically dominant caste after the creation of the separate Marathi-speaking state of Maharashtra with Bombay as its capital in the 1960s (Munshi and Rosenzweig 2006).

In a final blow to my theory, I decided to sample only Marathi men. In Mumbai – and arguably also in general – men are much more riot-prone and violent. The core support base of the nationalistic alliance of BJP and Shiv Sena in Maharashtra comes from the urban male youth (Palshikar and Deshpande 1999). Men in general also tend to favor ingroup members to outgroup members in laboratory experiments more than women (Yamagishi and Mifune 2009). If I can show that we can get rid of ingroup bias in men, this will be even more plausible for women.

Although there are no reliable demographic data collected at the level of neighborhoods, local experts helped me identify three predominantly Marathi slum neighborhoods: Bhoiwada, Magathane, and Shivaji Nagar. As is often the case in Mumbai – and in other ethnically heterogeneous parts of the world – these more or less homogenous areas were in close proximity to zones inhabited by other

ethnic groups. A good example is Shivaji Nagar with its alternation of exclusively "Hindu" and "Muslim" streets. Although all the buildings in Shivaji Nagar are the same corrugated iron huts, the ubiquitous saffron *or* green flags easily point a lost visitor to the right ethnic bastion.

As an official report concluded, "slums have a tendency of magnifying small issues and minor irritants into full–blown communal riots" (Srikrishna 1998). Most underworld dons are from slums and local violent nonstate actors, whether it be gangs in Muslim slums or extremist political parties in Hindu slums, enjoy impunity in their respective areas (Shaban 2010). At the same time, common civic institutions are securing communal peace in and between these neighborhoods. After the deadly riots in 1993, each police station has set up a *mohalla committee* consisting of local community leaders, who diffuse tensions at the time of crisis. Security is not the only public good that requires bottom-up contributions channeled through the civil society. The main vehicle for public goods provision in the slums of Mumbai are the so-called *community-based organizations* (CBO). Slums located on hillsides get their water not directly from the authorities, but through their own cooperative society that maintains pumps, collects money from members, and pays bills to the authorities (Mumbai 2010). World Bank-sponsored Slum Sanitation Programme involves local CBOs in planning, construction, and maintenance of "pay-and-use" toilets (Mumbai 2010). Due to the program's unprecedented success, almost all newly constructed public toilets belong to this category (Bhide 2011). CBOs also help with garbage disposal. According to Bhide (2011), this Slum Adoption Scheme now covers 45% of slum population.

Most public goods experiments to this day have been run in university laboratories with student convenience samples. Because of significant differences between American college students and the rest of the world, it is usually problematic to generalize such results to any meaningful population (Sears 1986; Henry 2008; Belot, Duch et al. 2010). In order to avoid this potential threat to external validity of my findings, I chose a less common yet more promising approach – random sampling of subjects representative of a population that is the most interesting in the context of this work. Following some of the leading experimentalists (Henrich, Boyd et al. 2001; Habyarimana, Humphreys et al. 2009; Henrich, Ensminger et al. 2010), I literally brought the laboratory to where subjects live. The experiments were conducted in rented Internet cafes in the areas, where the sampled individuals lived. This approach is often called a "lab in the field" experiment (Morton and Williams 2010).

402 subjects were randomly sampled using a "random route" technique. 210 individuals were interviewed, but they did not receive any experimental treatment. Following the usual practice in the

field, I refer to them as to the “control group” because their attitudes will serve as a baseline for some of the comparisons discussed in this study. Four treatment groups (T1-T4) corresponded to my four hypotheses (discussed in more detail below). Within each neighborhood, treatment assignment was randomized. Subjects in each area were first randomly assigned to either the control group (and filled the questionnaire) or to receive the treatment (and arranged an appointment for the experimental session, usually for the next day). At the beginning of each experimental session, a random number generator assigned the subject to one of the four treatment groups. Table 2 demonstrates that the groups were well balanced and that treatment assignment did not correlate with location.

Table 2: Sampled individuals per treatment group and neighborhood

Treatment Group	Bhoiwada	Magathane	Shivaji Nagar	Total
T1	14	17	17	48
T2	15	16	16	47
T3	16	17	16	49
T4	19	14	15	48
Control group	70	70	70	210
Total	134	134	134	402

The main goals of this experiment was to find out whether different expectations of reciprocity have an effect on cooperative behavior, whether people use group membership to estimate how likely reciprocity is, and whether a real record of behavior can change the stereotypical beliefs about groups' reciprocity. For this purpose, I employed the public goods game, which is frequently used by economists (see [Camerer 2003] for a review) and increasingly also by political scientists (for example [Habyarimana, Humphreys et al. 2009]) to investigate under what conditions people cooperate.

The payoff structure of the game that participants were playing during my experiments in India (Table 3) corresponds to the well-known s*tag hunt game*. The game got its name from a famous story in Jean-Jacques Rousseau's *Discourse on the Origin and Basis of Inequality Among Men*. The story, never actually spelled out by its author, can be envisaged as follows: Two or more hunters are hunting a stag. While they are waiting for the beast to appear, one of the hunters sees a hare passing by. The hunter faces a dilemma: If he shoots the hare, he will get its meat for himself. Unfortunately, the shot will also scare the stag off. The stag could provide him and his companions with much more meat than a single hare. However, if the hunter decides to wait for the stag, he risks that someone else will see the hare

and fire at it.

Table 3: Payoff Structure of the Stag Hunt Game

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	20 20	10 0
	Defect	0 10	10 10

In its general form, the game can be described by the following ordering of outcomes: CC > DC ≥ DD > CD. The game has two pure-strategy Nash equilibria: Either both players cooperate or they both defect. How the game will end thus depends on the level of trust in the group. A rational player will hunt the stag if he/she expects the other player to do the same; but he/she will go for the hare if he/she doesn't trust his/her companion. Both players are better off if they hunt a stag; this is a Pareto-optimal equilibrium. But to hunt hare is a risk-dominant strategy. There is also a mixed strategy equilibrium that depends on the payoff matrix. In the game used for my experiment, the mixed equilibrium would be to cooperate with the probability of 0.5.

The stag hunt game was hailed as the "prototype of the social contract" (Skyrms 2004: 1) and has found many applications in social sciences.[1] As early as in the 18th century, David Hume used the stag hunt game to model public goods provision. In his example, two neighbors wish to drain a meadow that they own in common. If they both work to drain it they will be successful, but if either fails in his part, the meadow will not be drained (Hume 2003). Incidentally, Hume's metaphor is the everyday reality for Mumbaikars. Every monsoon season, the 100-years old storm water drain system has to channel the flood water out of the city. When it becomes clogged, monsoon rains inevitably cause a disaster. For example, the heavy downpour of July 26, 2005 submerged several neighborhoods for long seven days (Arunachalam 2005). All rail and road traffic in the city came to a standstill and even the runways at the international airport were flooded. It took seven days to restore power and ten days to fully restore train traffic in India's financial capital. Mumbai's storm water drain system becomes clogged periodically because people throw garbage into open drains (Mumbai 2010) and slums are encroaching on the shrinking rivers (Arunachalam 2005). If everyone stops dumping garbage

1 For example, the stag hunt game has been used to model arms race and disarmament (Jervis 1978).

in drainage channels and stops building huts inside the natural waterways, everyone will greatly benefit from a functioning drainage system when the next flood comes. The problem is that people do not trust their neighbors enough to believe that the channels will stay clean. And if you believe that someone else will inevitably clog the drainage channel, why not to dump your garbage there as well?

