Social networks and intergroup conflict

Takács, Károly

SOCIAL NETWORKS AND INTERGROUP CONFLICT

RIJKSUNIVERSITEIT GRONINGEN

SOCIAL NETWORKS AND INTERGROUP CONFLICT

Proefschrift

ter verkrijging van het doctoraat in de
Psychologische, Pedagogische en Sociologische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. D.F.J. Bosscher,
in het openbaar te verdedigen op
donderdag 30 mei 2002
om 14.15 uur

door

Károly Takács

geboren op 8 juli 1973
te Veszprém (Hongarije)

Promotores:	Prof. dr. F.N. Stokman
	Prof. dr. W.B.G. Liebrand
Referenten:	Dr. R. Popping
	Dr. A. Flache

Beoordelingscomissie:	Prof. dr. G. Bornstein
	Prof. dr. M.W. Macy
	Prof. dr. W. Raub

ISBN 90-367-1610-1

Acknowledgements

The book you hold in your hand is the product of the research I conducted as a graduate student at the ICS in Groningen between 1997 and 2002. I am grateful for the support I have received from many people during this time. First of all, I would like to thank my supervisors, Frans Stokman, Wim Liebrand, Roel Popping, and Andreas Flache for their valuable, thorough, and inspiring advises. They provided guidance and helped with ideas that enriched this dissertation significantly. As examples I could mention the suggestion of using multilevel analysis for experimental data, the introduction to the team games literature and to the program language Delphi, or the suggestion of a new way of graphical representation of team games. Although I was advised to change the personal pronouns in the text from the old-fashioned plural to singular tense, I kept the plural in order to indicate their close involvement in the research. After all, I remain strongly indebted to them.

I owe my thanks to the members of the Manuscript Committee, Gary Bornstein, Michael Macy, and Werner Raub, who read the dissertation carefully. Justine Horgan and Linda Karr read and corrected the English text. I am thankful for Carlijne Philips, Rita Smaniotto and Ferry Koster who translated the summary into Dutch and for Roel Popping who controlled the translation.

Several people helped me with comments on earlier versions of different parts of the text. For the theoretical part of the research I received suggestions from Thomas Gautschi, Henk Flap, Harry Ganzeboom, Tom Snijders, Wolfgang Balzer and two anonymous referees of the Journal of Conflict Resolution. I thank Marcel Lubbers, Carlijne Philips, Rafael Wittek, Marcel van Assen, Alexander Gattig, Christian Steglich, René Torenvlied, Corine Hoeben, Uwe Matzat, Béla Janky, Marieke van der Wal, Rita Smaniotto, Martin van der Gaag, Alinda van Bruggen, Justine Horgan, Annelies Kassenberg, and Jan Kratzer for their comments on the experimental design and/or on the analysis. I would like to thank Sicco Strampel for developing the computer program for the experiments, Jeanette Loos, Hans van Gestel, and Rie Bosman for their practical help. All subjects deserve thanks for participation.

I benefited a lot from discussions with Michael Macy, Brent Simpson, Marianna Klotschko, Robb Willer, Mat Salganik, and Yosimitchi Sato during the winter of 2001 that I spent at Cornell. Special thanks to Michael Macy, who was an excellent host in Ithaca and a major motivating force to proceed with this research.

I am grateful to my roommate, Carlijne Philips and to Csilla Horváth, Zsolt Sándor, Csilla Bartha, Jan Kratzer, Marcel van Assen, Marieke van Onna, Béla Janky, Rafael Wittek, Alessandra Ligi, Justine Horgan, Peter Mühlau, Judit Koczok, Mátyás Hunyadi, Péter Földi, Laura Szabó, Emese Bíró, Tamás Bartus, Philip Wotschack, Rita Smaniotto, Uwe Matzat, Javier Arregui, Norel Rosner, Barend Schipper, and many others who made my years in Groningen enjoyable.

My sister, my parents, and my friends in Hungary also deserve thanks as they continuously supported me despite of the geographical distance. Last but not least, I would like to thank the encouragement I received from my teachers in Hungary, from Zoltán Szántó, György Lengyel, and from László Csontos.

Remaining mistakes are mine.

Nemesvámos, April 2002

Károly Takács

Contents

"…neither cultural, psychological, nor economic differences are ***necessary*** *for the rise of intergroup conflict… Nor are maladjusted, neurotic, or unstable tendencies necessary conditions for the appearance of intergroup prejudice and stereotypes. The* ***sufficient condition*** *for the rise of hostile and aggressive deeds …was the existence of two groups competing for goals that only one group could attain, to the dismay and frustration of the other group."*

Muzafer Sherif: In Common Predicament. Social Psychology of Intergroup Conflict and Cooperation (1966a: 85)

CHAPTER 1

COMPETITION AND CONFLICT BETWEEN GROUPS

Introduction

1.1 Aims of this study

I passed by a graffiti in the center of Budapest during the Serb-Croatian war, which claimed that “Milosevic is a pig”. Under this text, there was a sagacious response also in English stating that “Human beings are the problem”. Almost a decade later, the graffiti is still there. Nobody repainted the wall; probably in order to leave the graffiti as a memento for the senseless bloodshed in the Balkans.

Human beings are the problem in the sense that despite the painful consequences, they choose to make sacrifices for their group and participate in an action that results in intergroup conflict. There is no need to raise extensive arguments as to why these harmful conflicts should receive special attention in the social sciences. It is also needless to say that not all intergroup relations end up in conflict. Just the opposite, we find that groups in most cases can live peacefully together (Fearon and Laitin, 1996; Gould, 1999). Therefore, the most challenging question for scientific research is to determine the conditions under which peaceful relations prevail. Developments in this direction could offer methods of conflict resolution and hopefully, might lead to policy advice and implementation.

There is no doubt that policy makers are in a strong need of suggestions for handling intergroup relations of several kinds. Recent examples of civil war, ethnic hostility, fights between soccer fans or between urban gangs bitterly show that conflicts often remain unresolved or are only solved with tremendous effort. International warfare of alliances of nations is an analogy on a larger scale and confrontation between pupil groups often causes a similar headache on a smaller scale. And this is by far not the end of the list of poisonous intergroup relations.

For proper advice to politicians we need a better understanding and a deepened explanation of violent confrontations. Sometimes we need to argue from the opposite point of view and have to concentrate on the emergence of peaceful coexistence, like the rise of a live-and-let-live system in the trenches in World War I. Sometimes in similar circumstances we do not have to worry about individual participation, but rather about the lack of participants, like in the case of voting in two-party democracies. All of these and many more examples relate to the very same problem of *competitive intergroup relations* that are based on individual contributions.

To determine the conditions for the emergence of intergroup conflict and peace we have to ascertain under which circumstances group members are willing to make contributions to harmful intergroup competitions. Furthermore, in order to do this, we have to explore the *underlying mechanisms* that drive individuals towards sacrifices against another group. This sort of logic is grounded on the principles of *methodological individualism* (Coleman, 1990) and it is in strong contrast with previous explanations of intergroup conflict that were built on macro mechanisms and are still very much in fashion. The logic of explanation in this study is also in contrast with theoretical and empirical research that examines individual causes, but fails to reveal the mechanisms behind these effects.

In general, little has been said before now about the social mechanisms that lead to intergroup conflict. As a major contribution, this study aims to highlight causal chains that stem from the *embeddedness* of human action (Granovetter, 1985). Individual decisions in the intergroup context are interdependent, embedded in the social structure, and embedded in a historical context. They are *interdependent*, because contribution to harmful intergroup competition has an effect on intergroup relations and consequently on the well-being of others. Moreover, it has also a direct effect on actions that others undertake. Such influences are especially strong among friends, neighbors, colleagues, and family members, which means that the network of individual relationships constrains individual actions. This explains why *structural embeddedness* has an impact on intergroup relations. On the other hand, lethal clashes in the past and prospects of future relations have firm effects on present attitudes and actions in the intergroup context. This is the reason why *temporal embeddedness* plays a crucial role in intergroup competitions.

The question of how interdependence and embeddedness influence decisions in the intergroup context and consequently the outcome of intergroup relations is central in this study. This question frames the theoretical model building in the research, as well as the methodology we apply to detect fundamental processes. Among effects of structural embeddedness, particular attention is devoted to the effect of *segregation* and to the underlying *social control* mechanisms. With regard to temporal embeddedness, we are mainly interested in tracing typical *scenarios* of intergroup relations and individual *behavioral heuristics* that are responsible for the changes. We use *simulations* and laboratory *experiments* as research tools that are quite unconventional in this field. We do this with the aim of better understanding, explanation, and prediction of intergroup conflict and peace.

In this chapter we summarize the objectives of this study and give an introduction to the analysis of intergroup conflict and peace. Section 1.2 introduces the general research problem. In Section 1.3 we provide an overview of related theories of intergroup conflict and attempt to highlight some fundamental problems of existing approaches. In Section 1.4 we discuss why we line up behind the *team games model* of intergroup relations. A major scientific relevance of this study lies within the incorporation of effects of structural embeddedness into the team games model. The background of this model development is discussed in Section 1.5 and the cornerstones of the model extension are introduced in Section 1.6. A discussion about the effects of temporal embeddedness follows in Section 1.7. The main research questions of this study are formulated in Section 1.8. The primary motivations for using simulations and experiments as methodological tools are presented in Section 1.9. Finally, we provide an outline of subsequent chapters.

1.2 Research problem

1.2.1 The nature of intergroup conflict

In our attempt to determine underlying mechanisms that influence the likelihood of intergroup conflicts, we first have to clarify what kind of conflicts we study. There are many forms and reasons of intergroup conflict, too many and too different to provide a general theory for all of them. Being less ambitious, we restrict our interest to conflicts between two groups that are *unintended* consequences of conscious and voluntary actions of group members. Unintended conflicts are suboptimal outcomes compared to other outcomes, like peaceful coexistence. Examples are violent clashes between ethnic groups in residential neighborhoods or in villages and harmful competitions between pupil groups. We would like to explain such outcomes without assuming that group members actually seek harmful outcomes. As we have a limited scope of interest, we will use the term *intergroup conflict* only in the following restricted meaning.

Definition 1.2.1.1 Intergroup conflict is an aggregated consequence of individual contributions to an action of the group that hurts the interests of the other group.

The definition presumes that we consider intergroup conflict as an *aggregated* consequence of *individual* contributions. The definition applies to situations in which two distinct groups are involved. We emphasize that we do not consider groups as unitary actors, but as collections of individual decision makers who have the right to decide about their contributions. Furthermore, the definition also specifies the relationship that exists between the two groups. If an action of one group hurts the collective interests of the other group, there is a *negative interdependency* between the two groups. Negative interdependency occurs in *intergroup competition* situations, in which territory, power, economic and social incentives, or other scarce resources are at stake. For instance, opposite sides in civil war may fight for power or street gangs may contend to obtain social identity. These situations are symmetrical in the sense that a competitive action of one group hurts the other group and a competitive action of the other side also hurts the first group. If both sides act competitively, it results in a mutually harmful outcome that is worse for all than the lack of competitive action. For example, an endless war or bloody street battles is worse than living in peace for both parties.

Situations, in which intentional actions of individuals result in suboptimal outcomes for everyone, are described as *social dilemmas*. Consider the following more formal definition given by van Lange et al. (1992: 4):

Definition 1.2.1.2 Social dilemmas can be defined as situations in which each decision maker is best off acting in his own self-interest, regardless of what the other

> persons do. Each self-interested decision, however, creates a negative outcome or cost for the other people who are involved. When a large number of people make the self-interested choice, the costs or negative outcomes accumulate, creating a situation in which everybody would have done better had they decided not to act in their own private interest.[1]

Under certain conditions intergroup conflict is a social dilemma, but a special one in which people's *contributions* are in their own self-interest. Contributions are made to a *collective action* of the group. The collective action, however, creates a negative outcome or cost for members of another group and a gain for members of the own group. If a large number of members in both groups contribute to the group actions, then everybody is worse off than they would be in the absence of contributions. Hence, unlike in single group social dilemmas, *contribution is the undesirable action* from the collective point of view. For instance, in civil war, the choice of shooting with many comrades is a contribution to a collective action that hurts the interests of the enemy. When there are enough warriors in both groups, balanced fighting results in bloodshed.

The question now is how can contribution be the self-interested choice. This is a challenging question as in single group collective action *not contributing* is the unconditional best strategy. The structure of interdependency within the group is not different in intergroup related collective action. The difference is the competition between the groups. In order to explore the role of intergroup competition and examine under what conditions it can provide a structural solution for the collective action problem within the group (cf. Bornstein, Erev, and Rosen, 1990; Erev, Bornstein, and Galili, 1993), we need to integrate modeling interdependencies within the group and between the groups. These interdependencies cannot be separated, therefore should not be handled separately. It is quite surprising that this association has been hardly made before in research on intergroup conflict.

1.2.2 Embeddedness and intergroup conflict

The fundamental question is why do people join collective actions that are harmful in the intergroup context, if they have to make sacrifices. This is the question we need to answer in order to find an explanation for the emergence of lethal conflicts between groups.

Specification of all factors that possibly influence individual decisions and consequently the likelihood of intergroup conflict goes far beyond our possibilities. We concentrate on the *competition* between the groups for certain scarce resources and on the *structural* and *temporal embeddedness* of individual actions. First, individuals are encouraged to participate in conflict by rewards of intergroup competition and

[1] There are also less restrictive definitions of social dilemmas that do not require that following self-interest is the unconditional best strategy (Liebrand, 1983; Flache, 1996: 3).

comparison, such as territory, pride, or social identity. Second, they gain merits from social connections and relationships with others in their group. Third, they are mobilized by revenge, experience, and other factors based on the past, and expectations, fear, and other factors based on the future.

The analysis of *structural embeddedness* in intergroup conflict is an original aspect of this study. When individuals decide to participate in conflict, they are influenced by the presence, opinion, and behavior of their friends, neighbors, colleagues, and their family. They also receive relevant information through these connections. Actors and the interpersonal connections among them are referred to as the *social network* (Wasserman and Faust, 1994: 9). Surprisingly, research on intergroup conflict until now has disregarded social network effects almost completely (exceptions are Flap, 1988; Nelson, 1989; Gould, 1999; Bhavnani and Backer, 2000). By focusing on the relevance of networks of social relationships we can gain a better understanding of conflicts in different structural environments. Examples are how different residential structures affect the likelihood of ethnic conflict, how seating policies influence the likelihood of conflict between supporters in a stadium, as well as which settlement policies can help conflict resolution. Since there are many possible structural configurations and several network properties that can be important in intergroup relations, we should specify which structural characteristics are the focus of our exploration.

Inspired by empirical puzzles and debates, our research problem about social network effects is primarily centered around the effect of *segregation* on intergroup conflict. The level of segregation is conceptualized as the extent to which group members have ties among each other compared to all connections in the social network. Consequently, the question why and under what conditions segregation increases the likelihood of intergroup conflict is equivalent to asking why and under what conditions dense within group relations and scarce intergroup relations support the emergence of intergroup conflict.

There are already indications in classical sociology that segregation results in a higher level of intergroup-tension, which makes conflicts more likely (Simmel, 1955[1908]; Sumner, 1974[1906]; Coser, 1968[1956]). In the academic debate over school desegregation, for instance, Blalock (1986) has emphasized the effect of racial segregation on interracial tension. The implementation of school desegregation policy in the seventies in the United States brought improvements to interracial relations (Blalock, 1986; Granovetter, 1986). There is also empirical evidence of the effect of residential segregation on ethnic conflicts (cf. Harris, 1979; Diez Medrano, 1994; Hasson, 1996; Olzak, Shanahan, and McEneaney 1996).

On the other hand, some residential studies debate that segregation in cities would result in the emergence of ethnic conflicts (Duncan and Duncan, 1955; Lieberson, 1963). An outburst of ethnic violence at residential division lines with a mixed population and self-sustaining migration movements towards segregation evoked government policies supporting residential *separation* (Belfast, Jerusalem, Usti nad Labem, Mostar, or Kosovska Mitrovica are recent examples). Building walls and destroying or closing bridges together with a deployment of armed forces, although radical tools to achieve interethnic peace, they nevertheless provided a solution. This solution often seems only temporary as tension between the opposing ethnic groups is far from being extinguished. In other cases, separation is successful, because intergroup conflict ends with a disappearance of its object (cf. Simmel, 1955[1908]). If negative interdependencies are unavoidable and intergroup competition situations occur over and over again, separation policy can even support lethal conflicts.