Most laboratory public goods games to this day have used the prisoner's dilemma payoff matrix – making a strong implicit assumption that the real-life breakup of cooperation is motivated by people's preference for exploitation. I argue that defection in social exchange is not caused as much by the inherent desire to exploit others as by the lack of trust. Although a rational egoist should always defect in a prisoner's dilemma game, real people tend to cooperate if they believe that others are likely to cooperate (Ostrom 2000). For example, even in a one-shot prisoner's dilemma game, 40% of 136 subjects actually preferred (CC) to (DC) and 27% were indifferent between these two alternatives (Ahn, Ostrom et al. 2001). In other words, many people react as if they were playing a stag hunt game even in the situations more similar to the prisoner's dilemma.

During my experiment, each subject was seated in front of a computer and told that he would play a simple computer game with various people in two other areas of Mumbai: Bhendi Bazar (a typical Muslim enclave) and Dadar (a well-known Hindu Marathi neighborhood). Although religion or ethnic affiliation was never explicitly mentioned during the experiment, the computer screen showed a picture, the first name, and the neighborhood of the other player – three different unobtrusive clues of the partner's membership in the Hindu ingroup or the Muslim outgroup. During the experiment, players participated in 10 rounds of a public goods game in total with 5 distinct ingroup and 5 distinct outgroup members. Participants received financial rewards contingent on their performance in the experiment and the average compensation for a half an hour-long experiment represented 20% of the average daily household income.

In each round, the player received 10 Rupees and could decide whether to keep them or to invest them in a common project with the partner on the screen. The partner made the same decision simultaneously, without any communication with the player. The payoff structure corresponded to the stag hunt game (Table 3): If both the player and his partner invested in the common project, each of them received 20 Rupees (CC). If the player invested, but the partner kept his money, the project failed and the player lost his money. The payoff was 0 (CD). Or the player could simply keep the 10 Rupees. If he did it and the partner also kept his money, each got the payoff of 10 Rupees (DD). If the partner invested, but the player himself kept the money, the player's payoff was still 10 Rupees, while the

partner's payoff was 0 (DC). The best strategy clearly depended on whether the player expected his partner to cooperate or not. After each round, a screen showed the results of the round. But as each round was played against someone different, there was no way to know the individual's previous record of cooperation.

I used the subject's contribution to the common project in the last round played with an ingroup member and the contribution from the last round played with an outgroup member as behavioral measures of cooperation after the treatment. Most laboratory experiments only measure how the experimental treatment changes subjects' behavior in an artificial laboratory setting. In order to increase internal validity of my results, I also asked the subjects about their real-life hostile behavioral intentions in a post-treatment survey. I used two questions taken from different points of Bogardus's social distance scale (Bogardus 1925), which is frequently used by psychologists. The first question asked "If it was up to you, would you accept a Muslim to close kinship by marriage?" This question represents the most extreme pole of Bogardsus's scale, as even the most tolerant respondents in liberal democracies often discriminate against outgroups as long as family matters are concerned. The second question asked about a less prevalent type of avoidance: "If it was up to you, would you accept a Muslim as a neighbor?" We can think about these two questions as indicators of discriminatory attitudes, but also as an inverse measure of acceptance or tolerance.

Four treatment groups correspond to the cells of Table 1. The subjects in the T1 group always faced cooperators (both ingroup and outgroup members). Subjects in the T2 group faced cooperating ingroup members and defecting outgroup members. T3 played with defecting ingroup members and cooperating outgroup members. For the subjects in T4, both ingroup and outgroup members tended to defect. The manipulated frequency of defection in previous rounds of the game thus created a low or high expectation of reciprocity from other members of the ingroup and members of the outgroup. These four ideal types cover the whole spectrum of possible interactions.

31 percent of the laboratory experiments that have appeared in *American Political Science Review* used deception (Druckman, Green et al. 2006) and this tool is especially prevalent among projects researching prejudice (McDermott 2002). My experimental design is no exception. The element of deception was that participants believed that they were playing games with real people, while their partners were in fact generated by a computer program. Unfortunately, we would not be able to observe people's behavior across all four experimental conditions without deception.

Financial incentives tied to the well-known and simple payoff structure allowed me to hold material costs and benefits of the interaction constant. I also tried to hold the likely effect of social sanctioning constant at 0. The subjects interacted only through computers. They were also told that their partner knew only their neighborhood (a proxy for ethnicity), but could not see their name or their face. Therefore they had no reason to fear retaliation after the experiment ended. Furthermore, players were not allowed to talk to each other or to look at each other's screen – assuring that there was no peer pressure inside the laboratory.

Cooperation may have an intrinsic psychological value independent from material benefits and distinguishable from social pressure (Dovidio 1984 ; Piliavin and Charng 1990). This value consists of two different elements: General altruism is equally high regardless of the recipient's identity, while ethnocentric altruism gives more value to cooperation with ingroup members than with outgroup members. Although I could not manipulate the subjects' level of altruism, I could at least measure it. Before the public goods game described in a previous section, all subjects played one round of a dictator game. The dictator game is a one-stage game in which a subject decides how to allocate a sum of money between him/herself and other players (hence the “dictator” name). A purely self-interested person should keep all the money. However, empirical research found that people usually transfer considerable amounts of money to other players (Camerer 2003). Participants in my study played a version, in which the allocator is anonymous to other players, but the receivers' identities are known by the allocator. During the dictator game, each subject was given 10 Rupees and told that he could divide them between himself and two partners in any way he wished. The two partners were represented on the computer screen by a photograph, first name, and neighborhood. One of them appeared to be a Hindu Maratha (an ingroup member) and the other Muslim (an outgroup member).

How much money the player allocated to each of them was a direct behavioral measure of the subject's baseline altruism in relations towards ingroup and outgroup members. The difference between the amount of money allocated on average to an ingroup member (Rs. 1.79) and to an outgroup member (Rs. 1.86) was insignificant both substantively and statistically (p-value=0.457). In fact, the modal strategy was to give each of the two other players one Rupee and pocket the rest. When not observed by anyone, 52% of participants rewarded themselves with 8 out of 10 Rupees that they could allocate.

There may be and probably are other, unobserved confounding factors. Fortunately, randomization ensures that these factors are orthogonal to the treatment. Since randomization of the

treatment assignment was performed within each neighborhood separately, I checked for differences in various observable covariates in each of the three blocks. For this purpose, I regressed the assignment of a subject to a particular treatment group on different covariates. This very conservative randomization check revealed several imperfections. I ran this analysis for each group (T1, T2, T3, T4, C) and each neighborhood (Bhoiwada, Magathane, Shivaji Nagar), but I do not report all fifteen multinomial logistic regressions here. Instead, the following table summarizes the statistically significant results. The "plus" or "minus" sign shows how exactly a given covariate changed the probability of assignment to the given treatment group in the given neighborhood. Fortunately, Table 4 does not show any systematic problem with the randomization process. For example, there are more Hindus in the control group in Bhoiwada, but not in Magathane or Shivaji Nagar. Whereas subjects in the control group in Bhoiwada and Magathane appear to be older (holding all other variables constant), they are younger in Shivaji Nagar. Since the number of subjects in each cell of the table is quite low (e.g. 14 subjects in T1/Bhoiwada), the differences between them are likely to be caused by chance.