It is likely that there is no general answer for whether or not desegregation is an effective policy to decrease violence between groups. In this study we address the question *why and under what conditions segregation increases the likelihood of intergroup conflict*. Since segregation is the aggregation of ego-network attributes and intergroup conflict is an aggregation of intentional individual decisions, the proper way to find explanatory mechanisms is to look at the relationship between ego-networks and individual decisions. Therefore, we have to find an explanation to *why and under what structural conditions individuals are willing to contribute to an in-group collective action that hurts the interest of another group.* In this explanation, we have to reveal the micro mechanisms that cause the effects of structural relations. This also helps to specify how the segregation effect depends on the nature and content of interpersonal relationships.

What complicates our analysis is that groups rarely face only one single competition situation. As they live alongside one another, competition situations occur repeatedly. There might be changes in the nature of interdependency, but often similar situations are recurrent. On one hand, previous outcomes certainly play an important role in subsequent competitions, as they influence forthcoming decisions. On the other hand, foresight of future intergroup and social relations also poses limitations on current actions. Hence, our investigation cannot be complete without considering the effects of *temporal embeddedness*.

General observation in the literature about dynamic intergroup relations is that there are two likely scenarios. One, which fortunately occurs more often, is long lasting peace and the other is a lethal spiral of violence (Fearon and Laitin, 1996). We are interested whether and under what conditions these paths occur and what are the underlying mechanisms causing them. In order to help conflict resolution, we will attempt to show the conditions that best promote the evolution of peaceful scenarios and reduce the emergence of durable conflict. We are interested in *how and why these scenarios emerge*. Furthermore, we examine *how social networks and how the consideration of past and of expected future outcomes influence the dynamic process*. For answers, we have to consult again mechanisms at the micro level. We have to find an explanation to *how individuals behave in repeated intergroup competitions; why and under what conditions they choose a strategy that includes unconditional or conditional contributions to repeated collective action that hurts the interest of another group*.

1.3 Theories of intergroup conflict and peace: an overview

In the previous section, we specified our general research problem. In this section, we briefly discuss a selection of previous research on intergroup conflict. We will also consider studies that use the concept of intergroup conflict in a broader sense or in a different way than we do, since their findings might be instructive for this study. This overview is deliberately biased towards theories that are directed at intergroup competition, social networks, and historical effects as factors that are related to intergroup conflict and peace.

We structure this selection along the division line between macro and micro approaches, although the border between them is probably not as unambiguous as we propose. Furthermore, there is an overlap with the categorization offered by Lindenberg (1997). According to Lindenberg, interdependencies that make group relations interesting can be classified as *functional*, *cognitive*, and *structural* interdependencies. First, we discuss some classical theories in macrosociology that concentrate on functional interdependence in intergroup relations. Second, we turn to highly influential thoughts in the social psychology of intergroup relations that emphasize cognitive interdependencies. Third, we target microsociological and economic approaches that cope with functional and structural interdependencies.

Classical theories mainly focused on *macro explanations* and on *functional interdependence* in intergroup relations. The main concern of the seminal work of Simmel (1955[1908]; 1955[1922]) was the functional role of intergroup conflict on cohesion and solidarity within the group. Besides internal solidarity, another product Simmel emphasized was the integration of the group (Simmel, 1955[1908]: 91): "…the group as a whole may enter into an antagonistic relation with a power outside of it, and it is because of this that the tightening of the relations among its members and the intensification of its unity, in consciousness and in action, occur." Simmel also discussed the content of interpersonal relationships that drives towards participation in conflict. He argued that individuals have a limited choice in conflict situations, as the force to comply to a uniform action is very strong (Simmel, 1955[1908]: 92-93): "A state of conflict … pulls the members so tightly together and subjects them to such a uniform impulse that they either must completely get along with, or completely repel, one another."

Drawing on the work of Simmel, *realistic conflict theory* was oriented towards the functions of intergroup conflict (Coser, 1968[1956]; 1967). As Coser (1968[1956]) claimed, a primary function of conflict is that it establishes and maintains group identities and boundaries. Furthermore it is a basic source of internal solidarity. Besides the functions, the sources of conflict are also of central interest in realistic conflict theory. The starting point of realistic conflict is intergroup competition (Williams, 1947; Blumer, 1958; Coser 1968[1956]; Sherif, 1966a). Groups compete with each other for certain scarce resources and the scarcity of these resources makes intergroup conflict "realistic". Groups try to obtain these resources, because it is in their economic (or any other type of) interest. For instance, religions are in ideological conflict with each other when they claim to be the one and only true religion (Allport, 1954). In general, hostility towards the competitive group is just the means for obtaining the scarce resources for the in-group. Hence, groups involved in realistic conflict are in a certain sense rational. Still, this approach has nothing to do with rational individual choice (cf. Coenders, 2001: 14), as it considers the groups as the units of analysis and not the individual members of the groups.

Realistic conflict theory was extended in the individualistic direction by Blalock (1967). Blalock handled actual intergroup competition for scarce resources at the macro level separately from individual actions. In his model, individuals are mobilized if they *perceive* intergroup competition or an out-group threat. This perception does not necessarily have to coincide with an actual competition (e.g., LeVine and Campbell, 1972: 41).

The analysis of Richardson (1948a; 1948b) did not consider the group as the unit of analysis, but was based on statistical proportions of individuals with different moods in the rival groups. His study examined the effect of time on the number of persons in different *war-moods* in two opposing nations. Richardson explained changes in these numbers by an underlying mechanism that is analogous to epidemics. Friendly and hostile attitudes spread irreversibly as a disease or fashion and result in transitions from one mood to another. When appropriate transition rules and starting parameters are assumed, the typical phases of symmetrical war can be deduced. For certain values of the parameters a balanced state of armament levels is derived. The other two typical scenarios are escalation of a runaway arms race and complete disarmament (Richardson, 1951; 1960). However, just like the classical macrosociological theories, these derivations fail to reveal the micro foundations of intergroup conflict and peace.

Individuals and *cognitive interdependencies* are the focus of theories about intergroup conflict and attitudes in *social psychology*. On the other hand, these theories tend to neglect the importance of functional interdependencies between the groups that is at least recognized by macrosociological approaches.

The positive association between positive attitudes toward in-group members and negative attitudes toward out-group members is emphasized by the theoretical concept of *ethnocentrism* (Sumner 1974[1906]; LeVine and Campbell, 1972; Brewer, 1981). Sumner (1974[1906]) claimed that this association is universal and every group has a syndrome of ethnocentrism. In this respect, there is sufficient supporting evidence in the empirical literature, for instance, the studies of urban gangs (Suttles, 1967; Jankowski, 1991: 88).

Sherif in his "*theory of conflict*" focused on the emergence of negative out-group attitudes and intergroup hostility (Sherif, 1966b; Sherif and Sherif, 1969). In line with realistic conflict theory, he emphasized that competitions between groups have a fundamental impact on the evolution of group structures, on negative out-group attitudes, and on the emergence of hostile behavior (Sherif, 1966a; 1966b). In his field experiments in a young boys summer camp, he showed how sport competitions and segregation led to a greater distance between the groups, to prejudice, and even to occasional hostilities (Sherif, 1966b). Sherif also attempted to find possible resolutions of conflict. Contact between members of the groups that does not involve interdependence among them did not improve intergroup relations; occasional meetings just provided a place for hostile behavior and clashes. On the other hand, the establishment of superordinate, common goals fairly improved the situation. As groups faced a series of cooperative interdependencies, hostilities were disappearing (Sherif, 1966b; Brewer, 1996a).

For the emergence of in-group favoritism and out-group hostility it is not required that groups are organized or have an established set of norms. In the pursuit of minimal conditions that are sufficient to facilitate intergroup conflict, Tajfel and his colleagues conducted a series of *minimal group experiments* (Tajfel, 1970; Tajfel et al., 1971; Billig and Tajfel, 1973). In these experiments, there was no social interaction between the subjects, furthermore the groups were created on a cognitive basis and therefore they are referred to as “minimal”. *Categorization* based on the estimation of the number of dots in a drawing or on the preferences for the paintings of Klee and Kandinsky was already sufficient for the emergence of in-group favoritism. Besides this highly influential result, a pragmatic value of minimal group experiments was that they were able to distinguish and control for different effects, unlike field experiments, where it is also difficult and partly unethical to create “real” social groups.

The minimal group experiments stimulated the formation of *social identity theory* (Tajfel, 1981; 1982; Tajfel and Turner, 1986). The basic premise of social identity theory is that individuals strive to achieve or maintain a satisfactory image about them and an important aspect of self-definition is group identification. Social identity is primarily relational and comparative in nature as it is based on intergroup comparison. The polarization of the noble in-group image and the evil out-group picture provides positive social identity. The need for a positive psychological group distinctiveness can be achieved through social comparison. Intergroup competition provides this comparison, which explains the emergence of in-group favoritism. Individuals invest emotional energy to develop social identity and this might be the basis of their behavioral decision to participate in actions of the in-group. In these cases social identity is not taken into account purposefully, however the strive for social identity can indirectly explain why group members are mobilized (cf. Macy, 1997).

The follow-up of social identity theory has taken different routes. On one side, *self-categorization* theory emphasized intergroup behavior in terms of underlying cognitive representations. Radical views expressed that individuals are “transformed into groups” via the process of self-categorization (Hogg and Abrams, 1988: 21), meaning that this cognitive mechanism is sufficient for thinking in group terms. On the other side, critical remarks maintain that social identity is obtained conditionally, only if the group is seen as an acting social unit (Horwitz and Rabbie, 1982). This approach gives more attention to the instrumental character of intergroup relations in which one group competes with another group to achieve valued goals (Rabbie, 1982). Some critics in political science argue that the establishment of group identity might have no negative effect at all on the other group. An example is the distinction between the concept of “healthy” patriotism and harmful chauvinism at Kosterman and Feshbach (1989).

These approaches in social psychology emphasized the role of cognitive interdependencies in intergroup relations. On the other hand, they disregarded macro relations between the groups and functional and structural interdependencies.

The individualistic perspective of social psychology and the emphasis on functional and structural interdependencies between and within the groups is combined in some *microsociological* and *economic* theories. For this study, these research directions are most essential, as we intend to explore the micro foundations of intergroup conflict and we have only a side interest in analyzing cognitive interdependencies.

Segregation and its evolution were the focus of some pioneer models in *rational choice* sociology. Models of Schelling (1971; 1978) and Esser (1986) have demonstrated how residential segregation can evolve as an unintended result of intentional individual action. These models, however, concentrated on neighborhood preferences as a source of segregation and they did not specify what was wrong with segregation and how it was related to intergroup relations.

Economic incentives are emphasized in intergroup relations by *ethnic competition theory* (Gellner, 1983; Olzak, 1986; 1992). Ethnic groups are considered as effective coalitions that are formed to extract material benefits from others or to defend possessions. This is also the source of ethnic competition and violence. The stronger the ethnic competition, the more severe the attitudes towards the competitor group are (Olzak, 1986). A core hypothesis is that ethnic collective action is intensified by desegregation of the labor market, as employment possibilities are scarce and they are also the targets of ethnic competition (Olzak, 1992: 3). Ethnic groups (and also other groups), however, have fixed group boundaries and membership is not a question of choice. As entry and exit is blocked, ethnic networks enjoy several advantages, like trust, cohesion, and easier establishment of collective action (Wintrobe, 1995).

A substantial element of intergroup relations that was neglected by all models discussed so far is *strategic interdependence*. As a formal study of strategic interdependence, noncooperative *game theory* is ideally suited for sociological concerns (von Neumann and Morgenstern, 1944; Luce and Raiffa, 1957; Schelling, 1960). A contradiction between individual and collective interests is modeled by social dilemma games, including the Prisoner’s Dilemma (Rapoport and Chammah, 1965; Axelrod, 1984; Poundstone, 1992). Considering the wide range of societal applications, it is no surprise that for the study of individual behavior in these situations a new research field has evolved (e.g., Dawes, 1980; Liebrand, 1983; van Lange et al., 1992; Kollock, 1998). *Social dilemma research* uses game theory because it provides an adequate model for the interdependence of actions and not due to its key assumption of rationality (Macy, 1991b: 810).

The strategic interdependence of individual decisions in intergroup relations is emphasized by recent models of Fearon and Laitin (1996), Gould (1999), and Bhavnani and Backer (2000). These models also recognize that ways of resolving intergroup conflict are difficult to find, if explanations are only provided for conflict and not for intergroup (interethnic) peace. Furthermore, individuals have contradicting interests considering intergroup and within-group relations. Conflict between the groups often originates from an establishment of an in-group *collective action* (Gould, 1999; Bhavnani and Backer, 2000). Successful mobilization within the group is difficult to achieve; violent competitions are therefore less frequent than peaceful coexistence.

We follow this research path in this study. In order to explain the emergence of intergroup conflict and peace, we develop an explanation for individual participation in conflict. For this purpose, we base our theoretical model on the *team game* approach that considers strategic interdependence of individual actions both in the intergroup context and within the group. The subsequent section provides an introduction to this approach. In this way, this research diverts from explanations of group conflict that disregarded the purposefulness of individual action and also from studies that failed to recognize the conflict between individual and group interests.

1.4 Competitive intergroup relations modeled as team games

The theories introduced in the previous section are all lacking one or more aspects of what we believe an adequate theory on intergroup conflict should have. First of all, groups cannot be handled as unitary entities as they consist of consciously acting individuals who primarily care about their own gains and losses. The most challenging problem for research is when intergroup conflict is an aggregated outcome of voluntary and intentional individual actions. The theoretical foundations we choose for our analysis should take this into consideration.

The explanation should be supported by a theoretical model that gives meaningful predictions of the conditions under which conflict and peace are likely to occur. The appropriate model should represent the duality of within-group and intergroup interdependencies. Furthermore, the explanation of intergroup conflict should start from a simple and abstract theoretical model (Lindenberg, 1992). In this section, we argue what should be the key elements of a simple model of intergroup competitions and we introduce the theoretical framework that would be the departure point of the theoretical investigation in this study.

Considering within group interdependency as a *collective action problem* is one key element in the model of intergroup relations. Either looking at civil war, violence between football supporters, or fights between urban gangs, participation in conflict involves high costs and risk for group members. Individuals have to sacrifice time and sometimes money to take part. They can have serious injuries and might even risk their lives. If they do not participate and leave others "doing the job", they can still enjoy the benefits. Benefits of intergroup competition (for instance, public happiness, pride, or social identity) can be considered as group *public goods,* because there is no rivalry in consumption and no group member can be excluded from the consumption. This is the reason why individuals are able to free ride on the effort of others and the within group interdependency can be considered as a collective action problem. Individual contributions are costly, therefore group beneficial collective action is difficult to achieve.

What makes intergroup conflict different from other collective action problems is the *interdependence between the groups*. As it was suggested first by realistic conflict theory (see previous section), the origin of intergroup conflict is that the groups compete for obtaining certain scarce resources. Examples of such resources are territory or economic control. However, intergroup competition can also be heated by immaterial benefits, like pride, avoidance of shame, or social identity. Benefits to football hooligans are certainly only immaterial and consists of, for instance, the pride of being the most feared ultras. As an immaterial benefit that exists only in intergroup relations, the striving for social identity can explain group-beneficial behavior in minimal groups (Tajfel, 1970). It can also provide a reason for why groups are more competitive and aggressive than individuals (cf. Kramer and Brewer, 1984; 1986; Tajfel and Turner, 1986), which is referred as the *discontinuity effect* (Schopler and Insko, 1992; Insko et al., 1993; 1994). Experiments also confirmed that when they are playing alone and when they are members of groups, individuals behave differently (Bornstein, Erev, and Rosen, 1990; Schopler and Insko, 1992; Insko et al., 1994; Bornstein and Ben-Yossef, 1994). Subjects were more inclined to make sacrifices for their group to win from another group than they were for themselves in two-person situations.

Which group is able to realize the benefits of intergroup competition depends on an *intergroup comparison* of strength and efficiency of mobilization of group members. Practically, one of the major determinants is the number of people mobilized. Meanwhile the winning group obtains the benefits, the other group can suffer from negative consequences (for instance, loss of resources, shame, humiliation). For instance, successful recruitment of volunteers into paramilitary units creates possible gains for the in-group, but certainly hurts the interests of the out-group. Both sides are punished in the case of mobilization of a similar strength. Endless civil war, unsettled disputes, mutually harmful clashes are such examples where no benefits are realized. These outcomes are worse for everyone compared to no mobilization and peace.

This character of intergroup interdependence resembles the two-person *Prisoner's Dilemma* (e.g., Luce and Raiffa, 1957; Rapoport and Chammah, 1965; Axelrod, 1984; Poundstone, 1992). If groups were unitary entities and they could choose between mobilization and no mobilization, mobilization would be their *dominant strategy*. This means that regardless of the action of the out-group, the in-group would gain more with mobilization. Following the dominant strategy by both sides would result in a suboptimal outcome.