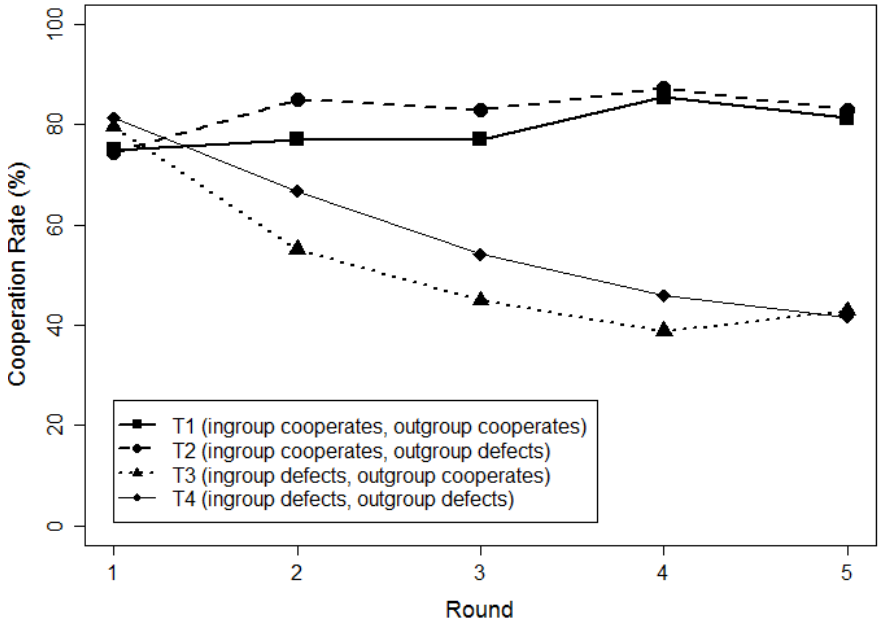
Table 4: Randomization Check

	Bhoiwada	Magathane	Shivaji Nagar
T1	Age – Education +		Income –
T2	Hindu – Income +	Age –	Age +
T3	Age –		
T4			
C	Hindu + Age + SHS&MNS voters –	Age +	Age – Education + Income +

Results

Figure 1 shows the percentage of players who cooperated in each of the five rounds played with other ingroup members. All four treatment groups started at a surprisingly high level of cooperation, with about 77% of participants investing real money in the common project with a stranger on a computer screen. In comparison, the average contribution in the first round in Western laboratories is typically between 40 and 60 percent of the endowment (Ostrom 2000). But if we break up the results by neighborhood, the poorest slum (Shivaji Nagar) had the initial cooperation rate of about 60%, Magathane of about 80%, and the wealthiest neighborhood (Bhoiwada) close to 100%.

Figure 1: Cooperation Rate in the Rounds Played with Ingroup Members

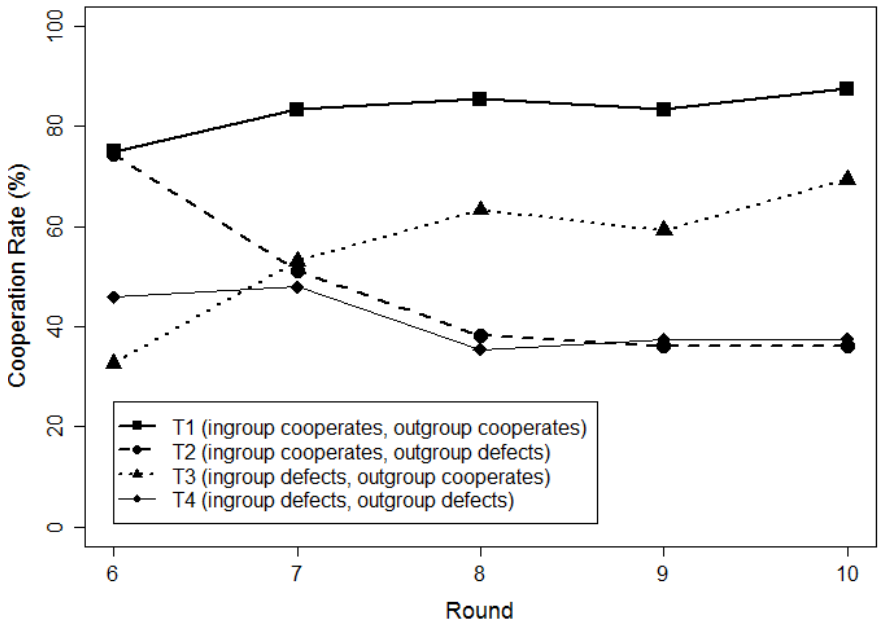


During the first five rounds, trajectories of the four treatment groups quickly diverged. The two groups that encountered defectors (T3, T4) started to defect too. By the fifth round, the number of people in these two groups willing to invest money in a common project dropped by half, from 80% to 42% (0.95 movement in standard deviation of the initial cooperation rate). On the other hand, the two groups that faced cooperative players (T1, T2) retained the initial high level of cooperation. In fact, the proportion of cooperators increased from 75% to 82% (or 0.17 movement in standard deviation of the initial cooperation rate). The difference between T3 and T4 on the one side, and T1 and T2 on the other

side was highly significant (p-values in a t-test with Holm adjustment lower than 0.001). The difference-in-means between the two groups with increased expectations of reciprocity and the two groups with lowered expectations of reciprocity is 40 percentage points. This difference corresponds to the 0.8 of standard deviation of the cooperation rate in T3 and T4.

The story of what happened after the participants were paired with outgroup members is even more interesting. As we can learn from Figure 2, 75% of people who have experienced cooperative ingroup members (T1, T2) immediately started to cooperate with their first Muslim partner too. Despite a different ethnicity, subjects in the T1 group (i.e. those with cooperative Muslim partners) kept increasing their cooperation rate, ultimately by 13 percentage points, from 75 to 88% (0.29 movement in standard deviation of the initial cooperation rate in the sixth round). However, subjects in the T2 group quickly learned that their partners did not reciprocate cooperation. The proportion of cooperators in this group dropped from 75 to 36% (0.87 of standard deviation).

Figure 2: Cooperation Rate in the Rounds Played with Outgroup Members



Subjects in the T3 and T4 groups started at a much lower level of cooperation because of all the defections they experienced in the previous five rounds. The T4 group predictably retained a low level of cooperation, with the cooperation rate in the last round lower than 38%. What is interesting, both groups facing defectors (T2 and T4) exhibited essentially the same pattern in the last four rounds

despite their very different initial conditions. The T3 group started the sixth round of the game at the lowest level of cooperation (33%). But subjects eventually learned that Muslims are more cooperative than Hindus. Cooperation rate kept climbing until it reached 69%. This increase by 47 percentage points corresponds to a 0.78 move in standard deviation of the cooperation rate in the sixth round. A pairwise comparison of means shows that the average final contributions of T1 And T3 are significantly different from those by T2 and T4 (p-values<0.002), while there is no statistically significant difference between average final contributions in T1 and T3, and in T2 and T4 respectively. The treatment effect is, again, more pronounced in the slums of Magathane and Shivaji Nagar than in Bhoiwada.

Disregarding ethnic affinity, participants cooperated with those whom they expected to reciprocate cooperation. They also correctly classified people into likely cooperators and defectors using clues of their ethnicity. Expectations of reciprocity changed by a single individual were subsequently generalized to his group and applied to the group's other members. Information updating also turned out to be surprisingly fast. Whenever, people observed that a particular group tended to defect, the number of cooperators dropped by half in just five rounds of the public goods game. When the T3 group observed that the outgroup cooperated more than the ingroup, their own average contribution increased by more than 100%. The highest level of cooperation in the final round of the game (88%) was achieved in the group in a situation of generalized reciprocity (T1). Most importantly, we did not observe any ingroup bias apart from that induced by the experimental manipulation of expectations of reciprocity (ingroup bias for T2 and the “mysterious” outgroup bias in T3).