However, groups do not fight with each other in every situation. They are often unable to establish collective action because of ramifying interests within the group. This way, within group interdependencies can help to solve conflicts between the groups. Hence, there are two different perspectives of the situation. Conflicts between groups can be resolved by interdependencies within the groups. On the other hand, intergroup competition can be considered as a possible structural solution to social dilemmas within the group (Bornstein, Erev, and Rosen, 1990; Erev, Bornstein, and Galili, 1993). To summarize, the model of intergroup competition should capture this duality of within-group and intergroup interdependencies and should represent the mixed motives of individual decisions.

Until now, however, *theories of group conflict hardly made any connection to collective action problems within the groups and theories of collective action overlooked intergroup relations*. A framework that would allow for a combined modeling is offered by the *team games approach* (Palfrey and Rosenthal, 1983; Rapoport and Bornstein, 1987). In team games, competitive group relations are represented by a game. There are two levels of the game. At the intergroup level, groups as aggregates face an interdependent situation. Group action is determined as a consequence of individual actions within the group. Individuals decide either to contribute to the group action (make a sacrifice for their group) or not.[2] Individual decisions are strategically interdependent, which is represented as *n*-person games at the within-group level.

In the *Intergroup Prisoner's Dilemma* (IPD) game (Bornstein, 1992; Insko et al., 1994; Bornstein, Winter, and Goren, 1996; Goren and Bornstein, 2000) individuals are always better off when they do not participate in the group action. Public rewards for group members increase with the difference between the number of participants in the own group and in the other group. Besides, if the numbers of participants are equal, higher rewards are distributed, in case more people are mobilized. In this game, no participation is a dominant strategy equilibrium that is suboptimal in comparison to the outcome in which everyone participates.

Similarly, in the *Intergroup Public Goods* (IPG) game (Rapoport and Bornstein, 1987; Bornstein, 1992) intergroup competition is based on the number of contributors. Public good benefits are distributed in the group with more contributors. In case the numbers of contributors are equal, scarce public good rewards are divided between the groups. In the IPG game, individuals do not have a dominant strategy, since there are situations in which a single decision changes the result of the competition. Still, in most cases, not contributing is a more beneficial option.

The IPG game nicely represents the dichotomy of interdependencies within the groups (provision of a public good) and between the groups (intergroup competition).

[2] In several intergroup competitions individual decisions are not binary, but continuous (for instance, how accurately warriors shoot in civil war).

On the other hand, the IPD game can illustrate the social trap character of harmful intergroup competitions. If both groups are able to mobilize their members effectively, the result of the competition is harmful for both sides. A modification of the IPG game with this character would result in a model that describes more closely lethal intergroup conflicts.

In this section, we introduced the team games approach as a model of intergroup relations. We discussed the key elements of this model and its advantages compared to other approaches. Based on these arguments we choose the team games model of intergroup relations as a departure point for theory development in this study.

Team games, however, *fail to specify the basic mechanisms that generate participation in intergroup conflict*. As the primary objective of this study is to explain why individuals participate in harmful intergroup collective actions, we need to overcome this deficiency of the team games model. Furthermore, *team games completely disregard the structural embeddedness of individual actions*. Behavior is embedded structurally, as everyone's behavior is to a large extent constrained by neighbors, friends, and the family, regardless of group affiliations. Not only the team games approach, but also other prominent theories of intergroup conflict neglected the role of interpersonal relations (including all that was discussed in Section 1.3). As we are interested in the explanation of intergroup conflict and peace, we need a model extension that can deal with behavioral constraints and influences of social networks.

For this reason, in this study *we will incorporate structural embeddedness into the team games model of intergroup relations*. Considering structural embeddedness in the analysis of intergroup conflict is a new development and a major contribution of this research. Before going into details about how we implement this, in the next section we will provide a brief overview of previous research that emphasized effects of structural embeddedness.

1.5 Effects of structural embeddedness

In order to reveal how structural embeddedness constrains individual behavior, in this section we discuss some theoretical suggestions and results of previous research. These recommendations might support the explanation of how interpersonal ties influence individual participation in intergroup related collective action.

With regard to interpersonal contact and interaction between members of the *opposite groups*, the *contact hypothesis* of Allport (1954) suggested that these are crucial to preserve good intergroup relations. In the presence of extensive interpersonal ties across group borders, the emergence of negative attitudes towards the out-group would be less likely. The contact hypothesis indirectly implies an increasing effect of *segregation* on the likelihood of intergroup conflicts. Supportive findings in empirical research showed highest intolerance and most competitive intergroup attitudes in homogenous ethnic enclaves (Anderson, 1983; Denitch, 1994; Massey, Hodson, and Sekulic, 1999). Although contact across group borders seems to be the key to the reduction of intergroup conflict, it works only conditional on the quality, frequency, and extensiveness of the contact and on the context in which it takes place (Brewer and Miller, 1984; 1996: 132; Brewer, 1996a; 1996b; 1999).

Interpersonal ties between members of the *same group* also influence the outcome of intergroup competition. As intergroup conflict involves collective action problems within the groups (see previous section), this study might build on the recommendations of research about social network effects in *collective action* situations. The importance of the structure of interpersonal contacts for mobilization in collective action is emphasized by both theoretical (e.g. Marwell, Oliver, and Prahl, 1988; Macy, 1991; Opp, 1991; Bonacich and Schneider, 1992; Gould, 1993a; Heckathorn, 1993; Flache and Macy, 1996; Chwe, 1999) and empirical studies (e.g. McAdam, 1986; Fernandez and McAdam, 1988; Chong, 1991; Finkel and Opp, 1991; Gould, 1991; 1995; McAdam and Paulsen, 1993; Opp and Gern, 1993; Sandell and Stern, 1998). It is widely believed that dense in-group relations help the establishment of collective action (Marwell, Oliver, and Prahl, 1988; Coleman, 1990: 318-320; Marwell and Oliver, 1993: 102; Gould, 1993a; Opp and Gern, 1993).

As social networks are known to greatly affect behavior in collective action situations, it is quite surprising that *there is almost a complete lack of research that addresses social network effects and the underlying micro mechanisms in intergroup conflict*. The main drawback of empirical studies is that they are purely descriptive and they do not test micro processes of social network effects and in particular, of segregation effects. On the other hand, empirical results show that segregation can be associated with higher likelihood of intergroup conflict (Blau and Schwartz, 1984; Whyte, 1986; Diez Medrano, 1994; Olzak, Shanahan, and McEneaney, 1996).

Explanations of intergroup conflict should not be satisfied with discovering merely associations between social network properties and intergroup conflict. For a deeper understanding, *underlying mechanisms* of social network effects should be revealed. As these mechanisms work at the interpersonal level, the role of social contacts and their influence on participation decisions in intergroup competition should be specified.

Network ties between individuals work as constraints to individual decisions. These constraints are represented as interpersonal interdependencies in *local interaction games* (e.g., Ellison, 1993; Berninghaus and Schwalbe, 1996; Mailath, Samuelson, and Shaked, 1997; Chwe, 1999; Morris, 2000). Local interaction games are played between connected individuals and might have different payoff structures, for instance, competitive structures between enemies or coordination games between fellows. This approach is highly relevant for the purpose of this study, as local interaction games also deal with individual decisions that are embedded in a network of dyadic interdependencies.

There are several other models and theories that might provide suggestions for how to conceptualize underlying mechanisms and constraints posed by network relations. Prominent examples are the theory of social impact (Latané, 1981; Nowak, Szamrej, and Latané, 1990), models of social influence (e.g., Carley, 1991; Friedkin and Johnsen, 1999), and the model of dissemination of culture (Axelrod, 1997b). A concept that embraces a wide range of micro mechanisms and specifies the content of dyadic interdependencies is *social control* (Heckathorn, 1990; 1993; Macy, 1993a).

Definition 1.5.1 Social control is defined as a constraint on individual decision posed by the influence as well as the presence, opinion, expectations, and behavior of relevant other individuals.

In this section, we discussed that individual actions are embedded in the social structure. Social control and other mechanisms explain why structural embeddedness has an influence on individual actions. Hence, the model of intergroup relations should consider effects of social networks and the underlying mechanisms. However, these mechanisms have not been related yet to intergroup relations. For this reason, *as a major contribution of this study, we will incorporate micro mechanisms of structural embeddedness into the team games model of intergroup relations*. The next section will specify the key elements of this model development.

1.6 Social control mechanisms and intergroup conflict

As we discussed in Section 1.4, structural embeddedness is missing from existing explanations of intergroup conflict. This is a major deficiency since individual behavior is embedded in the social structure also in the intergroup context (see Section 1.5). In this study, a more comprehensive theoretical model of competitive intergroup relations will be developed by an incorporation of different *social control mechanisms* into the team games approach.

We will concentrate on three different social control mechanisms. Social control in these forms constrains individual action by creating positive and negative incentives. First, the close social environment is a source of distribution of *selective incentives*, including social norms (Sandell and Stern, 1998). Durkheim (1984[1893]) and Parsons (1937), well before the rise of social dilemma research, have claimed that social norms help to solve the conflict between individual goals and common interest in favor of collective goals. The concept of selective incentives originates from Olson (1965), who showed the conditions under which the provision of selective incentives solves social dilemmas in a world of rational actors. Selective incentives from group fellows help the establishment of collective action as they reward participation and punish free riders within the group (Olson, 1965). The provision of these incentives is completely dependent on choices made in the intergroup context and does not require separate decisions. Therefore, it is more appropriate to consider them as additional incentives rather than produced normatives that raise a second order free rider problem (Heckathorn, 1989).

Second, behavior is constrained by social control as individuals strive for *behavioral confirmation* (Lindenberg, 1986). Behavioral confirmation is received for an action that is identical to behavior of related individuals. Finkel and Opp (1991) have found that participation in collective political action can be largely explained by willingness to conform to the expectations of important others. Furthermore, empirical evidence shows that people do not participate in collective actions in isolation, but together with friends and neighbors (see McAdam 1986; Gould 1991; Opp and Gern 1993). Individuals decide to participate, if they are assured of the participation of their friends (Chong 1991; Oberschall 1973; 1994). Chong (1991) and Oberschall (1994) described this as an assurance process. In the assurance process, behavioral confirmation has a two-fold effect. Confirmation by participating fellows provides an incentive for contribution and confirmation by free riders works against contribution.

Third, in intergroup relations, ties that connect members of the opposite groups pose a different constraint to individual behavior. Friends from competing groups have contradictory interests in the intergroup context. Since their friendship is valuable for them, they reward the other's action that is against the own group's interest (e.g., Kuran, 1995, 9-10). For instance, supporters surrounded by fans of the other club are esteemed for *traitor* behavior and for remaining silent, if their team scores.

As a result of dyadic social control, individuals can be mobilized to participate in collective action that has harmful consequences in the intergroup context. This raises further interesting questions about how the relative strength of social control mechanisms influences structural effects on intergroup conflict. We will explicate these research questions in Section 1.8. The new theoretical model that incorporates these forms of social control in the team games model of intergroup relations will be discussed in detail in Chapter 2.

1.7 Temporal embeddedness and intergroup conflict

Besides the structural embeddedness of behavior we should not forget the role of *temporal embeddedness* in intergroup relations (Granovetter, 1985). In this section, we will discuss the societal relevance of this factor. Moreover, we provide a brief overview of micro mechanisms specified by previous research that explain how temporal embeddedness influences intergroup relations. This study will rely on these mechanisms at the formulation of hypotheses for repeated intergroup encounters.

There is no doubt about the *societal relevance* of temporal embeddedness for empirical applications. Interdependencies arise repeatedly, giving a chance for the offended party to revenge. Examples are hostile relations between families in Corsica, Montenegro, Albania, or medieval Iceland, where insults are endlessly retaliated and the accepted norm of vengeance creates a spiral of clashes (Frank, 1988; Hardin, 1995). Similar dynamics of revenge can be traced between urban gangs, football supporters, nations, or ethnic groups. The Hutus and Tutsis have a long history of ethnic slaughter, in which killings are reciprocated by killings (Bhavnani and Backer, 2000). Fortunately, in many contexts, third party interventions or institutional solutions can break the spiral of violence. But the intention for revenge cannot be subdued easily. Hatred can elicit conflicts later in time, perhaps even a generation later, as in the case of Bosnia (Kaplan, 1993). Once conflict is established, it is difficult to bring it to an end. As Hardin notes (1995: 121) the really interesting question is not how the groups became enemies, but how their conflict is maintained. Counterexamples of competitive, but peaceful group relations are plentiful (cf. Fearon and Laitin, 1996; Gould, 1999). This leads us towards the inquiry of the determinants of the emergence of lethal and of peaceful scenarios.

In this inquiry, it is inevitable to discover *how* temporal embeddedness influences intergroup competition. A fundamental question of whether *the past or the future* governs individual decisions is a sharp division line in the theoretical literature and it is strongly related to debates over individual rationality.

Models that assume perfect rationality exclude the influence of the past. Individuals have a clearly forward-looking perspective, they choose for alternatives that offer the best future consequences. Such models are widespread in research on social dilemmas (e.g., Olson, 1965; 1982; Taylor, 1987; Sandler, 1992; Bicchieri, 1993; Buskens, 1999; Flache, 2002). Despite their theoretical value, these models have serious deficiencies in describing individual behavior in empirical and in experimental social dilemma situations (e.g., Dawes and Thaler, 1988; Caporael et al., 1989; Camerer, 1997). The problem is not merely that people have no knowledge or have misbeliefs about preferences and knowledge of others (Kreps et al., 1982). They also often fail to choose the alternative that would be optimal based on their beliefs and expectations (Rapoport, Bornstein, and Erev, 1989; Orbell and Dawes, 1991; Smelser, 1992). However, the fact that expectations often do not match with choices does not mean that expectations do not influence behavior at all. At the other extreme, a wide variety of models that assume backward-looking actors make this mistake, including several models of learning, imitation, reciprocity, and regret.

Comparison of extreme models of forward-looking and backward-looking behavior provided important insights for social dilemma research (Flache, 1996; Chen and Tang, 1996; Ho and Weigelt, 1996; Erev and Roth, 1998; Flache and Hegselmann, 1999b). However, in less extreme versions, forward-looking action and backward-looking behavior are not obviously in contradiction. In case of incomplete information forward-looking action can use experience from the past as indication of intended behavior. On the other hand, backward-looking action can also be purposeful and intentional in the sense that behavior is updated based on the past in order to gain future success (Vanberg, 2002).

There are some models that combine elements of forward-looking and backward-looking perspectives (Zeggelink, 1993; Stokman and Zeggelink, 1996; Buskens and Weesie, 2000), and there is also illustration for how strategic decision making may include rational learning from experience (Friedman, 1986: 126-136). The concepts of bounded and procedural rationality just deny the assumption of perfect rationality, but not the relevance of future (Simon, 1976; 1981). These concepts also dispute the logic of backward-looking models with long-term memory and a wide horizon. Instead, they assume that individual behavior is guided by a set of simple heuristics or rules of thumb (Orbell and Dawes, 1991) that might contain the influence of the past as well as the shadow of the future. The question is what might these simple heuristics be?

In a *forward-looking perspective*, assuming bounded rationality, there are several strategies that are based on expectations about subsequent outcomes and might be relevant in intergroup encounters. Shortsighted strategies vary in their assumptions regarding individual rationality, calculating capabilities, and access to information. Below we just mention some examples that will be used in the theoretical analysis of Chapter 2. Individuals might recognize their *dominant strategy*, if they have any. If they have sufficient information and calculating capabilities, they might give their *dominant reply* to dominant strategies of others. Even more extensive information is assumed, if there is *common knowledge* between friends or neighbors about the dominant reply of relevant others. Furthermore, in the lack of dominant reactions, individuals might decide on the basis of their *expectations* about the outcome of intergroup relations and about the decisions of related persons.

A simple forward-looking decision heuristic that is also based on expectations about the subsequent outcome is the *criticalness* principle. In case an individual expects that his or her decision would be a decisive one in the subsequent situation, which means that his or her decision would change the outcome of the intergroup competition, he or she is guided to participate in the collective action of the group. The median voter theorem of Downs (1957) was probably the first one to recognize the importance of objective criticalness. Later, research on step-level public goods showed that *perceived* criticalness is an important predictor of contribution (Rapoport, 1987; Kerr, 1989; Erev and Rapoport, 1990; Marwell and Oliver, 1993; Chen, Au, and Komorita, 1996; Au, Chen, and Komorita, 1998).

In a *backward-looking perspective*, strategies are conditional on previous events and decisions, that is, they are built on *past* experience. However, decision rules that build on the full record of past experience are not likely to be followed in practice. For instance, people rarely make as extensive calculations as it is assumed by fictitious play (Brown, 1951; Fudenberg and Kreps, 1993). It is plausible to assume that individuals act according to simple heuristics that use experiences about the outcome and about the behavior of a limited set of players from the recent past only. These heuristics can be interpreted as programs that trigger *conditional responses* to earlier events (Vanberg, 2002).