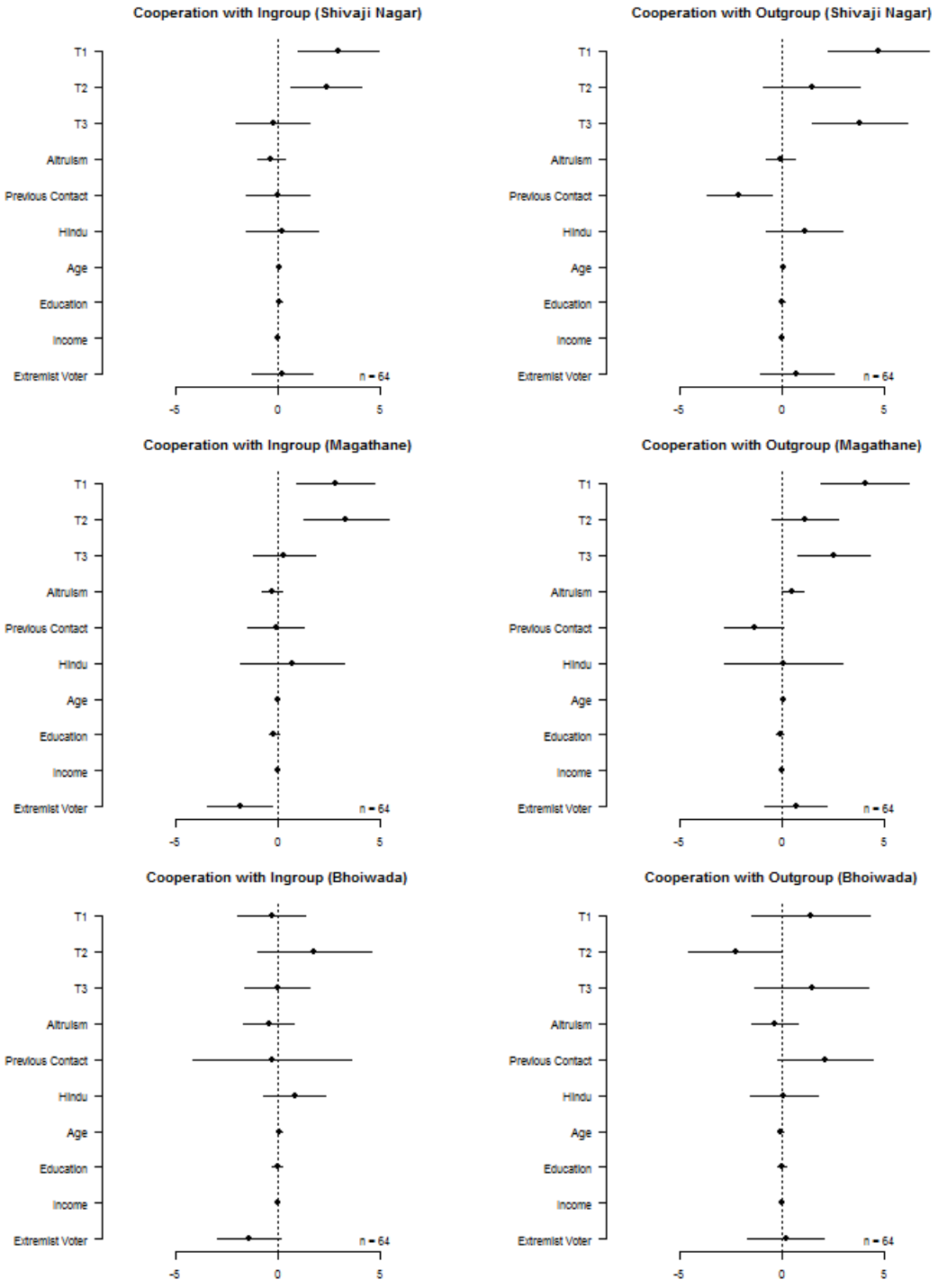
Figure 3 reports the results of multinomial logistic regressions predicting whether a person cooperated or not in the last round of the public goods experiment. I conducted this analysis in order to control for various covariates that could correlate with the outcome. The models in the left column predict cooperation in the last (fifth) round played with a partner from the ingroup. The models in the right column predict whether the subject cooperated in the last (tenth) round played with an outgroup member. Both models use cooperation in the T4 group as a baseline, so we can more easily compute how much different expectations of reciprocity increased cooperation compared to this uncooperative default situation. I ran the regression for each neighborhood separately. This approach follows old Fisher's rule “you analyze as you randomize”. All six regressions were estimated using a Bayesian alteration of the logistic regression described in Gelman, Jakulin et al. (2009). I could not use the standard maximum likelihood because of complete separation in Shivaji Nagar, where all subjects in the T4 group defected in the last round. An unmodified logistic regression would produce a coefficient

equal to negative infinity for T4, which would be, of course, flattering to my hypothesis, but also quite unhelpful. In the remaining five regression equations (i.e. those with no separation problem), coefficients remained essentially the same regardless of whether I used the maximum likelihood or the Bayesian approach.

Altruism is measured by the subject's contribution to the ingroup member (left column in Figure 3) or to the outgroup member (right column) in the dictator game. I also added a control variable (called “Previous Contact” in Figure 3) for the subject having acquaintances in Dadar (left column) or Bhendi Bazar (right column). If the contact hypothesis is true, some subjects may be more cooperative with partners from Dadar not because they share the same ingroup identity (Hindus), but because they had encountered other people from that neighborhood before the experiment. This variable was self-reported by the subjects in a post-treatment survey and recollection of this information might have been possibly changed by the treatment itself. However, the Tukey Honest Significant Difference (HSD) method did not reveal any significant difference in values between the treatment groups and the control group. Therefore, a treatment effect on this variable is highly unlikely and it should be safe to use it in the model. Finally, I added covariates, which appeared to be slightly imbalanced across treatment groups (see Table 4). These variables as well were measured by a post-treatment survey. Nevertheless, it is difficult to imagine how my treatment could change the subjects' age, religious affiliation, education, and income. The only covariate that might have been changed by the treatment is the person's recollection for which party he/she voted in the last election. But, again, the HSD did not show any statistically significant differences between the control group and the treatment groups.

The models predicting cooperation with ingroup correctly predicted 75 (Shivaji Nagar) to 86 (Bhoiwada) percent of observations. What is more informative than this high value of PCP is the percentage reduction in error (PRE) in comparison to the null model (that is the one with only an intercept). Using this metric, the full regression equation improved the explanatory power of the null model by 65% in Magathane, 44% in Shivaji Nagar, and 20% in Bhoiwada. This indicates substantive predictive power of the included variables. The Nagelkerke / Cragg and Uhler's pseudo R2 was also high (from 0.44 in Bhoiwada to 0.60 in Magathane).