In this study, we will test the presence of some simple backward-looking heuristics in repeated intergroup encounters. One simple heuristic that received broad attention and empirical support in previous social dilemma research is *reinforcement learning*. Reinforcement learning describes that the chance of remaining at the previous individual decision increases after that particular action resulted in a satisfactory outcome. Otherwise this likelihood decreases. This mechanism has been found adequate in describing animal and human behavior in simple experimental situations. Its theoretical origin goes back to Thorndike's (1898) 'law of effect' and to Skinner's (1938) behaviorism. Bayesian updating and stochastic learning (Bush and Mostseller, 1955) are mathematical formulations of this simple rule. In slightly different versions, reinforcement learning has been useful to model and understand human behavior in social dilemmas (Nowak and Sigmund, 1993; Macy 1989; 1990; 1991b; 1993a; 1993b; 1995; Messick and Liebrand, 1995; Roth and Erev, 1995; Flache, 1996; Buskens and Snijders, 1997; Erev and Roth, 1998; 1999).

Another simple heuristic that has received enormous attention in the literature on social dilemmas is *reciprocity* (Axelrod, 1984; 1987; 1997a; Komorita, Hilty, and Parks, 1991; Komorita, Parks, and Hulbert, 1992; Kollock, 1993; Wu and Axelrod, 1995; Bicchieri, 1997; de Vos and Zeggelink, 1997; Wagner, 1998; Komorita and Parks, 1999; Watanabe and Yamagishi, 1999). There are theoretical arguments for reciprocity being a rational individual strategy (Friedman, 1971; Axelrod, 1984; 1997a). Reciprocity is also an important mechanism in intergroup relations (Fearon and Laitin, 1996; Goren and Bornstein, 1999). All empirical and anecdotal evidence of vengeance supports the consideration of this heuristic. Peaceful attitudes of the rival group are tolerated, but harmful collective action is retaliated in the subsequent encounter, if reciprocal strategies are applied.

Besides the aggregate action of the other group, the behavior of neighbors can also be retaliated (Whatley et al., 1999). Such *local reciprocity* is often named by journalists as the "main ingredient" of violent ethnic conflict (Hardin, 1995: 148). Local reciprocity is a direct way of imitation as it is a reaction to actual behavior in the past. The intergroup context provokes vengeance in interpersonal relations between members of the opponent groups, but there are also indications that intergroup relations support in-group reciprocity (Brewer, 1981).

In this section, we provided a brief introduction and overview of the forward-looking and the backward-looking behavioral mechanisms we will investigate in this study as a basis of dynamic intergroup relations. Our research questions about these micro mechanisms of temporal embeddedness and their macro consequences will be formulated in the next section.

1.8 Research questions

In the previous sections, we discussed the general research problem and the theoretical foundations of this study. In this section, we formulate our main research questions that can be classified along three dimensions. First, we are interested in the effects of *structural* and *temporal embeddedness* and their *interactions* on intergroup conflict. Second, we are interested in *macro effects* on intergroup conflict and in the *underlying micro mechanisms*. Third, research questions are open for a *theoretical* or for an *experimental* investigation. There are only effects of structural embeddedness in single intergroup encounters. In repeated intergroup relations, effects of structural and temporal embeddedness and also their interactions are present. Table 1.8.1 provides a simplified overview of this classification.

Table 1.8.1 Overview of research interests

<table>
<tr><td></td><td>Chapter 2
theoretical</td><td>Chapter 3
experimental</td><td>Chapter 4
theoretical</td><td>Chapter 5
experimental</td></tr>
<tr><td></td><td colspan="4">structural embeddedness</td></tr>
<tr><td></td><td colspan="2"></td><td colspan="2">temporal embeddedness</td></tr>
<tr><td rowspan="2">macro effects on
intergroup conflict</td><td colspan="4">segregation effect</td></tr>
<tr><td colspan="2"></td><td colspan="2">dynamics of intergroup
conflict and peace</td></tr>
<tr><td rowspan="2">micro effects on
individual decisions</td><td colspan="4">social control mechanisms operative</td></tr>
<tr><td colspan="2"></td><td colspan="2">behavioral heuristics</td></tr>
</table>

In order to gain a transparent overview, interaction effects of structural and temporal embeddedness are omitted from Table 1.8.1. This overview also expresses our objectives to have a close correspondence between the content of research questions that are formulated for theoretical and for experimental investigation. Keywords in the cells indicate our specific research interests. Main research questions for these interests are as follows.

Segregation effect on intergroup conflict:

- *What are the* ***structural conditions*** *under which the likelihood of intergroup conflict is higher and what are the conditions under which peaceful coexistence might be expected? Particularly, why and under what conditions does* ***segregation*** *promote intergroup conflicts?*

These research questions will be addressed in a theoretical investigation in Chapter 2. We test some of the derived model predictions under experimental conditions. Experimental analysis will address these questions in single intergroup competitions in Chapter 3 and in repeated encounters in Chapter 5.

Social control mechanisms operative:

- *What are the **underlying mechanisms** of network effects at the interpersonal level? How do different forms of **social control**, namely **selective incentives, behavioral confirmation,** and **traitor rewards** contribute to intergroup conflict and what is the impact of their relative size?*

These questions aim to highlight why and under what conditions individuals participate in collective action that hurts the interest of another group and might result in mutually harmful consequences. In particular, we will predict that different forms of social control influence individual decisions conditional on the local network neighborhood and consequently are responsible for the segregation effect on intergroup conflict. We will test these predictions in single-shot encounters in Chapter 3 and in repeated intergroup relations in Chapter 5. In the theoretical analysis of Chapter 2, we will also investigate whether the effects and macro consequences of social control mechanisms are dependent on assumptions about *individual rationality* and access to information or not.

Regarding effects of temporal embeddedness, we aim to answer the following research questions.

Dynamics of intergroup conflict and peace:

- *How do intergroup relations change over time? Are there **typical scenarios**, like an endless regression of conflict or a spiral of peace?*

Behavioral heuristics:

- *What are the simple **heuristics** that guide individual choice and as a consequence, are responsible for the emergence of macro scenarios? Are **criticalness, reinforcement,** and **reciprocity** important determinants of individual action in repeated intergroup relations?*

These questions will be investigated in repeated intergroup competitions. We will elaborate more on the exact nature of behavioral heuristics and their relation to aggregated scenarios in Chapter 4. Derived hypotheses will be tested in experiments in Chapter 5. Furthermore, we will also test interaction effects of structural and temporal embeddedness. At the macro level, we will investigate whether typical scenarios differ with structural conditions. At the individual level, we will analyze if we find different behavioral heuristics in different structural environments and we will test if experiences from the local neighborhood can trigger conditional responses.

To answer our research questions, we will apply simulations and experiments as research methodology. In the next section, we will give arguments why do we choose for these methodological tools.

1.9 Research methodology: simulations and experiments

Due to the complexity of real settings, behavioral rules and social network effects are often untraceable and our predictions about them cannot be tested. We would need to analyze very simple (for instance, ancestral or tribal) environments in order to investigate individual level mechanisms that are responsible for social network effects and dynamic intergroup processes.

A powerful alternative to prehistoric or anthropological testing is computer simulation and conducting laboratory experiments. In *simulations*, computational capacities allow to investigate the effect of parameter values and explanatory variables in any deliberately chosen environment. Using a computer the properties of a higher level system can be derived from simple laws governing the lower level units (Epstein and Axtell, 1996). Simulations are also experiments in the sense that different combinations of parameter values can be tried out to investigate the theoretical research questions. Simulations have the aim to explore general answers, but only in a simplified environment that is not a true representation of reality (Gilbert, 1995). They are favored to analytical methods when the formal derivation of system properties is difficult from the model assumptions.

A criticized feature of simulations is that they require simplified assumptions and cannot cope with the complexity of real situations. A sound reply to this criticism is expressed by Axelrod (1997a: 5): "… if the goal is to deepen our understanding of some fundamental process, … then simplicity of the assumptions is important and realistic representation of all the details of a particular setting is not." The lack of complexity is a virtue of simulations, not a vice. Based on simple assumptions they are perfect tools of theoretical investigations and abstract model building. As Macy (2002) summarizes, "Making these models fit actuality would add complexity that undermines their usefulness as theoretical tools."

In our case, we intend to use simulations for exploration of underlying mechanisms that cause the macro phenomenon of intergroup conflict. In this respect, our study is one of the first attempts to use simulation methodology to understand intergroup conflicts.[3] Furthermore, we will use the simulations to derive hypotheses for our experiments.

The experimental design is constructed in order to test these hypotheses. Laboratory *experiments* have the advantage that by manipulating conditions and controlling for disturbing variables it is possible to detect the presence of predicted mechanisms. For this reason and because the values of other parameters can be fixed, experiments are ideally suited to test hypotheses about mechanisms of the simulation model. In laboratory conditions, interdependencies can be created, in which individual

[3] Existing simulation studies deal with different kinds of intergroup relations than we do (cf. Suleiman and Fischer, 1996; 2000; Fischer and Suleiman, 1997; Rapoport and Amaldoss, 1999; Jager, Popping, and van de Sande, 2001).

payoffs are measurable. Non-strategic uncertainty can be kept to a minimum and control of information can be easily achieved (Crawford, 1997: 216). Non-monetary incentives can be separated by experimental manipulations (Roth, 1995). Furthermore, laboratory experiments do not face serious ethical concerns that field experiments of intergroup conflict do.

Experiments also have their drawbacks. It is difficult to generalize their results, because subjects are most often students, therefore the pool of participants does not represent the entire population. This would be a serious concern, if there had been a major difference in the effects of structural and temporal embeddedness for students and for the general population. Since the number of experimental sessions and participants is vastly limited, only effects of few variables can be tested and only a small set of parameter values can be used. However, this is only a disadvantage compared to simulations. In field experiments or in case studies there is no way to vary parameter values at all.

Furthermore, there is a rather large gap between the simulations and the experiments with regard to their behavioral assumptions. This follows from the difference in the objectives of these research methods. Simulations are devoted for understanding the theoretical relationship between structural embeddedness and intergroup conflict. This relationship is explored in general when auxiliary assumptions on individual rationality and on access to information are introduced and varied. Interactions between these assumptions and effects of structural embeddedness are certainly interesting, but questions about individual behavior are not as central as in experiments. In the experiments the actual behavior of subjects is studied and hypotheses about decision heuristics that appear in repeated encounters are tested. Moreover, there are also differences in how the results are analyzed. In the simulations, simple rules are assumed about individual behavior and macro phenomena are derived through the application of these rules. In the experiments, behavioral assumptions as well as emergent macro phenomena are tested using statistical analysis.

1.10 Outline of this study

In *this chapter*, we provided an introduction to the analysis of intergroup conflict and peace. We presented an overview of the societal relevance and theoretical background of this problem. We discussed the necessary theoretical developments and the major contributions of the present study. We formulated our research questions and we provided arguments for the chosen simulation and experimental methodology.

The structure of subsequent chapters follows the classification of research interests as they were provided in Table 1.8.1. *Chapters 2* and *3* are devoted to the analysis of effects of structural embeddedness in single intergroup encounters. The emphasis is on the explanation of why and under what conditions intergroup conflicts are promoted by segregation and on the underlying mechanisms of this macro phenomenon. In Chapter 2, the Intergroup Public Good (IPG) game model of competitive intergroup relations is extended by dyadic forms of social control that can be responsible for social network effects. Next to the theoretical contributions, this chapter aims to answer theoretical research questions by using simulation methodology. In Chapter 3, a new experimental design is introduced to test the segregation effect and the underlying social control mechanisms in a series of single-shot IPG games.

In *Chapters 4* and *5*, we analyze repeated intergroup encounters. Besides the effects of structural embeddedness, this allows for the investigation of how and as a result of which mechanisms intergroup relations change over time. In Chapter 4, we specify elementary behavioral rules that might be relevant for individual decisions in the repeated IPG game and we derive hypotheses about micro and macro effects of temporal embeddedness. Model predictions about effects of structural and temporal embeddedness, and about their interactions are tested in Chapter 5.

In *Chapter 6*, research problems and results of this study are summarized and conclusions are drawn. Furthermore, limitations of the study are discussed and suggestions are given for future research. It is argued that policy implications should be handled with reservations as there is no general remedy for real intergroup conflicts and resolutions have to take into consideration the unique circumstances of the given situation.

"…it is social structure that can transform free ridership into zeal."
James S. Coleman: Foundations of Social Theory (1990: 275)

CHAPTER 2

SEGREGATION AND INTERGROUP CONFLICT

Theoretical developments and analysis of single-shot situations[1]

[1] Parts of this chapter are drawn from Takács (2001).

2.1 Introduction

In this chapter we present a theoretical model that provides predictions about the structural conditions under which harmful intergroup conflict is likely to occur and when peaceful coexistence might be expected. We focus on a factor that has been surprisingly neglected in previous research: the effect of structural embeddedness (cf. Granovetter, 1985). Previous models have not taken into account that individual network ties within *and* between the groups transmit social and cognitive rewards that influence participation in intergroup related collective action (see Section 1.5). In particular, while it is widely believed that dense in-group relations help the establishment of collective action (Marwell, Oliver, and Prahl, 1988; Marwell and Oliver, 1993; Gould, 1993a; 1993b), in the intergroup context, we do not know much about why and under what conditions dense in-group and scarce out-group relations (segregation, clustering) support harmful collective action. We will investigate this question in this chapter, considering competition situations between two distinct groups. As a first step, we focus on single-shot situations. Since history effects are quite crucial in empirical examples, Chapter 4 will be devoted for the development of a dynamic model that considers repeated intergroup relations.

In Section 1.4 we discussed that if individual contributions are costly, group beneficial collective action is difficult to achieve. This partly explains why lethal conflicts are more the exceptions than the rule (Fearon and Laitin, 1996; Gould, 1999). On the other hand, experimental findings have revealed that a presence of an out-group, especially in the case of competitive intergroup relations might help to establish collective action within the groups. This duality of intergroup and within group interdependencies and their interrelation is represented in the *Intergroup Public Goods* (IPG) game (Rapoport and Bornstein, 1987; Bornstein, 1992).

The original IPG model disregards the *structural embeddedness* of individual actions. As it was discussed in Section 1.5, everyone's behavior is to a large extent constrained and influenced by neighbors, friends, and the family, regardless of their group membership. In this chapter, we would like to demonstrate the underlying mechanisms of *social control* behind these network constraints. In the close social environment *selective incentives* are distributed (Sandell and Stern, 1998). Selective incentives from fellow group members help the establishment of collective action, but friends from the other group provide *traitor incentives* that suppress contribution. *Behavioral confirmation* that accompanies almost every kind of human interaction (Lindenberg, 1986) is also transmitted by network ties. In large-scale collective action people are not likely to conform to everyone's behavior or to be fair to everyone. Behavioral confirmation is distributed in the social network environment, which appraises expectations about the behavior of friends and neighbors.

These social control mechanisms can be represented as overlapping dyadic interdependencies. Each dyadic relation is subject to playing local coordination games (cf. Ellison, 1993; Morris, 2000), but in a form that is inseparable from participation choice in collective action. Surrounded by extremist fellow group members, people are highly constrained to participate. On the other hand, peaceful friends or many friends from the rival group can provide enough confirmation pressure to avoid contribution to the harmful collective action. Such mobilization process is also referred to as “block recruitment” (Oberschall, 1973) and can provide the micro foundation for collective actions such as demonstrations, urban gang fights or civil war.

In this chapter, we incorporate these different mechanisms in the model of intergroup conflict. Intergroup competition, structural embeddedness, and strong social control in dyadic connections might mobilize group members to participate in intergroup related collective action. A theoretical investigation in this chapter supported by computer simulations aims to show under what structural conditions and due to which forms of social control mobilization occurs.

As a consequence of these factors, a social trap of a different kind might arise. Should both groups become involved in a collective action, then this could be mutually harmful. From the community perspective, this is an outcome that is best avoided. Among structural conditions, it is particularly interesting to study under which circumstances *segregation* leads to intergroup clashes. In this chapter, we will investigate which forms of social control are responsible for the strengthening of the segregation effect. We will demonstrate how the relative size of selective incentives and behavioral confirmation (normative pressure versus confirmation pressure) influences the effect of segregation on the likelihood of conflict and hence the chance that residential policy can help conflict resolution.

Furthermore, for exact model predictions, we have to specify assumptions about individual consciousness and access to information. Assumptions in the relevant literature contain perfect rationality (e.g., Chwe, 1999) as well as considerations of motivations that are beyond egoistic incentives (cf. Caporael et al., 1989). There are various models that assume bounded rationality or limited access to information in different forms (e.g., Ellison, 1993; Morris, 2000; Fearon and Laitin, 1996; Gould, 1993a; Rapoport and Bornstein, 1987; Macy, 1991b). Since there is no generally accepted view on what level of rationality can be assigned to individual action and on what kind of information people use for their decisions, we will consider four models with different levels of consciousness and access to information. We do not claim that any of these models reflect the appropriate view on the logic of individual action. We also think that there is no such a view: the level of rational consciousness in individual action varies, depending on the framing and importance of situations (Lindenberg and Frey, 1993). Individual access to relevant information also alters in different circumstances.

The effect of network structure on intergroup conflict will be compared under the different specifications. Different behavioral models can also be considered as robustness tests to check whether normative pressure in general strengthens the segregation effect on intergroup conflict more than confirmation pressure. Examples of typical structures will demonstrate how the success of local mobilization can be dependent on the behavioral assumptions on individual “rationality”. We will investigate whether under certain structural conditions “rational” individuals are more willing to contribute to collective action or are they the ones who always abstain from participation.

In the following section, we introduce the Intergroup Public Goods (IPG) game as a model of competitive intergroup relations. As a major theoretical contribution of this chapter, we incorporate dyadic effects of social control in the model in Section 2.3. This new, structurally embedded model provides an explanation as to why segregation often supports the emergence of intergroup conflict. In Section 2.4 we show that participation in conflict can even be a dominant individual strategy that leads to a suboptimal outcome of harmful clashes. In order to derive exact predictions, we introduce four different behavioral models in Section 2.5. These models differ in their assumptions about how much rationality is assigned to individual actions.

The conditions under which segregation leads to intergroup conflict are investigated in the four behavioral models by using computer simulation. Section 2.6 describes the simulation design. Section 2.7 clarifies the interpretation of segregation and its measurement in the simulations. Simulation results are presented in Section 2.8. Examples of other structural effects are discussed in Section 2.9. Finally, we summarize the simulation results and conclude this chapter in Section 2.10.

2.2 The Intergroup Public Goods game

There are two exclusive groups A and B of size n_A and n_B ($n_A \geq 2$ and $n_B \geq 2$), with contradictory collective interests. Inside both groups, members face the dilemma of providing a step-level public good (cf. Bornstein, 1992). A step-level public good is only provided, if a certain level of contribution has been reached. This level is determined by the number of contributors in the other group.

We assume that group members are anonymous and can gain (lose) the same rewards from the intergroup context with identical action. For instance, we suppose that everyone is equally proud after a victory and equally ashamed after a defeat.[2] If the number of contributors in group A exceeds the number of contributors in group B, then each member in A receives a reward v (a piece of a *victory*-cake, temptation

[2] More precisely, since we do not make any interindividual comparison, it is enough to assume that everyone values his or her share relatively the same way in comparison to other rewards and costs of him or her.

reward) and members of B receive a reward of d (*defeat*, the sucker's payoff). If the number of contributors is equal, then everyone receives a reward c (*clash*, reward for a draw). The relation between these rewards is $v>c>d$. For instance, let us consider a team sport in which members of the two teams decide whether they help their team with low or high effort. High effort can be regarded as contribution to the provision of the public good that is a win in the competition.

We assume that a free riding action (low effort in the previous example) results in an extra positive reward of e (*endowment*, $v>e>d$). As an example, consider that the group of Republican voters are all happy if a Republican president is elected, but those who refrain from voting gain more, they could otherwise occupy themselves instead of going to the polls (opportunity cost of voting). Table 2.2.1 represents the payoffs from the IPG game for player $i \in A$.

Table 2.2.1 Possible payoffs from the IPG game (Rapoport and Bornstein, 1987)

	(I)	(II)	(III)	(IV)
Outcome	*Unconditional defeat*	*One for the draw*	*One for the victory*	*Unconditional victory*
Conditions	$k_{A-i}<k_B-1$	$k_{A-i}=k_B-1$	$k_{A-i}=k_B$	$k_{A-i}>k_B$
Contribution	d	c	v	v
No contribution	$d+e$	$d+e$	$c+e$	$v+e$

Notes: k_{A-i}= the number of contributors in group A (excluding player i); k_B= the number of contributors in group B.

As Table 2.2.1 shows the IPG game is not a perfect social dilemma. In a game of continuous provision of public goods not to contribute is a dominant strategy, but this is not the case for a step-level public good (cf. Frohlich and Oppenheimer, 1978; Hardin, 1982). In the IPG game shown in Table 2.2.1, when $c>d+e$, in state (II) i is better off by contributing and when $v>c+e$, in state (III) contribution is a better choice. Given the lack of a dominant strategy Rapoport and Bornstein (1987) assumes that individuals order expected values to the actions and maximize this expected value. Critical probabilities (cf. Caporael et al., 1989) concerning states (II) and (III) show the likelihood that the single individual action of i has an influence on the overall outcome. Denote the subjective probabilities for i of the four states with P_I, P_{II}, P_{III} and P_{IV}, respectively ($P_I+P_{II}+P_{III}+P_{IV}=1$). It can be shown, that contribution is a better choice for player i if

$$P_{II}\,c + P_{III}\,(v - c) > e. \qquad (2.2.1)$$

If group sizes are large, then critical probabilities are small and the structure of the game is close to being a pure in-group social dilemma (cf. Bornstein and Rapoport, 1988, 127). Consequently, peace is very likely to be the outcome of intergroup opposition (cf. Fearon and Laitin, 1996; Gould, 1999).

However, the social trap character of a stalemate is not incorporated in the original IPG model. Therefore, that model can be applied to different intergroup competition situations in which a tied outcome does not have a harmful character. An example is an election in two-party democracies. A tie can be embarrassing and can lead to lengthy and costly recounting and law suites, but it is not worse than if nobody voted. Similarly, in team sports a scoreless draw at a boring match is definitely not better than a draw with many goals, where all players gave their best. Competitions between ethnic groups, urban gangs, or pupil groups have a different character. A draw means a mutually harmful clash that is worse than the lack of collective action (peace). To explain the emergence of such outcomes, we implement modifications that are borrowed from the IPD game (see Section 1.4).

Another change is a generalization of the original model. In the Rapoport and Bornstein (1987) model, for all draws the size of public payoffs are equivalent, irrespective of how many individuals contribute. Since in the situations we would like to model, mutual mobilization significantly differs from mutual lack of mobilization, we should specify how the payoffs depend on the number of contributors, if contributors in the two groups are equal in number. If a couple of Serb civilians shoot at Bosnian civilians, this would not yet be a civil war situation, but would be treated as an attempt of murder. Ethnic clash starts, if the number of contributors is large enough. For this reason, besides the endogenous threshold (the number of contributors in the other group), we introduce an exogenous minimal contributing set (MCS) in the game (cf. van de Kragt, Orbell, and Dawes, 1983). This means that less than a specified number of contributors no collective action will be established in the group and consequently provision of the victory and clash public goods is not possible. Since groups can differ in their internal structure of norms, one group can exhibit a more patient attitude in intergroup relations. Therefore we allow minimal contributing sets to be different in the groups and we denote them by k_A^* ($0 \leq k_A^* \leq n_A$) in group *A* and by k_B^* ($0 \leq k_B^* \leq n_B$) in group *B*.

If the number of contributors is equal and both groups are above the minimal contributing set, then everyone receives a negative reward *c* (*clash*, punishment payoff). We suppose that the clash of collective actions is worse than the outcome of *peace*. Peace is the collective outcome in which no collective action is established in the groups. For the sake of simplicity the reward for peace *p* is a reference value and assumed to be zero. Hence, the relation between the different payoffs is $v>p=0>c>d$. If groups were unitary entities and they could choose between either collective action or no action, collective action would be their dominant strategy. Both sides following the dominant strategy would lead to a suboptimal outcome. On the basis of Definition 1.2.1.1, we will call all outcomes in this game *intergroup conflict*, where collective action is established at least in one of the groups.

We retain the assumption that a free riding action results in an extra individual reward of *e* (*endowment*, $v>e>0$). Table 2.2.2 gives a complete typology of possible outcomes of the IPG game and represents the payoffs for player $i \in A$.

The IPG game in this form is intended to model group competition situations in which collective action of equal strength leads to mutually harmful outcomes (clash punishment). Examples are civil war, conflicts between pupil groups, fights between football supporters or urban gangs. In the case of only a few initiators, nothing happens and the status quo is preserved. If, however, the number of contributors exceeds a certain threshold, a collective action is established and this can mean victory for the group. A group wins if more contributed to the collective action than in the other group. Defeat is the worst case scenario: just imagine the frustration experienced by gang members having lost a street battle.

Table 2.2.2 Possible payoffs from the IPG game with clash punishment[3]

	(I) *Peace*	(II) *Unconditional defeat*	(III) *One for the clash*	(IV) *One for minimal contribution*	(V) *From defeat to victory*	(VI) *One for the victory*	(VII) *Unconditional victory*
	$k_{A-i}<k_A^*-1$ and $k_B<k_B^*$	($k_{A-i}<k_B-1$ or $k_{A-i}<k_A^*-1$) and $k_B \geq k_B^*$	$k_{A-i}=k_B-1$ and $k_B \geq k_B^*$ and $k_{A-i}+1 \geq k_A^*$	$k_{A-i}=k_A^*-1$ and $k_B<k_B^*$	$k_{A-i}=k_A^*-1 \geq \geq k_B \geq k_B^*$	$k_A^* \leq k_{A-i}= =k_B \geq k_B^*$	($k_{A-i}>k_B$ or $k_B<k_B^*$) and $k_{A-i} \geq k_A^*$
C	*0*	*d*	*c*	*v*	*v*	*v*	*v*
D	*e*	*d+e*	*d+e*	*e*	*d+e*	*c+e*	*v+e*

Notes: k_{A-i}= the number of contributors in group *A* (excluding player *i*); k_B= the number of contributors in group *B*.

We need assumptions about how individuals reach their decisions so as to predict which outcome will be realized. Similar to the original IPG game, there is no dominant strategy in the modified game. In states III, IV, V, and VI (cf. Table 2.2.2) contribution is a better choice. In these states a single individual decision is *critical* for the outcome. Using the argument of Rapoport and Bornstein (1987) again, given the lack of a dominant strategy we could assume that individuals order expected values for actions and maximize this expected value.

Denote the subjective probabilities for *i* of the seven states in correspondence with states (I)-(VII) in Table 2.2.2 by P_z's ($\sum_{z=\text{I..VII}} P_z = 1$). For the sake of simplicity, let us assume that rewards are numerical and individual utility is a linear function of rewards. The expected value (*EV*) of contribution for player *i* is:

[3] State (V) is the exceptional case and only relevant if $k_A^*>k_B^*$. In this state there are an equal number of contributors in both groups. Collective action is established only in group *B*, but the contribution of player $i \in A$ means that the outcome is victory of group *A*. State (I) is not relevant, where less than two contributors can establish group collective action and state (IV) is not relevant, where the minimal contributing sets are zero. States (VI) and (VII) are irrelevant, if $k_A^*=n_A$.

$$EV(C) = P_{II}d+P_{III}c+(P_{IV}+ P_V+ P_{VI}+ P_{VII})v,$$

whereas the expected value of not contributing is

$$EV(D) = e+(P_{II}+P_{III}+P_V)d+ P_{VI}c+ P_{VII}v.$$

Contribution is a better choice if $EV(C) > EV(D)$, thus

$$P_{III}(c-d)+P_{IV}v+ P_V(v-d)+ P_{VI}(v-c) > e,$$

where P_{IV} is zero, if $k_A^*=0$, P_V is zero, if $k_B^* \geq k_A^*$, and P_{VI} is zero, if $k_A^*=n_A$.

In this model, critical probabilities concerning states III, IV, V and VI indicate the likelihood that the decision of *i* has an influence on the overall outcome. If group sizes are large and the minimal contributing sets are relatively high, then the likelihood is extremely small and the structure of the game is close to being a perfect social dilemma. Similar to the original IPG game, peace is still the expected outcome of intergroup opposition (cf. Fearon and Laitin, 1996; Gould, 1999).

A specific example of the game is represented graphically in Figure 2.2.1. In this example group sizes and the minimal contributing sets are equal. Bullets indicate Nash equilibria. In general (if minimal contributing sets are larger than one), pure strategy Nash equilibria are the situations in which there are {0; 0}, {k_A^*; 0}, or {0; k_B^*} contributors. Overall defection is an equilibrium, because a single contribution cannot break the peace, but involves the loss of endowment *e*. Another Nash equilibrium is when the number of contributors in one group equals to the MCS and in the other group there are no contributors. In this case, no contributor would be better off by free riding, because $v>e$. For defectors, it would not make sense to change their decisions, because they cannot improve on the outcome alone. In addition to these equilibria, clash with overall participation is also a Nash equilibrium, if group sizes are equal and $d+e<c$. This equilibrium is never Pareto-optimal.

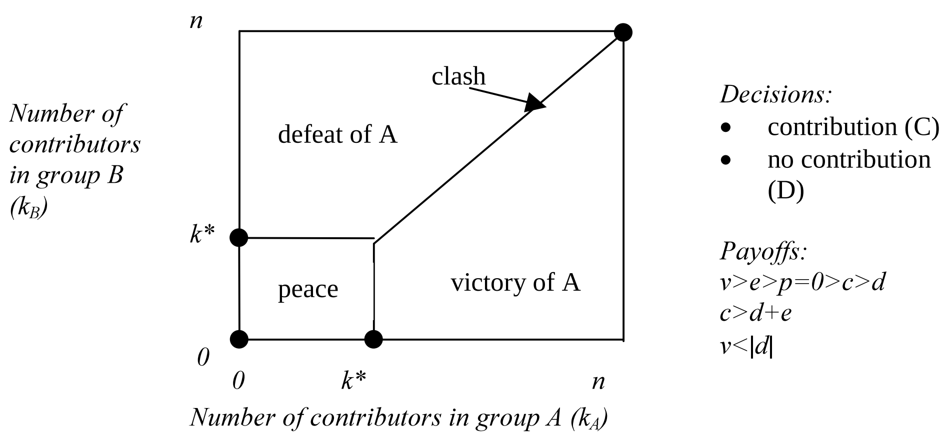


Figure 2.2.1 Graphical representation of an example and Nash equilibria (bullets)

Although concerned about the simplifications made, we will nevertheless use this simple model throughout this study as the solid ground of modeling intergroup conflict. We will leave issues such the size of payoff parameters, additional incentives, non-linear utilities, risk preferences, social orientations, third party or institutional interventions for later discussion. These are all factors that can introduce complications to the model. Discussion of these will be however strictly limited, as they are not in the main interest of this research.

2.3 Interdependencies between neighbors and friends: the structurally embedded IPG game

As individuals are able to free ride on the effort of others, beneficial collective action is difficult to achieve within the groups. However, individuals can be mobilized for participation, if effective social control, norms, or selective incentives exist in the group (Olson, 1965; Coleman, 1990; Heckathorn, 1990). In this section we will incorporate three different forms of social control that are transmitted by network ties in the model of intergroup conflict. These are *traitor rewards, behavioral confirmation*, and *social selective incentives*. As a consequence of dyadic social control, under certain network structures harmful collective action is likely to emerge.

If their neighbors or friends are from the other group, individuals are rewarded for not participating in the collective action. We assume that everyone receives a $t>0$ *traitor* reward in the case of no contribution for each tie that connects this person to members of the opposite group. Hence the traitor payoff is a selective incentive rewarding defection and distributed locally conditional on the number of ties with the other group. The traitor reward provides an additional incentive for people who, living close to members of the opposite group to restrain from participation in collective action. For instance, supporters surrounded by fans of the other club are rewarded for remaining silent in a stadium.

Ties connecting members of the same group transmit different social incentives. People receive *behavioral confirmation* ($b>0$) from each relation by acting similarly to their fellow friends. This reward is a mutual positive externality, which drives towards uniform action. Irrespective of behavioral confirmation, contribution is rewarded by fellow neighbors or friends. They appreciate group-beneficial action by social *selective incentives.* We assume that all contributors receive a selective incentive $s>0$ from each fellow neighbor. The provision of these incentives does not require separate decisions as they always accompany choices made in the intergroup game. This assumption is plausible for certain social rewards such as respect or status that can be by-products of intergroup relations. The relationship between fellow neighbors or friends can be represented as a local coordination game (see Table

2.3.1).[4] Unlike public goods ("bads") *v*, *c* and *d*, bestowing the three types of social incentives is not conditional on the outcome of the intergroup competition.

Table 2.3.1 Local coordination game between fellow neighbors

		neighbor from the same group	
		contribute	do not contribute
individual i	contribute	$b+s$, $b+s$	s, 0
	do not contribute	0, s	b, b

The structurally embedded IPG game is the *extension of the IPG game* (Table 2.2.2 and Figure 2.2.1) *with the incentives* (*t*, *b*, and *s*) *from the network environment*. Individuals must choose a single action (contribution or no contribution) and cannot tailor their behavior to each neighbor.[5] A formal expression of individual payoffs in the structurally embedded IPG game can be found in the Appendix 2.A.1.