Figure 3: Estimated Coefficients with 95% Confidence Intervals



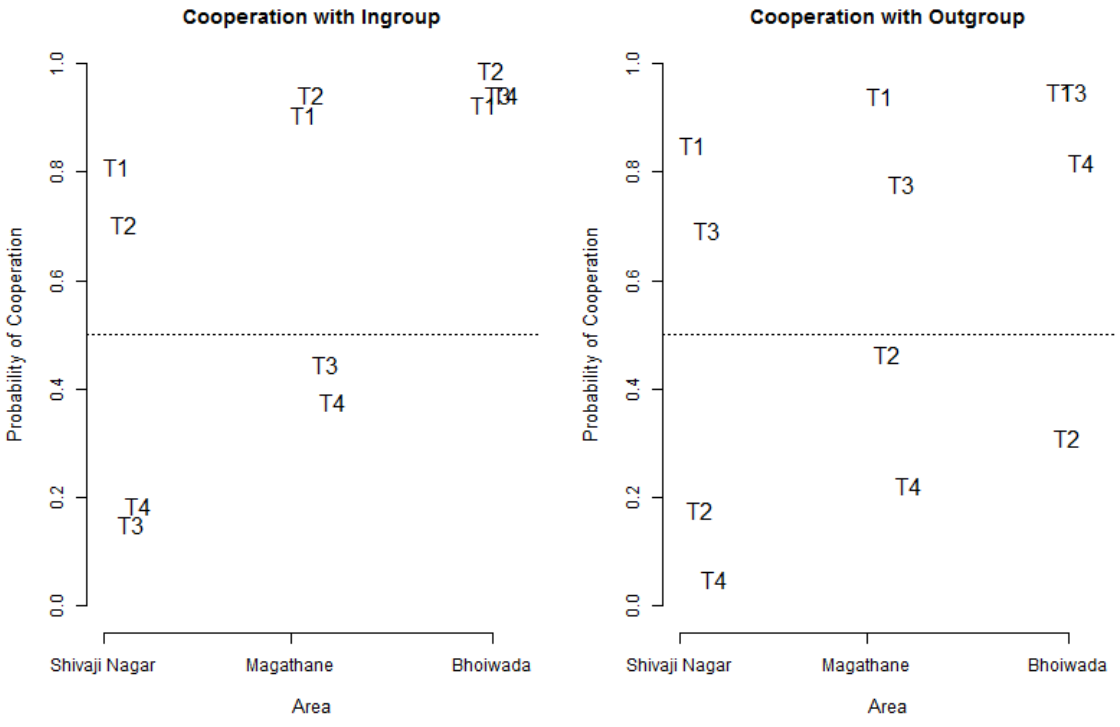
Three regression equations summarized in the left column of Figure 3 confirm what we learned from the difference-in-means estimate (Figure 1). Increased expectations of reciprocity in the T1 and T2 groups retained their strong, positive, and highly significant effect on cooperation in the public

goods game in two slums even when we control for all the covariates. The treatment effect could not be distinguished from zero in Bhoiwada, however. This was probably caused by a high baseline cooperation rate in this neighborhood and slower divergence in the first five rounds.

The updated expectation of reciprocity was also a strong and statistically significant predictor of cooperation with outgroup members (right column of Figure 3). The PCP reached 86% (Magathane) to 93% (Bhoiwada). The full model reduced error in comparison to the null model by 70% in Magathane, 68% in Shivaji Nagar, and 43% in Bhoiwada (PRE). The pseudo R^2 ranged from 0.59 to 0.74. As predicted by hypotheses 1 and 3, an increase in the reciprocity expected from the outgroup (T1 and T3) also led to more cooperation with this group. The treatment effect was statistically highly significant in the two slums, but, despite its considerable size, did not pass the usual 95% threshold of significance in Bhoiwada.

Expectations of reciprocity had a strong and statistically significant effect on cooperation. The following figure shows fitted probabilities of cooperation as a function of treatment and the subject's neighborhood, with all other variables held constant at their mean value (or median value in case of binary variables) as measured in the given neighborhood.

Figure 4: Predicted Probability of Cooperation Depending on Treatment Group



Probability of cooperation with a member of a reciprocal group is generally higher than with an nonreciprocal group and this result holds true regardless of whether the partner was a member of the ingroup or the outgroup. For example, the probability of cooperation with a member of an uncooperative group in Shivaji Nagar was lower than 20% for both Hindus and Muslims. If the group was seen as reciprocal, this probability increased more than threefold – again for both groups. In the slums of Shivaji Nagar and Magathane, the probability of cooperation is always higher than 50% (the dotted line in Figure 8) for the groups seen as reciprocal and always lower than 50% for those with lowered expectations of reciprocity. In Bhoiwada, treatment largely failed to influence behavior in the first five rounds. The results from this neighborhood became more in line with my hypotheses by the tenth round of the game, except for an unusually high cooperation rate in the T4 group. Possible explanations of this curious discrepancy are discussed below.

I argued that updating of expectations of reciprocity may be the real causal mechanism behind Allport's contact hypothesis. The experiment provides more evidence for this explanation. As we can see from Figure 4, cooperation in the public goods game varied widely depending on expectations of reciprocity. On the other hand, three of the optimal conditions of a positive contact (common goal, cooperative interdependence, support of the authorities) were held constant at a positive value across all four treatment conditions. Both subjects and their on-screen partners shared the common goal to earn money, depended on each other in this pursuit, and positive interaction was sanctioned by the experimenter. Due to this lack of variation, these three conditions clearly could not explain why the cooperation rate in the T4 group was typically three to four times lower than in the T1 group.

Let us now look more closely at the fourth optimal condition of a positive contact, equal status. In the post-treatment survey, participants were asked to put various ethnic groups living in Mumbai on different rungs of the standard MacArthur Scale of Subjective Social Status. Marathi people were perceived on average as having a significantly (p-value<0.001) higher social status (5.9 out of 10) than Muslims (4.4 out of 10). So, the fourth condition of a positive contact could not be truly satisfied in the case of these two groups. Since some individuals might have perceived Muslims as more or less equal in social status to the Marathi people, I also regressed cooperation with Muslims on the difference in social status between these two groups. The coefficient of relative social status was close to zero and statistically insignificant. This result rejects the idea that equality in social status leads to more cooperation. Nor did the social status differential mediate the causal effect of updated expectations of reciprocity, as interactions between social status and treatments were also insignificant. As three of

Allport's conditions did not vary and the fourth one did not exhibit the effect assigned to it by the contact hypothesis literature, the striking variation in the cooperation rate in my experiment can be attributed to the experimental manipulation of expectations of reciprocity, and not to intergroup contact – or positive intergroup contact – itself.

The effect of altruism on cooperation with Muslims could not be distinguished from 0. I tried to replace this direct behavioral measure of altruism with two attitudinal proxies for ethnocentric altruism from a survey: attachment to Maharashtra (and thus to the Marathi identity) and frequency of praying (indicating the person's attachment to the Hindu religion).[2] Coefficients of these two alternative measures remained statistically insignificant too.

As the final check, I looked whether the treatment effect was homogenous across different subgroups. Since I wanted to test my theory on the least-likely crucial case, it would be problematic if I found out that extremist voters were immune to the treatment for instance. That is why I replicated my regression analysis, but this time including interactions between treatments and two covariates of interest (religious affiliation and past voting for extremist parties). The effect of increased expectations of reciprocity seemed to be slightly stronger among Hindus, but none of the interaction terms was statistically significant. Moreover, none of the models with interaction terms improved the corresponding null model without them. The safest interpretation is that the treatment effect was homogenous for both Hindus and Buddhists. When I looked at the voters of extreme nationalist parties (Shiv Sena and MNS), they seemed to respond to the effect even more strongly than other subjects as far as cooperation with Muslims (but not Hindus) was concerned. But, again, the p-value of the interaction terms was far too large to reject the null hypothesis. This analysis demonstrates that the causal effect of updated expectations of reciprocity was indeed homogenous across various subgroups of my sample. This is an encouraging result in terms of external validity.