2.4 Social dilemma of a different kind

In the structurally embedded IPG game, contribution can even be the dominant strategy. For this, selective incentives have to exceed rewards for defection in the "worst case" scenario, i.e. when no fellows neighbors are contributing and a single contribution does not change the outcome. That is, contribution is a dominant strategy of player $i \in$ A, if

$$g_i t + e < f_i (s - b), \tag{2.4.1}$$

where f_i denotes the number of fellow neighbors of i and g_i stands for the number of neighbors from group *B*. Although the decision of *i* is not likely to be critical, contribution can be highly beneficial due to social incentives. For instance, many individuals join tribal wars although the gains from these conflicts are only symbolic and single contributions make no difference. One reason is that warriors can attain high status in the group and can easily become "heroes". We can argue in a similar way in order to explain redundant contribution choice (Caporael et al., 1989). People seek social rewards when they sacrifice their contribution to the production of a public good that has already been established.

[4] For the sake of simplicity, we assume that selective inventives and confirmation payoffs are held constant through all pairwise games. Although it is sufficient to assume that all individuals relate social rewards to other rewards and costs in the same way.
[5] It is also the assumption of the literature on local interaction games (see Morris, 2000: 57).

At the other extreme, no contribution is the dominant strategy of *i*, if defection provides higher rewards than contributing even if all fellow neighbors are contributing and a single additional contribution would change the outcome of the game. That is, defection is a dominant strategy of player $i \in A$, if [6]

$$g_i t + e > f_i(s+b) + v - d. \tag{2.4.2}$$

Nash equilibria in the structurally embedded game can be very different from the original IPG game, depending on the exact network structure. Social networks decisively shape conditions under which social incentives can generate solutions for the in-group collective action problem. In a highly segregated network with dense in-group and scarce out-group relations overall participation is likely to be an equilibrium. Full contribution can be a dominant strategy equilibrium and a suboptimal outcome in which every individual payoff is smaller than in overall peace. The unusual social dilemma that traps groups in harmful contribution emerges, if

$$|c| > f_i s - g_i t - e > f_i b \tag{2.4.3}$$

holds for every individual.

2.5 Model predictions under different decision algorithms

In order to derive exact model predictions, we need auxiliary assumptions on individual behavior. In this section we introduce four behavioral models and describe the effect of segregation on intergroup conflict under the different assumptions. The discussion of these four models has an illustrative purpose and does not mean a commitment to related behavioral disciplines. In all four models, however, we assume a certain level of rationality. We also bring strategic thinking back into consideration, which would not be present if we had assumed that people order expected values for actions and maximize this value (cf. Rapoport and Bornstein, 1987). On the other hand, the models also deviate from classical game theory by avoiding the assumptions of complete information and perfect rationality.

---

[6] If $k_B^* \geq k_A^*$, then the less strict conditions

$$g_i t + e > f_i(s+b) + c - d$$

and

$$g_i t + e > f_i(s+b) + v - c$$

are sufficient to hold for defection (no contribution) to be a dominant strategy.

By considering four decision models, we can analyze the effects of individual consciousness and access to information on the likelihood of contributing and the interaction effect of behavioral assumptions and segregation on intergroup conflict. In Model 1, we only assume that individuals choose their dominant strategy, if they have any. This is a simple strategic model, in which individuals take those actions that are beneficial under all circumstances. In Model 2, in addition we assume that individuals are also able to recognize the dominant replies to the dominant strategies of their neighbors. This is still a simple strategic model that introduces local information into the analysis. In Model 3, we assume that such obvious actions are common knowledge between neighbors and optimal replies are chosen accordingly. In Model 4, an expected value element is added to all these assumptions. This model still assumes bounded rationality, since we control for a tendency of overestimation of criticalness (cf. perceived efficacy at Kerr, 1989). These four models are easily comparable as their assumptions are added step by step. The most rigid assumptions on individual behavior are used in Model 4 and the least rigid ones in Model 1. All the assumptions will be discussed in detail in this section. We intend to show that rigid assumptions strengthen the predicted relationship between segregation and the likelihood of intergroup conflict.

2.5.1 Model 1: Dominant strategy rule

In the first model, only a limited rationality of players is assumed. We presume that actors choose their dominant strategy, if they have one. Everyone lacking a dominant strategy contributes with a fixed probability. These behavioral assumptions allow for a derivation of hypotheses from equations (2.4.1) and (2.4.2). As far as the main payoff parameters are concerned, a smaller difference between victory and defeat and a smaller reward for free riding will increase the likelihood of intergroup clash. With regard to structural effects, extensive fellow connections (larger f_i's) support contribution, therefore collective action will be more likely in a clustered population.

Already this simple model generates empirically plausible implications. However, there are empirical examples that contradict our predictions and show that isolation can sometimes be an effective way to avoid intergroup clash. Isolation in these cases could mean a termination of the interdependent situation (e.g., building a wall in Belfast, destruction of a bridge in Mostar, or blocking a bridge in Kosovska Mitrovica). These are external or artificial solutions of intergroup conflict that might require the deployment of armed forces for monitoring. When there are no external solutions and interdependencies are unavoidable, *we predict a strong effect of segregation on the likelihood of conflict between groups, where selective incentives are relatively more important than behavioral confirmation* (see equation 2.4.1).

The model has interesting implications for group size effects. In a physical clash or battle, larger groups can obtain success easily. If the minimal contributing sets are equal in the groups, the larger group has a higher chance to win from the intergroup opposition. The lower the minimal contributing set, the more likely it is that collective action emerges. In this case, only a little noise is needed to destroy the peaceful equilibrium. Empirical examples of noise are mistakes, misinterpretations, drunkenness, or sudden passions (Fearon and Laitin, 1996). It is more remarkable that even if minimal contributing sets are given proportionally to group size, the larger group still has the advantage. It follows from the fact that if group *A* is larger than group *B*, then the expected proportion of fellow ties in group *A* is higher than the relative size of group *A*. Hence, the chance of being in a neighborhood, in which normative pressure restricts the individual to contributing action, is exponentially higher by increasing group proportion. For instance, there is evidence of nonlinearly increasing voting participation (and votes) with higher levels of residential segregation (Butler and Stokes, 1974). In other cases this prediction may contradict real life experience. Larger groups tend to be more sparse and less organized (Olson, 1965). Furthermore, if there is a large inequality in the strength of the groups, the minority may try to avoid intergroup opposition by choosing assimilation. As a consequence, group borders might fade away.

2.5.2 Model 2: Dominant reply rule

In the second model we formulate more rigid assumptions about individual behavior by introducing access to local information. Every actor follows his or her dominant strategy, if there is one. Furthermore, since people know their neighbors to some extent, they can also attain information about their possible actions. Let us assume that people can recognize when their neighbors have a dominant strategy and can give an unconditional best (dominant) reply, if there is one.[7] Denote the number of fellow neighbors who have a dominant strategy of contribution by f_{ic} and the number of fellow neighbors who have a dominant strategy of defection by f_{id}. From Table 2.2.2 and equations (2.4.1) and (2.4.2) it can be derived that contribution is the unconditional best (dominant) reply of $i \in A$, if

$$g_i t + e < f_{ic}(s+b) + (f_i - f_{ic})(s-b) \qquad (2.5.2.1)$$

holds and defection (no contribution) is the dominant reply, if

$$g_i t + e > (f_i - f_{id})(s+b) + f_{id}(s-b) + v - d \qquad (2.5.2.2)$$

[7] We call a strategy unconditional best (dominant) reply, if it is a pure best reply against all pure strategy profiles that contain the dominant strategies of neighbors. It involves the assumption that every actor is capable of assessing information about the number of ties and about the relative value of social rewards *b* and *s* for all fellow neighbors.

is satisfied.[8] Everyone without a dominant strategy or dominant reply is assumed to contribute with a fixed probability.

Model 2 generates further insights into structural effects. Compared to Model 1, the existence of relatively closed “ghettos” increases the likelihood of conflict. The periphery of these network segments acts in line or together with the initiators as they have dominant reactions. The higher the relative size of social incentives ($s+b$), the more likely it is that the periphery will also be encouraged to contribute. A large relative difference between selective incentives and behavioral confirmation ($s-b$) helps key contributors to arise (see Model 1), but their additive value ($s+b$) is important for the mobilization of the periphery (Model 2). On the other hand, if confirmation rewards are relatively important, peaceful behavior might diffuse to radical defectors. Hence in Model 2, centralized networks are efficient in spreading both behavioral patterns (cf. Marwell, Oliver, and Prahl, 1988; Gould, 1993a).

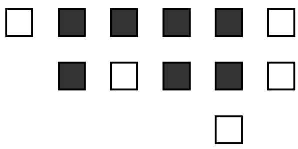


Figure 2.5.2.1 Imaginary map of a small village with mixed population

Consider for instance the following example. An imaginary map of a small village is represented in Figure 2.5.2.1. Five members of group *A* (white houses) and seven members of group *B* (black houses) inhabit the village. In this example members of group *A* live at the periphery of the village. Assume that groups are involved in a competition situation that can be described by the structurally embedded IPG game. Let us suppose that everyone is in close connection only with neighbors to the East, West, South, and North. Group *B* has the advantage of size but also has a structural advantage, since its members are located mainly in the center of the village. Consequently, collective action is more likely in this group. One player has a dominant strategy of contribution in group *B*, if $e+3b<3s$. Three other members living in black houses have a dominant reply of contribution in this case, if $e+t<2s$. If a stronger condition of $e+t+2b<2s$ is satisfied, then five members of group *B* have

[8] If $k_B^* \geq k_A^*$, then the less strict conditions

$$g_i t + e > (f_i - f_{id})(s+b) + f_{id}(s-b) + c - d$$

and

$$g_i t + e > (f_i - f_{id})(s+b) + f_{id}(s-b) + v - c$$

are sufficient to hold for defection to be a dominant strategy.

contribution as a dominant strategy. If the condition of $e+t<s+b$ is also met, then the remaining two members of group *B* may contribute to the collective action, since it is their dominant reply. *In none of these cases does anybody from group A have a dominant strategy or a dominant reply of contribution*. In this village only high opportunity costs of contribution (high rewards for free riding) and low importance of selective incentives can help to avoid conflict and the exploitation of group *A*.

2.5.3 Model 3: Local common knowledge

If we assume extensive contacts in the neighborhood, we can also suppose that individuals are not only capable of recognizing dominant strategies of their neighbors, but also dominant replies, best replies to dominant replies, and so forth. In Model 3 we assume that people cannot only anticipate obvious actions of their neighbors, but also the neighbors' perceptions about their own behavior and the neighbors' perceptions about their own perceptions. That is, having a dominant strategy or reply of any order is common knowledge between neighbors. Under this decision rule, it is also assumed that everyone who has a dominant strategy or reply of any order acts in accordance with this strategy. In the absence of such a strategy, individuals are assumed to contribute with a fixed probability. Model 3 goes far beyond the previous models in the sense that it also takes indirect network influence into consideration. In this model, hesitating people can be influenced by neighbors who have in turn been convinced by their neighbors. At high levels of clustering, contribution spreads easier and at low levels defection does. Consequently, in this model we expect a stronger effect of segregation on the likelihood of conflict. The stronger relationship originates in the more exhaustive recruitment of peripheral areas next to the initiators of collective action. Completely isolated individuals are not assured by the action of fellow group members and are not rewarded for traitor action, therefore they remain indifferent in the intergroup game (their decision is probabilistic).

2.5.4 Model 4: Expected value calculation

Decision models 1-3 allowed individuals to make strategic calculations, but involved a pure probabilistic element in the case of the absence of a dominant strategy or a dominant reply. However, under certain circumstances it is reasonable to presume a calculative choice for these individuals. In Model 4, we assume that everyone who has a dominant strategy or reply of any order acts in accordance with this strategy, just as in Model 3. Those who do not have such a strategy will base their decision on an expected value calculation that involves estimating the number of contributing fellow neighbors and a subjective forecast of probabilities of possible outcomes (similar to Rapoport and Bornstein, 1987).

To make the model more realistic, we incorporate a certain tendency in the model that is found in experiments and is in accordance with bounded rationality. Social psychologists often claim that people usually overestimate the criticalness of their own decision (e.g., Kerr, 1989). Even if their beliefs about critical probabilities are correct, they contribute to the collective action more likely than what would follow from expected value calculations based on these probabilities (Rapoport, Bornstein, and Erev, 1989). This striking gap is also present in experimental conditions, where confirmation incentives can be excluded (cf. Caporael et al., 1989). Such a positive error can originate in people's preferences for not being responsible for the group decision and can be labeled as responsibility aversion. We incorporated this tendency in Model 4. The exact specification of the expected value rule that is adjusted for responsibility aversion can be found in Appendix 2.A.2.

In Model 4, collective action might be established in a segregated setup, even where rewards of intergroup opposition are not salient. In less segregated settings, not only direct neighborhoods, but also fellows of a larger network distance can be enforced to contribute because they might forecast contribution in the close neighborhood of initiators. On the other hand, highly mixed networks are still likely to avoid conflict.

We are also able to derive predictions about the effect of a certain type of cognitive interdependency between the players. When at least in one group there is a widespread belief that the local area is a leader in the establishment of group collective action (for instance, in many districts of the city, Serbs believe that only "good" Serbs live in that district), collective action will be more likely realized. The more people who expect a high level of contribution (conflict), the more likely it is that conflict will happen. On the other hand, expectations of peace will help the occurrence of a peaceful outcome. Hence, cognitive beliefs have an inflating effect in both directions.

2.6 Simulation design

In the previous section we introduced four models of individual behavior and we discussed general model predictions under different assumptions. We noticed that segregation increases the likelihood of intergroup conflict in all models, especially in the presence of strong selective incentives. Besides the derivation of transparent analytical results, we use simulations to derive precise predictions and provide comparative statics for all possible networks in specific settings. In the simulations, network ties represent relations between neighbors and other relations are omitted.[9] People are seldom able to escape interacting with neighbors and being influenced by them. As a consequence, different neighborhoods have different influences on

[9] Simulation programs were written in Delphi 3.

individual behavior. The empirical relevance lies in the fact that unlike other ties, neighboring connections are symmetrical (undirected) and easily mapped. Residential structures are visible and therefore our results can be interpreted easily. Furthermore it is known, that residential segregation often goes together with other forms of segregation (e.g., Whyte, 1986). The simplicity of neighborhood maps makes it possible to apply a *cellular automata* (von Neumann, 1966) alike design. Cellular automata have become useful in understanding the relation between structural embeddedness and individual choice in an interdependent situation (e.g. Messick and Liebrand, 1995; Hegselmann, 1996).

In the simulations, a grid of a rectangular grid modeled residential locations. Every location in the grid (each cell) could have three different states: occupied by a member of group *A*, by a member of group *B*, or empty. No restrictions were applied about the location of the empty cells. For instance, corner areas of the rectangle could be empty. In this way the model could also resemble cross-shape or amorphous settlements. However, simulation could provide only a simplification of residential configurations observed in reality. Simplification was also made with respect to an upper bound on neighborhood size. The usual assumptions of cellular automata based research were embraced and at most four (South, West, North, East) or eight (also SW, NW, NE, SE) adjacent cells were considered to represent neighbors (von Neumann or Moore neighbors). As in reality, at the edges of the grid, neighborhood was smaller in size. Empty adjacent cells at central locations could represent, for instance, uninhabited buildings, squares, or parks.

2.7 Measures of segregation

Since the central interest of this chapter is the relationship between structural configurations and the likelihood of conflict, it is important to describe the network structure with appropriate measures under the model settings. Therefore we briefly summarize the measurements used in computer simulation.

In the limited scope of simulations, ties connect two adjacent cells. We measured *density* by the proportion of ties that connect two nonempty cells. This measure of density approaches a simple quadratic function of the proportion of nonempty cells as grid size increases to infinity, regardless of the definition of neighborhood. The proof of this is outlined in Appendix 2.A.3 (Lemma 2.A.3.1).

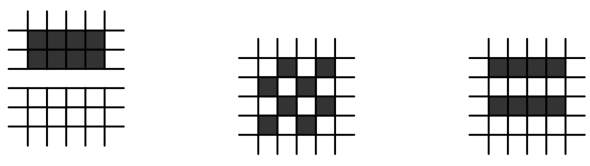
The concept of segregation is central in this research. In general, we mean by the level of segregation in a network the following:

Definition 2.7.1 The *level of segregation* in a network is defined as the proportion of relations between two members of the same group relative to all ties in the network. If this proportion is higher, then segregation is also higher.

Segregation levels are comparable given a certain level of density in the network. The definition corresponds with how the concept of *clustering* is used in the empirical literature (Lieberson and Carter, 1982). The level of clustering measures the extent to which members of one group have connections only among each other (cf. Willms and Paterson, 1995; Lieberson, 1980). On the other hand, segregation is also related to the level of *exposure* that indicates to what extent members of one group are exposed to members of the other group (Lieberson, 1980). If exposure is higher, then segregation is lower in the network. In empirical research, segregation indexes are computed from grouped data (e.g., group proportions in census tracts). As Grannis (1998) stresses, in this way the indexes provide a biased measure of neighborhood compositions. Individual behavior is influenced mainly by contacts embedded in smaller units of residential structure, hence tertiary residential-type streets or merely the closest neighbors comprise the lionshare of the explanatory focus.

In the simulations, we need to specify exactly how we measure segregation. As our simulations contain individual-level data, we can rely on micro-level indexes that are close in interpretation to the empirical isolation and exposure measures. The proportion of fellow ties (from all non-empty relations) will be used as an index of *segregation* (clustering). This measure is closely related to the individual f_i values. As in the empirical isolation index, high values indicate high levels of clustering. As grid size increases to infinity, the expected proportion of fellow ties approaches the sum of squares of the group proportions. The proof of this statement can be found in Appendix 2.A.3 (Lemma 2.A.3.2).

If a grid is more clustered than another under the von Neumann neighborhood definition, this does not necessarily mean that this grid is also more clustered under the Moore neighborhood definition. A striking example is a chessboard-like settlement, in which black and white fields represent members of the two groups respectively. In such a residential structure the segregation index is zero, if neighborhood is defined by von Neumann neighbors, but it is close to the average level, if neighborhood is defined by Moore neighbors (see Figure 2.7.1).



neighborhood definition (maximum number of neighbors)	4	8	4	8	4	8
isolation index	1	1	0	0.5	0.5	0.25
exposure index	0	0	1	0.5	0.5	0.75

Figure 2.7.1 Simple patterns of different levels of clustering

Note: Indexes are calculated for infinite grid sizes with these patterns

The proportion of opposite ties that connect members of the competing groups could be used as an index of *exposure*. This measure is closely related to the individual g_i values. High values of exposure indicate high levels of mixing. It is important to note that extreme mixing is not equivalent to a random residential structure. The proportion of fellow and opposite ties (segregation and exposure) always sums to one. The constructed segregation and exposure indexes fulfill the proposed criteria and are appropriate for the simulation analysis based on complete information. In our next examples clustering will be operationalized by the value of the segregation index.

2.8 Simulation results

2.8.1 The effect of group size and neighborhood definition

The aim of this section is to provide precise predictions about the effect of segregation on the expected likelihood of conflict under different behavioral models. In the simulations, structurally embedded IPG games were played between two groups of equal size. We considered a 10×10 grid, in which 90% of the cells were inhabited. Thus, there were 45 members in both groups. Under each decision model, a homogenous population was assumed in the sense that every player used the selected decision rule. The probability element of each decision rule was fixed to 25% of contribution.

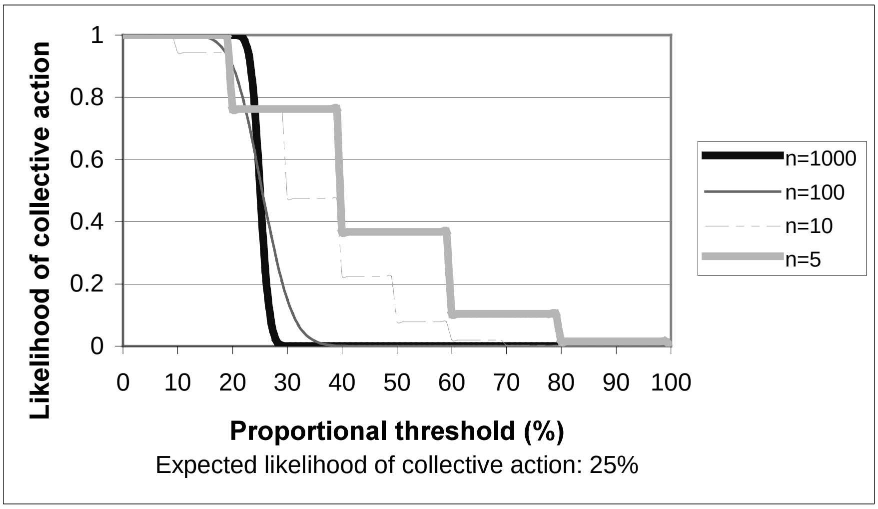


Figure 2.8.1.1 The effect of proportional threshold (MCS) on the expected likelihood of collective action

Note: Individual contribution chance is 25%.

Although we fix the parameter values of group size and minimal contributing sets in all simulations, we can briefly demonstrate the effect of these parameters on the likelihood of intergroup conflict. For this purpose, let us consider a probabilistic world in which every individual contributes with a 25% chance. Figure 2.8.1.1 shows the effect of varying the minimal contributing set on the expected likelihood of collective action of the group. The minimal contributing set is given in proportion to the group size and different lines indicate different group sizes. If the minimal contributing set is higher than the expected value (25%), then small groups can establish collective action easier than large groups. The analytical calculation of the likelihood of conflict becomes highly complicated if individual decisions are not randomly taken. This is an additional reason why we need to use computer simulation.

Before discussing the effects of segregation and individual decision rules, let us also illustrate how the definition of neighborhood influences the expected likelihood of conflict. If the same levels of clustering are considered, the expected likelihood of conflict is usually higher in the von Neumann neighborhood. Figure 2.8.1.2 shows two comparisons in decision Model 1. As a grid with Moore neighborhood represents a denser network, the examples illustrate that given a certain level of segregation, contributions and *intergroup conflict is more likely in a sparse network.* The reasons can be found in the structural conditions of having a dominant strategy. In the confirmation pressure condition only defection can be a dominant strategy. Furthermore, it can be a dominant strategy only under the Moore neighborhood definition, which means a higher likelihood of peace in this case. If local selective incentives are important, then contribution is more likely to be a dominant strategy in dense networks (cf. equations 2.4.1 and 2.4.2).

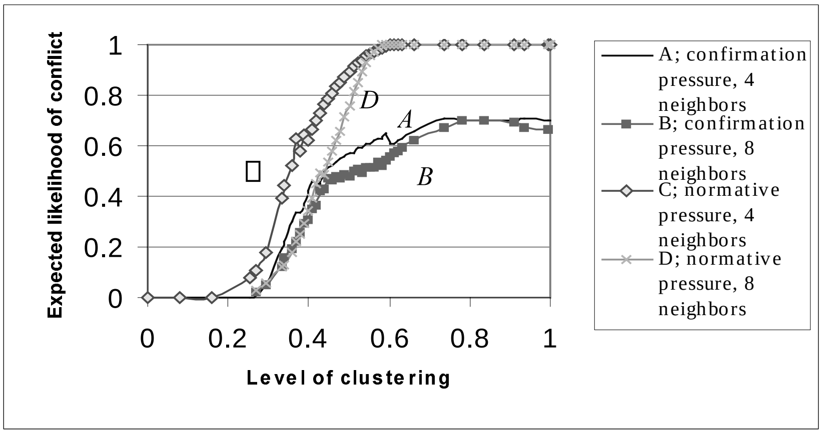


Figure 2.8.1.2 The effect of segregation on the expected likelihood of conflict, if neighborhoods are defined differently

Notes: In confirmation pressure condition: *b=2, s=1*; in normative pressure condition: *b=1, s=2*. Other parameter values: *v=5*, *e=2*, *c=-1*, *d=-5*, *t=2*. Minimal contributing set: 12 members (25%) in both groups.

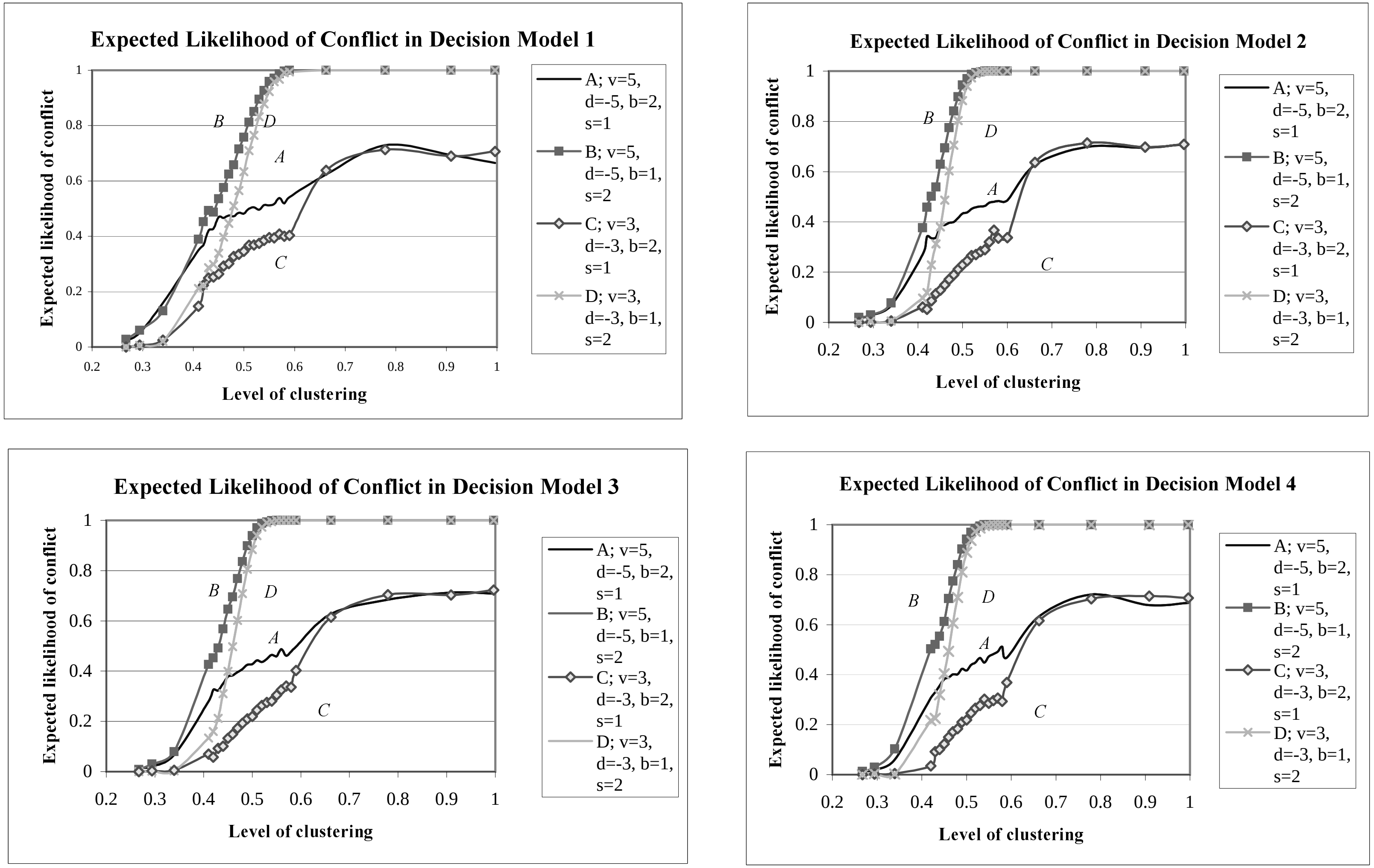


Figure 2.8.2.1. The Expected Likelihood of Conflict by Clustering, if the Minimal Contributing Set is 12 Members

Note: Curves A, B: salient; curves C, D: non-salient rewards of intergroup opposition. Curves A, C: confirmation pressure condition; curves B, D: normative pressure condition.

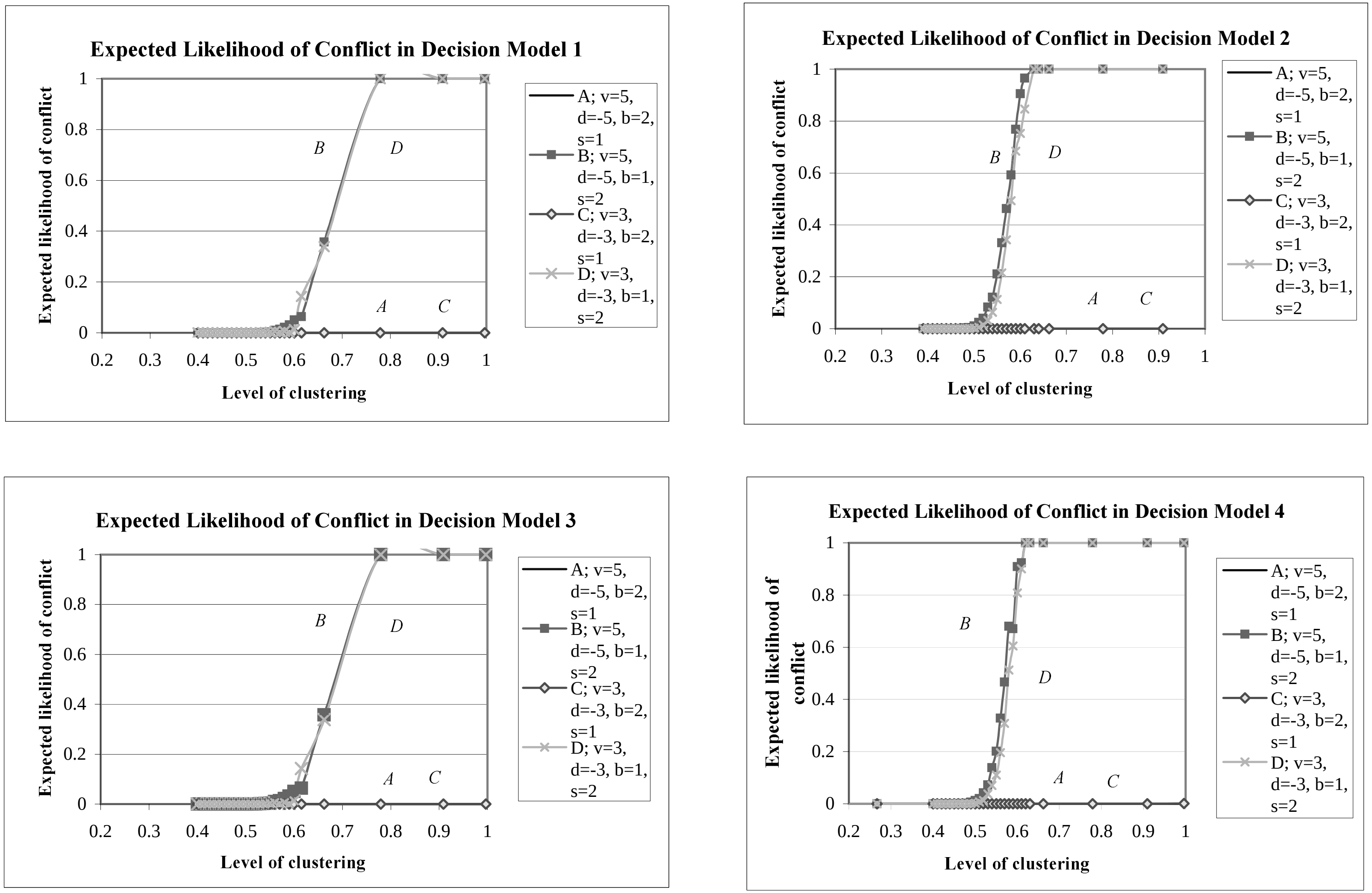


Figure 2.8.2.2. The Expected Likelihood of Conflict by Clustering, if the Minimal Contributing Set is 23 Members

Note: Curves A, B: salient; curves C, D: non-salient rewards of intergroup opposition. Curves A, C: confirmation pressure condition; curves B, D: normative pressure condition.

Curves in Figure 2.8.1.2 and in subsequent figures connect discrete cases of clustering levels. It is possible that when the average number of fellow neighbors is higher, the number of people who have enough fellow neighbors to have contribution as a dominant strategy is smaller. This causes quite a big fluctuation, especially for high and low ranges of clustering. To avoid graphic confusion, averages of expected likelihood of conflict are shown by 0.01 interval sizes in the medium range of clustering and by 0.025 interval sizes for extreme cases. However, it could be interesting to investigate what structural configurations are behind these fluctuations and what are the structural conditions that result in a high likelihood of conflict in a mixed setting or the opposite, resulting in a low likelihood of conflict in a segregated setting. We will return to this point in Section 2.9.