The post-treatment survey showed that updated expectations of reciprocity not only increased cooperation with individual members of the outgroup in the laboratory, but also dramatically changed attitudes about the outgroup in general. Although the experiment was ostensibly about “cooperation between people from different neighborhoods” and did not mention “Muslims” as a group at all, subjects generalized their experience with individual seemingly Muslim players on the computer screen to the Muslims as a social category. Surprisingly, changed prejudices could be easily discerned even

2 “Attachment” to one's identity is considered to be the best measure of identity salience (Sinnott 2006) and the wording of my question is routinely used by World Values Survey. For the model with the frequency of prayer, I included only Hindu subjects.

from the answers to Bogardus's harshest indicator of social distance, that is the question whether the participants would "accept a Muslim to close kinship by marriage". In India, marriage is one of the most revered institutions and it is characterized by strong ethnic endogamy. Persistence of the caste is clear whenever you open the "matrimonials" classifieds section of any Indian newspaper. Most of them still read like this one from the Times of India:

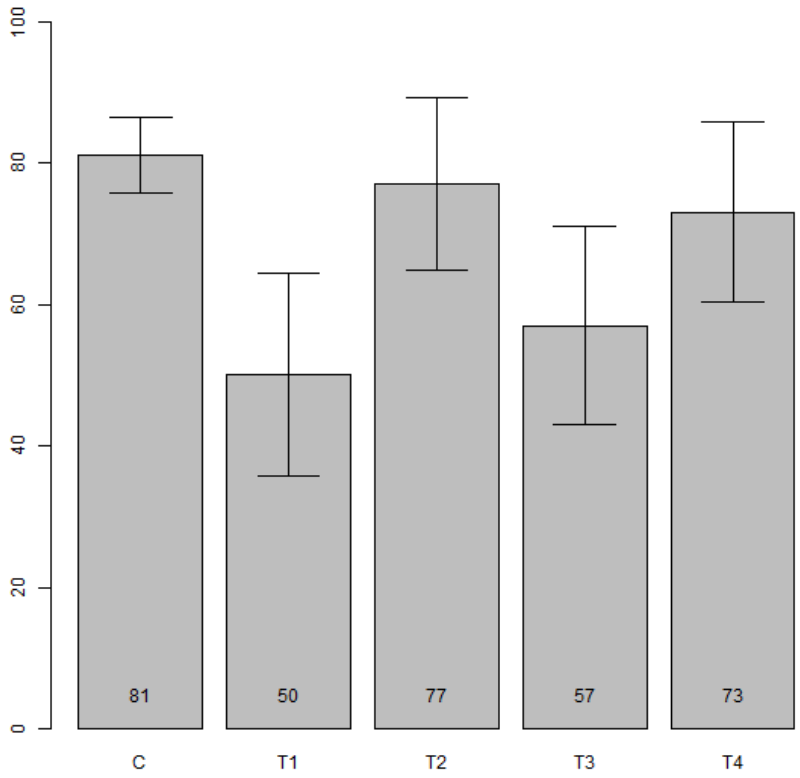
MBA 26 Yrs/5'4" fair & b'ful
Saraswat Brahmin Girl work-
ing in Bank-Mumbai, Seeks
equally qualified Veg. Brahmin.

This short classified mentions the caste ("Brahmin") and sub-caste ("Saraswat Brahmin") of the potential bride, as well as the required caste of the potential groom ("Brahmin"). As if it was not enough, it emphasizes a physical quality associated with the Brahmin caste ("fair" skin) and its religiously inspired dietary custom ("vegetarianism").

Given the sacrosanct character of marriage and importance of ethnicity in the selection of a marriage partner, I did not expect any effect of my experimental treatment on behavioral attitudes concerning marriage. And the results from the control group only reinforced this stereotypical view of the Indian society: 170 out of 210 respondents (81%) said that they would not accept a Muslim marrying into their family, while 209 out of 210 respondents would accept a Marathi. Although treatment did not change noticeably people's attitudes to marriage within their caste, it definitely increased their acceptance of intermarriage with Muslims (see Figure 5).

There is no statistically significant difference in attitudes between people in the control group and those, who encountered uncooperative Muslims during the experiment (T2, T4). However, participants, whose expectation of reciprocity from Muslims rose because of the experimental treatment (T1, T3), showed much more acceptance. The difference-in-means estimator of the average treatment effect (ATE) is fairly large – 27 percentage points or a 0.70 movement in standard deviation for the control group. There is no significant difference between the two groups with the increased expectations of reciprocity (T1 and T3). But if we compare only treatment groups that received positive (T1, T3) and negative (T2, T4) treatment, the difference between them is 21 percentage points, corresponding to 0.48 standard deviation movement from T2 and T4.

Figure 5: Percentage of Subjects Who Would Not Accept a Muslim Marrying into Their Family, with 95% Confidence Intervals



In the Indian context, this relaxed attitude to intermarriage with a group that is often seen as the sworn enemy in this conflict-ridden city only after a 15-minute positive interaction with alleged Muslims on a computer screen is truly astonishing. Rejection of Hindu-Muslim marriages is typically widespread even among people who generally support intermarriage between different Hindu castes. In the control group, 56% of those who would accept a Gujarati into a family, would still not accept a Muslim. (And 94% of those who would not accept a Gujarati would not accept a Muslim either.)

I replicated the same analysis using a softer measure of social distance, the question whether the respondent would accept a Muslim as a neighbor. Regardless of the different operationalization of the dependent variable, the treatment effect remained strong. Again, 71% of the control group said a resolute "No" to the question whether they would accept a Muslim as a neighbor. At the same time, 99% would accept a Marathi neighbor. As far as attitudes towards the Muslim outgroup are concerned, there is no statistically significant difference between the control groups and the two groups with the negative treatment (T2, T4), but the difference between the control group and the two groups with a

more positive experience (T1, T3) is highly significant. The proportion of people who would not accept a Muslim neighbor dropped by about a half, from 71% to 31% (T1) or to 41% (T3).

Although experimenters rarely ask their subjects about their inner motivations, I was interested in getting a glimpse of why respondents answered this question as they did. The very limited qualitative data (N=8) basically support the story that arose from the quantitative analysis. Only one respondent was among those, who would not mind a Muslim living in his neighborhood. His explanation nicely illustrated the causal argument proposed by this study: "We allow the Muslim people to be our neighbours because they are very honest and *help us* all the time in all the situations" (italics mine). Seven respondents offered a variety of explanations of why they did not desire any Muslims in their neighborhood: three did not believe in Muslims' reciprocity,[3] two were afraid of a perceived threat,[4] and two highlighted cultural differences between ingroup and outgroup.[5]

Potential Problems and Caveats

Qualitative data gathered right after the public goods game suggested that players were not as risk averse as I expected. Most of the subjects interviewed after the experiment confessed that they made their contribution hoping that they would get the maximum profit possible. This is how one member of the ill-fated T4 groups explained his strategy: "I went on with the investment of Rs. 10 every time in order to get the maximum profit. My partners didn't invest more than Rs. 0 and I ultimately ended in gaining no profit at all." What is intriguing, some subjects kept investing despite no contribution from their partners. As one of the subjects said: "My partner didn't invest but I didn't feel anything for that. Still I went on investing the amount." These dispassionate responses suggest that my study may have a problem with the house money effect: As the subjects were playing with the money given to them by the experimenter, they were more willing to take risks than if they had to earn the money by working (Thaler and Johnson 1990). The good news is that the potential bias introduced by higher propensity to take the risk at worst dilutes the estimated treatment effect. Risk-seekers probably

3 According to one respondent, Muslims "are not good people", "cannot behave nicely", and "always quarrel with their neighbours". According to another, they "are not good people by nature" and "cannot behave in good manner". Another respondent even called Muslims "goons by nature".

4 The most appalling, but also somewhat understandable answer in the context of the violent ethnic conflict in Mumbai was: "I do not accept the Muslim people as my neighbours because it is very difficult to survive beside them in the time of riots." The other respondent concerned with his security said: "Muslims are very cruel by nature which makes them unsuitable for anybody."

5 According to these two respondents, "their living standard is not well according to our culture" and "they give too much importance to their religion".

cooperate more even if they should not (i.e. when they expect the other player to defect). As a result, the difference between people who do and who do not expect reciprocity will at worst appear smaller than it is in reality.

This is what may have happened in Bhoiwada, the neighborhood with an untypically high cooperation rate in all the treatment groups. As this neighborhood is also more affluent than the two slums (with the average income more than twice as high as in Magathane and Shivaji Nagar), I checked whether income increased cooperation at the individual level. When I included dummy variables for neighborhoods in the regression, the coefficient of individual income became indistinguishable from zero. Whatever was driving the unusually high cooperation rate in Bhoiwada, it was not a low payout relative to the average income. It is possible that residents of Bhoiwada were more cooperative for some other reason. The civil society in the neighborhood was more active, with strong housing cooperatives and CBOs. Approximately two thirds of the respondents in Bhoiwada were aware of their community's effort to organize public goods provision (compared to about a half of respondents in Magathane and more than a third in Shivaji Nagar). However, it is impossible to test any alternative explanation without additional data.

Another problem frequently arises in experiments, when their research design violates the non-interference assumption and the treatment of one subject somehow influences the outcome of another subject (Gerber and Green 2012). The research design used in this study minimized possible spillover effects that may have otherwise lead to biased estimates of the average treatment effect in several ways. No communication between subjects was allowed and the research team was very vigilant in enforcing this rule. Moreover, anonymity was ensured by privacy screens around computers, so subjects could not just copy someone else's strategy.

In general, randomized experiments should satisfy two basic assumptions: non-interference and excludability, better known under the acronym SUTVA. The excludability assumption requires that each potential outcome depends solely on the received treatment and there is no other causal agent (Gerber and Green 2012). When we compare the T1-T4 groups, both assumptions are fully satisfied. But we must be more careful while comparing the four treatment groups to the control group. Clearly, subjects in the control group did not receive any treatment. But this was not the only difference. They did not play the behavioral games. They did not sit in front of a computer. They did not fill the questionnaire on a computer, but were interviewed by an interviewer. Finally, they knew that they would not receive any monetary reward for their participation in the survey. Due to all these

differences, the control group is not a true placebo group like in medical research. I believe that the survey administered in the control group reflects the real attitudes in the population. But it would not be safe to blindly attribute the measured differences between the control group and the treatment groups exclusively to the treatment.

On a happier note, both non-compliance and attrition rate were equal to 0. This wonderful result was achieved because subjects were motivated by nontrivial financial compensation and perhaps even more by an interesting and novel activity. During the debriefing, many subjects expressed their gratitude not as much for the earned money, as for the first opportunity to use a personal computer in their lives. That is why they took the experiment very seriously.

Unlike subjects sampled from usual subject pools (such as at universities or over the Internet), my subjects had no prior experience with behavioral experiments. There was only one experimental manipulation involved, which eliminated possible cross-effects. Experimental cross-effects occur "when subjects' choices in an experiment are influenced by the manipulations they have received in previous experiments" (Morton and Williams 2010). As the likelihood that any of the sampled individuals will ever participate in a behavioral experiment again is microscopic, my sampling strategy also prevents the so-called "pollution" of the sample by deception (Henrich 2001).

Another potential trouble with experiments conducted on a computer is that subjects very often do not believe that the other subjects, with whom they are interacting over a computer network, exist at all (Morton and Williams 2010). Although this is a frequent problem in developed countries and among "professional" subjects, it was not an issue in the slums of Mumbai. Among 192 participants who interacted with the computer, none expressed disbelief or doubts about existence of other players. Again, this result increases internal validity of my research because it makes the laboratory experiment more realistic (Aronson, Brewer et al. 1985).

Conclusion

The experiment showed that: (1) people cooperate if they believe that their partner will reciprocate their cooperative behavior; (2) people use their partner's ethnicity as an information shortcut to predict how likely reciprocity is; and, most importantly, (3) observation of individuals' real behavior can change the stereotypical beliefs about groups. By manipulating expectations of reciprocity, we achieved very different levels of cooperation in the laboratory. The post-treatment

survey showed that the effect of increased expectations of reciprocity in the behavioral games actually reduced prejudice in totally unrelated domains – ethnic segregation of housing and even the strictly regulated marriage market. This indicates that expectations of reciprocity are not only generalizable from the individual to the group, but also across situations.

The optimistic conclusion of this paper naturally leads to the question why we see so much ingroup bias and so little outgroup bias in the real life. One possible explanation is that it is easier to find and punish someone who exploited one's trust if the person belongs to one's own group (Habyarimana, Humphreys et al. 2007). In addition to this straightforward mechanism, information about the person's untrustworthiness can spread through his/her social network and reach potential partners that have not had any encounter with him/her. Since social networks are usually denser within than across ethnic groups, it should be easier to obtain information about the past of a coethnic (Fearon and Laitin 1996). Bounded groups with a clearly defined membership can also use the implicit threat of total exclusion from the group and thus from all the benefits associated with the membership (Hardin 1995). Laboratory experiments in different countries detected more cooperation between coethnics only when they were seen by others, and no preference for ingroup members in an anonymous setting, in which no punishment of defectors was possible (Yamagishi and Mifune 2008; Habyarimana, Humphreys et al. 2009). The dictator game described in this paper only confirmed these earlier results.

There are not many known ways to overcome ingroup bias. Even social psychologists themselves admit that recategorization and decategoriazation of social identities are hardly applicable outside scientific laboratories (Brewer 1997). Positive intergroup contact seems to work better (Pettigrew and Tropp 2006), but its restrictive conditions are difficult to achieve in practice. This study suggests that we should focus on one particular aspect of intergroup contact, which is reciprocity. If we build institutions that encourage positive reciprocity in indirect exchange between groups, we may be able to reduce their prejudices and, consequently, also discrimination. Although normative changes typically take many years, generalized reciprocity may be also ensured by strong rule of law and better internal security.

References

Ahn, T. K., E. Ostrom, et al. (2001). "Cooperation in PD games: Fear, greed, and history of play." Public Choice **106**(1): 137-155.

Alesina, A., R. Baqir, et al. (1999). "Public Goods and Ethnic Divisions." Quarterly Journal of Economics **114**(4): 1243-1284.

Alesina, A. and E. L. Ferrara (2005). "Ethnic Diversity and Economic Performance." Journal of Economic Literature **43**(3): 762-800.

Alesina, A., E. Glaeser, et al. (2001). "Why doesn't the United States have a European-style welfare state?" Brookings Papers on Economic Activity **2001**(2): 187-254.

Allport, G. W. (1954). The nature of prejudice. Cambridge, Mass., Addison-Wesley Pub. Co.

Aronson, E., M. B. Brewer, et al. (1985). Experimentation in Social Psychology. Handbook of Social Psychology. G. Lindzey and E. Aronson. New York, Random House**:** 441-486.

Arunachalam, B. (2005). "Drainage Problems of Brihan Mumbai." Economic and Political Weekly: 3909-3911.

Axelrod, R. M. (1984). The evolution of cooperation. New York, Basic Books.

Bajaj, V. and J. D. Goodman (2011). 3 Bombs Hit Crowded Sites in Mumbai. New York Times.

Belot, M., R. Duch, et al. (2010). Who should be called to the lab? A comprehensive comparison of students and non-students in classic experimental games. Discussion Papers, University of Oxford, Nuffield College.

Bhide, A. (2011). From The Margins: The Experiences of Low Income Communities in Mumbai vis-a-vis Changing patterns of Basic Service Delivery. Mumbai Reader '09. P. Joshi and I. Mathew, Urban Design Research Institute**:** 222-241.

Blalock, H. M. (1967). Toward a theory of minority-group relations. New York, Wiley.

Bogardus, E. S. (1925). "Measuring Social Distance." Sociology and Social Research **9**: 299-308.

Brewer, M. B. (1997). "The Social Psychology of Intergroup Relations: Can Research Inform Practice?" Journal of Social Issues **53**(1): 197-211.

Camerer, C. (2003). Behavioral game theory : experiments in strategic interaction. New York, N.Y. ; Princeton, N.J., Russell Sage Foundation ; Princeton University Press.

Dovidio, J. F. (1984). "Helping behavior and altruism: An empirical and conceptual overview." Advances in experimental social psychology **17**: 361-427.

Druckman, J. N., D. P. Green, et al. (2006). "The growth and development of experimental research in political science." American Political Science Review **100**(4): 627-635.

Easterly, W. and R. Levine (1997). "Africa's Growth Tragedy: Policies and Ethnic Divisions." Quarterly Journal of Economics **112**(4): 1203-1250.

Eckstein, H. (1975). Case-Study and Theory in Micro-Politics. Handbook of Political Science. F. Greenstein and N. Polsby. Reading, Addison-Wesley. **7:** 79-138.

Engineer, A. A. (1993). "Bombay Riots: Second Phase." Economic and Political Weekly **28**(12/13): 505-508.

Fearon, J. D. and D. D. Laitin (1996). "Explaining interethnic cooperation." American Political Science Review **90**(4): 715-733.

Gelman, A., A. Jakulin, et al. (2009). "A Weakly Informative Default Prior Distribution For Logistic And Other Regression Models." The Annals of Applied Statistics **2**(4): 1360–1383.

Gerber, A. S. and D. P. Green (2012). Field Experiments: Design, Analysis, and Interpretation. New York, W.W. Norton.

Gerring, J. (2007). Case study research : principles and practices. New York, Cambridge University Press.

Habyarimana, J., M. Humphreys, et al. (2007). "Why Does Ethnic Diversity Undermine Public Goods Provision?" American Political Science Review **101**(4): 709-725.

Habyarimana, J., M. Humphreys, et al. (2009). Coethnicity: Diversity and the Dilemmas of Collective Action. New York, Russell Sage Foundation Publications.

Hansen, T. B. (2001). Wages of violence : naming and identity in postcolonial Bombay. Princeton, N.J. ; Woodstock, Oxfordshire, Princeton University Press.

Hardin, R. (1995). One for all : the logic of group conflict. Princeton, N.J., Princeton University Press.

Henrich, J. (2001). "Challenges for everyone: Real people, deception, one-shot games, social learning, and computers." Behavioral and Brain Sciences **24**(3): 414-415.

Henrich, J., R. Boyd, et al. (2001). "In search of homo economicus: behavioral experiments in 15 small-scale societies." The American Economic Review **91**(2): 73-78.

Henrich, J., J. Ensminger, et al. (2010). "Markets, religion, community size, and the evolution of fairness and punishment." Science **327**(5972): 1480.

Henry, P. J. (2008). "College sophomores in the laboratory redux: Influences of a narrow data base on social psychology's view of the nature of prejudice." Psychological Inquiry **19**(2): 49-71.

Hinkle, S. and R. Brown (1990). Intergroup comparisons and social identity: Some links and lacunae. Social identity theory: Constructive and critical advances. D. Abrams and M. A. Hogg. New York, Harvester-Wheatsheaf**:** 48-70.

Hobhouse, L. T. (1906). Morals in evolution; a study in comparative ethics. London,, Chapman and

Hall, Ltd.

Hume, D. (2003). A Treatise of Human Nature.

Jervis, R. (1978). "Cooperation Under the Security Dilemma." World Politics **30**(2): 167-214.

Lawler, E. J. (2001). "An affect theory of social exchange." American Journal of Sociology.

McDermott, R. (2002). "Experimental Methodology in Political Science." Political Analysis **10**(4): 325-342.

Miguel, E. (2004). "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania." World Politics **56**(3): 327-363.

Miguel, E. and M. K. Gugerty (2005). "Ethnic diversity, social sanctions, and public goods in Kenya." Journal of Public Economics **89**(11-12): 2325-2368.

Molm, L. D., G. Peterson, et al. (2003). "In the eye of the beholder: Procedural justice in social exchange." American Sociological Review **68**(1): 128-152.

Morton, R. B. and K. C. Williams (2010). Experimental political science and the study of causality : from nature to the lab. Cambridge ; New York, Cambridge University Press.

Mumbai, M. C. o. G. (2010). Mumbai Human Development Report 2009. New Delhi, Oxford University Press.

Munshi, K. and M. Rosenzweig (2006). "Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy." The American Economic Review: 1225-1252.

Nowak, M. A. and K. Sigmund (2000). "Shrewd investments." Science **288**(5467): 819-820.

Nowak, M. A. and K. Sigmund (2005). "Evolution of indirect reciprocity." Nature **7063**: 1291-1298.

Ostrom, E. (2000). "Collective Action and the Evolution of Social Norms." The Journal of Economic Perspectives **14**(3): 137-158

Palshikar, S. (2004). "Shiv Sena: A Tiger with Many Faces?" Economic and Political Weekly **39**(14/15): 1497-1507.

Palshikar, S. and R. Deshpande (1999). "Electoral Competition and Structures of Domination in Maharashtra." Economic and Political Weekly **34**(34/35): 2409-2422.

Pettigrew, T. F. (1971). Racially Separate or Together? New York, McGraw-Hill.

Pettigrew, T. F. and L. R. Tropp (2006). "A meta-analytic test of intergroup contact theory." Journal of Personality and Social Psychology **90**(5): 751-783.

Piliavin, J. A. and H.-W. Charng (1990). "Altruism: A review of recent theory and research." Annual Review of Sociology **16**: 27-65.

Posner, D. N. (2005). Institutions and ethnic politics in Africa. New York, Cambridge University Press.

Rabasa, A., R. D. Blackwill, et al. (2009). The Lessons of Mumbai, RAND Corporation.

Schuman, H., C. Steeh, et al. (1985). Racial attitudes in America : trends and interpretations. Cambridge, Mass., Harvard University Press.

Sears, D. O. (1986). "College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature." Journal of Personality and Social Psychology **51**(3): 515-530.

Sears, D. O. and C. L. Funk (1999). "Evidence of the long-term persistence of adults' political predispositions." The Journal of Politics **61**(1): 1-28.

Shaban, A. (2010). Mumbai: political economy of crime and space. New Delhi, Orient BlackSwan.

Sinnott, R. (2006). "An Evaluation of the Measurement of National, Subnational and Supranational Identity in Crossnational Surveys." International Journal of Public Opinion Research **18**(2): 211-223.

Skyrms, B. (2004). The stag hunt and the evolution of social structure. Cambridge, UK ; New York, Cambridge University Press.

Srikrishna, B. N. (1998). Report of the Srikrishna Commission Appointed for Inquiry Into the Riots at Mumbai During December 1992 and January 1993, Govt. of Maharashtra.

Tajfel, H. (1969). "Cognitive Aspects of Prejudice." Journal of Social Issues **25**(4): 79-97.

Thaler, R. H. and E. J. Johnson (1990). "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice." Management science **36**(6): 643-660.

Varshney, A. (2002). Ethnic conflict and civic life : Hindus and Muslims in India. New Haven Conn., Yale University Press.

Wilkinson, S. (2004). Votes and violence: electoral competition and ethnic riots in India. Cambridge studies in comparative politics. Cambridge ; New York, Cambridge University Press.

Yamagishi, T. and N. Mifune (2008). "Does Shared Group Membership Promote Altruism?" Rationality and Society **20**(1): 5-30.

Yamagishi, T. and N. Mifune (2009). "Social exchange and solidarity: in-group love or out-group hate?" Evolution and Human Behavior **30**: 229–237.