2.8.2 The segregation effect under different decision algorithms

Figures 2.8.2.1 and 2.8.2.2 illustrate how different individual decision rules (Models 1-4) influence the effect of segregation on the expected likelihood of conflict. The figures demonstrate this effect for two minimal contributing sets. In these examples, we considered Moore neighborhoods. In both figures, behavioral models are represented separately. On each figure, curves display the expected likelihood of conflict for four combinations of parameter values: whether rewards from the intergroup opposition are *salient* ($v=5$, $d=-5$) or *non-salient* ($v=3$, $d=-3$), and whether behavioral confirmation is more important than selective incentives (*confirmation pressure* condition; $b=2$, $s=1$) or not (*normative pressure* condition; $b=1$, $s=2$). Values of other parameters were fixed in all cases ($e=2$, $c=-1$, $t=2$).

Figures 2.8.2.1 and 2.8.2.2 show that segregation has a crucial effect on the expected likelihood of conflict in the normative pressure condition.[10] Salient payoff parameters are always associated with a higher expected likelihood of conflict. This effect is never as crucial as the difference between the confirmation pressure and normative pressure conditions. This is not surprising, since social incentives originate from network relations; meanwhile payoff parameters of the IPG game are independent from social structure. If contribution can be a dominant strategy ($s>b$), then the relationship between segregation and the expected likelihood of conflict is best described by a steep S-shape curve. In the normative pressure condition, clustering has a crucial effect on conflict in a certain critical range. This range *ceteris paribus* moves to the right (compare Figures 2.8.2.1 and 2.8.2.2) if the minimal contributing set for

[10] In the confirmation pressure condition, contribution cannot be a dominant strategy (cf. equation 2.4.1). In these cases the expected likelihood of conflict is smaller. If $b=1$ and $s=2$, then contribution is a dominant strategy for individual i, if $g_i=0$ and $f_i \geq 3$; or if $g_i=1$ and $f_i \geq 5$. If rewards from the intergroup opposition are salient ($v=5$, $d=-5$), then not to contribute is a dominant strategy for individual $i \in A$, if $f_i=0$ and $g_i \geq 3$; or if $f_i=1$ and $g_i \geq 4$; or if $f_i=2$ and $g_i=6$. For non-salient payoffs, not to contribute is a dominant strategy for individual $i \in A$, if $f_i=0$ and $g_i \geq 2$; or if $f_i=1$ and $g_i \geq 3$; or if $f_i=2$ and $g_i \geq 5$.

collective action is higher, which means that the overall likelihood of conflict is always smaller. Under certain conditions, there is no critical range. There are examples in which conflict is certain even in a grid of minimum clustering. There are also cases in which peace is expected with certainty even in a grid of maximum clustering (cf. confirmation pressure in Figure 2.8.2.2).

Within Figures 2.8.2.1 and 2.8.2.2, comparisons can be made between the effects of segregation under different decision models. The segregation effect somewhat increases and the critical range of clustering decreases, as we go towards models with more rigid behavioral assumptions. In the low ranges of clustering, the expected likelihood of conflict is lower, if a rigid decision rule is applied. In these cases, peace can be achieved easier, if the community consists of “rational” individuals with extensive information attainment. This success of calculative action can be explained by “negative” block recruitment. In the high ranges of clustering, the opposite process (positive block recruitment) can be traced. In the normative pressure condition, more and more people will have a dominant strategy (and a dominant reaction) of contributing in a segregated network. These two processes of block recruitment result in steeper curves on the figures.

However, curves only become *slightly* steeper. The processes discussed above are only present in some networks. In most of possible network structures very few individuals have a dominant reply of any order. This implies that assumptions on individual consciousness and local information are *not crucial* to determine segregation effects on conflict. However, we will see in the next section that in some specific structures more rigid assumptions on rationality definitely change predictions.

2.9 Anomalies: when segregation does not have the predicted effect

In this section we try to illustrate with examples when rational consciousness and access to additional information makes a significant difference for our model predictions. We also try to provide an explanation for the fluctuations in Figures 2.8.2.1 and 2.8.2.2. These curves are constructed by connecting discrete data points. One data point shows the expected likelihood of conflict under a given level of clustering. There is a high variation of how many network configurations belong to the same level of clustering. Furthermore, the expected likelihood of conflict might be very different for two networks with the same level of clustering. Obviously, segregation is not the only network characteristic that influences the likelihood of conflict between groups. As a short illustration, we mention here the relevance of bridges, minority hostages, and subgroups.

Bridges are believed to be of central importance in social network analysis (Granovetter, 1973). Chain reactions in collective action also require bridges that link socially distant actors (Macy, 1991a). A bridge is defined as a connection between otherwise separated units or subgroups of the network. In the intergroup context, bridges are connecting otherwise separated portions of the *same* group. Our analysis shows that bridges do not always help to diffuse contribution or defection and sometimes do not play any role in intergroup conflict. Whether bridges play brokerage or not, depends on the importance of social rewards, on the width of the bridges, and also on the environment of bridging ties. The size of social rewards determines what compositions of the individual neighborhood would allow dominant strategies or replies. Let us consider the normative pressure condition of Figures 2.8.2.1 and 2.8.2.2 with the same parameter values. The first three network segments in Figure 2.9.1 are examples of low clustering levels. In these segments, a single bridge, a double bridge, and a double bridge with bridgeheads are represented for the black group in a residential setting in which network ties are assumed to exist between Moore neighbors.

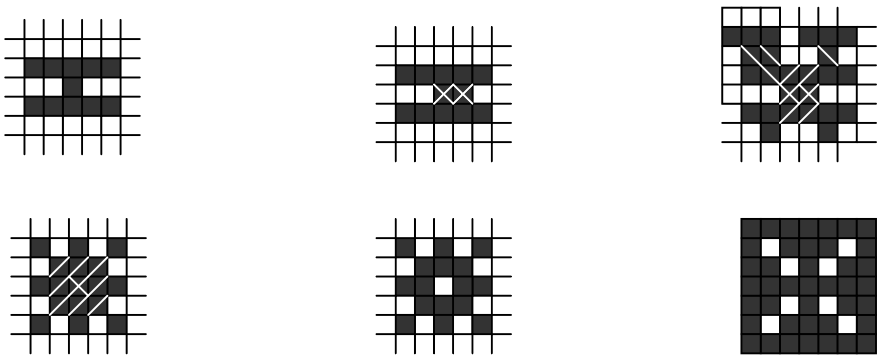
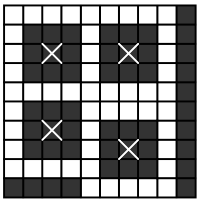


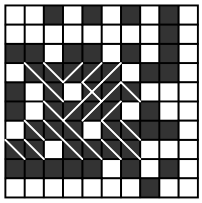
Figure 2.9.1 Examples of a single bridge, a double bridge, and bridgeheads (above) and examples of suppression of collective action by minority hostages (below)

Note: A white X denotes a dominant strategy of contribution, a white "/" stands for a dominant reply of a first order and a white "\" indicates a dominant reply of a higher order.

The examples demonstrate that under the given parameter conditions a single bridge "does not make a summer" (cf. McAdam, 1986; McAdam and Paulsen, 1993). If mediators are not alone, at least they will be active contributors to collective action. However, they can only influence the connected subgroups, if sufficient people receive their message. That is, bridges are capable of transmitting contribution incentives, if they are "wide" enough and bridgeheads are built to receive and forward these incentives. From Figure 2.9.1 we can also get an idea how a flow of contribution can be established between loosely structured subgroups that are connected by wide bridges and bridge-heads, if the local common knowledge rule (Model 3) is applied.

Another central issue in social network analysis is the role of structural holes (Burt, 1992). In the intergroup context, we redefine structural holes as empty or minority connections in a local environment that is dominated by one group. Empty cells in a homogeneous neighborhood harm neither contribution, nor defection, unless there are not many of them. In a dense structure, abandoning some fellow ties and creating structural holes has a low marginal influence on contribution, since dominant strategy and reply is more dependent on the homogeneity than on the size of the neighborhood. However, if structural holes are filled with minority "hostages", then they form a serious threat for contribution. If they have central location, they can nip collective action in the bud. As an example let us take a look at the normative pressure condition, with the same parameter values as in previous figures. In this case, minority hostages have a good chance of suppressing collective action if they are not completely alone (see Figure 2.9.1, network on the bottom right). The network on the bottom left of Figure 2.9.1 indicates the situation where the direct influence of a single individual is maximal (all neighbors have a dominant reply of contribution). The feature in the middle of the second row shows that contribution is suppressed if a structural hole is inserted at a central location.





segregation index	0.5088	0.5088
Model 1	0.88	0.49
Model 2	0.88	0.96
Model 3	0.88	1
Model 4	1	1

Figure 2.9.2 Expected likelihood of conflict in networks with equal level of clustering

Notes: The expected likelihood of conflict is indicated for salient intergroup rewards in the normative pressure condition with a minimal contributing set of 13 members (Moore neighborhood). In each decision model, the probabilistic element is 25%.

Left side: a network with small cohesive subgroups, right side: a network with a large, loosely connected subgroup. A white X denotes a dominant strategy of contribution, a white "/" stands for a dominant reply of a first order and a white "\" indicates a dominant reply of a higher order.

For the discussion of anomalies, we would like to emphasize another important point: the role of subgroup structures. What helps intergroup collective action more: many small, cohesive, but isolated subgroups or few large, loosely connected subgroups? The answer depends on the behavioral assumption we apply. If people can recognize only their dominant strategies or have very limited local information (Model 2), then isolated, but dense subgroups are more efficient in the establishment of collective action. However, if people are capable of assessing high quality information regarding their neighbors (Model 3), then large, loosely connected subgroups are more effective in mobilizing group members. Individuals at the periphery of the large group can be convinced to participate in collective action. However, key contributors are always necessary to initiate mobilization. Two networks with equal level of clustering are represented in Figure 2.9.2 as an example. In the network on the left, there are two small cohesive subgroups of black cells. In the right network, there is a loosely connected large subgroup. The expected likelihood of conflict is larger in the left structure under Model 1, but it is smaller under the assumptions of Models 2 and 3.

2.10 Discussion

The aim of this chapter was to construct a theoretical model of intergroup conflict that is based on the interdependency of individual decisions and integrates sociological insight into the analysis. The latter was achieved by focusing on social incentives that, besides the rewards from the intergroup context, influence individual decisions. Social incentives are transmitted through social ties; consequently the network structure has a strong impact on the emergence of group conflicts.

Previous research found that intergroup competition, the local assurance process, and the application of selective incentives are possible structural solutions to social dilemmas. By integrating these different ideas in a general model, we showed that *a different social dilemma might occur in which overall contribution traps the groups in a mutually harmful outcome*. Further investigation in the chapter concerned structural conditions that can help to avoid lethal conflicts. Particular attention was paid to the direct and interaction effects of network segregation. *Results indicate that segregation is likely to increase the likelihood of conflict, but not under all circumstances. Depending on other parameters, certain ranges of clustering are decisive in the determination of the outcome of intergroup competition*. This result will help us to derive hypotheses for the experiments that are discussed in the subsequent chapters. Furthermore, our findings might also have implications for real conflict management, e.g. for residential policy.

Besides aggregated consequences, our main research questions concerned underlying mechanisms on the interpersonal level (cf. Section 1.8). We were interested in the effects of social control on individual decisions and we intended to explore how different forms of social control are responsible for the segregation effect and what is the impact of their relative size. In general, *our model predicts a strong positive effect of segregation, if normative pressure is more important than confirmation pressure of neighbors and friends*. We were also interested if our results were valid under different behavioral specifications. We demonstrated that the intensified effect of segregation under strong normative pressure was present in all of our behavioral models. We showed that by assuming higher consciousness of individual decisions and better access to local information, the segregation effect becomes stronger. On the other hand, *more rigid assumptions on "rationality" made a difference only in certain network configurations*.

Possible applications of the model include ethnic conflict in neighborhoods, villages, or cities under different residential structures; conflict between football supporters in a stadium or between pupils in a classroom under different seating patterns; and participation in voting in two-party democracies. Empirical evidence from different areas provides support for many of our model predictions. For instance, residential segregation and separate education was found highly responsible for repeated conflict in Northern Ireland (Whyte, 1986). In studies of voting behavior, the classical work of Tingsten (1963 [1937]) has shown that socialist party choice is disproportionally more likely in working class districts. Further evidence of a nonlinearly increasing effect of segregation on voting was found by Butler and Stokes (1974) and Ragin (1986).

The model might have many important societal implications that will be discussed in Section 6.2. On the other hand, there is a valid concern about the limited applicability of the model to empirical situations. The ecological validity would be enhanced significantly if some of the parameter values were based on empirical data. However, the measurement of payoff parameters (especially social selective incentives and behavioral confirmation) is highly problematic. Numeration of public good rewards (e.g., social identity and nationalist pride) is also often impossible. Furthermore, the model is based on far too simple assumptions to be sufficiently competent to describe the complexity of reality. We should also mention some of the limitations. The focus on single-shot games results in the neglect of time. In the model, simultaneous actions of individuals are assumed, although in many empirical situations there are long-term delays and people can obtain information about the decision of others. The introduced dominant reply and common knowledge decision rules could however be interpreted as decisions with a certain time lag. Using this interpretation our model allows new insight also regarding the threshold models of collective action (Granovetter, 1978; Oliver, Marwell, and Teixeira, 1985; Macy, 1991a; Gould, 1993a; Chwe, 1999). More strikingly, we did not attempt to include the history of intergroup relations, which is the root of many empirical conflict situations. The neglect of history will be relaxed in Chapters 4 and 5 by iterating the game over time.

Appendix

2.A.1 Model specification using matrix algebra

Individual payoffs in the modified IPG game can be expressed as follows. Formally, the set of possible strategies (ω_i) for player $i \in A$ are {0,1}, where 1 stands for the contributing action. Let us denote the number of contributors in group A and B respectively by k_A and k_B and the minimal contributing sets (MCS) by k_A^* and k_B^*. Given the strategy combination vectors $\boldsymbol{\omega}_\mathbf{A}$ and $\boldsymbol{\omega}_\mathbf{B}$ (dimensions of $n_A \times 1$ and $n_B \times 1$), the payoff for i is determined by:

$$h_i(\omega_A, \omega_B) = e(1 - \omega_i) + \iota_{k_A+1}'\mathbf{H}\iota_{k_B+1},$$

where ι_{kA+1} is a $(n_A+1)\times 1$ vector of zeros with 1 in the k_A+1-th place. The expression $\iota_{kA+1}'\mathbf{H}\iota_{kB+1}$ points to the $\{k_A+1;\ k_B+1\}$ element of matrix **H**. The $(n_A+1)\times(n_B+1)$ dimension **H** matrix contains the payoffs from the intergroup context for individual $i \in A$. If there are k_A contributors in group A and k_B contributors in group B, then the $\{k_{A+1};\ k_{B+1}\}$ element of the **H** matrix shows the reward for individual $i \in A$. In general, the **H** matrix can be partitioned into four submatrices. The $k_A^* \times k_B^*$ dimension $\mathbf{H}_1$ submatrix contains zeros as elements, the $k_A^* \times (n_B - k_B^* + 1)$ dimension $\mathbf{H}_2$ submatrix contains d's as elements, and the $(n_A - k_A^* + 1)\times k_B^*$ dimension $\mathbf{H}_3$ submatrix contains v's as elements. In the $(n_A - k_A^* + 1)\times(n_B - k_B^* + 1)$ dimension $\mathbf{H}_4$ submatrix, there are c's, if the row number equals the column number in matrix **H**. Furthermore, there are v's if the row number exceeds the column number and there are d's if the column number is higher. For example, if there are five members in group A, seven in group B, and the minimal contributing set is the minimal majority, then the 6×8 size **H** matrix is:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \end{bmatrix} = \begin{bmatrix} 0^* & 0 & 0 & 0 & d^* & d & d & d \\ 0 & 0 & 0 & 0 & d & d & d & d \\ 0 & 0 & 0 & 0 & d & d & d & d \\ v^* & v & v & v & d & d & d & d \\ v & v & v & v & c & d & d & d \\ v & v & v & v & v & c & d & d \end{bmatrix}$$

Payoffs with an asterisk indicate rewards from a Nash equilibrium outcome. There are three Nash equilibria: one in which nobody contributes, one in which there are three contributors in group A and none in group B, and one in which there are no contributors in group A and there are four in group B. To facilitate the interpretation of the **H** matrix, this latter equilibrium belongs to the fifth element in the first row of the matrix. The corresponding payoff d denotes that group A is defeated.

In the structurally embedded IPG game, these payoffs are extended by rewards from interpersonal relations (local coordination games). The strategy set of player $i \in A$ is again {0,1}, where 1 stands for the contributing action. Let us denote the number of i's neighbors from group A by f_i and the number of neighbors from group B by g_i. The individual payoff for $i \in A$, given the strategy combination ($f_i \times 1$) vector ω_f for the fellow neighbors, is determined by:

$$h_i(\omega_A, \omega_B, \omega_f) = (e + g_i t)(1 - \omega_i) + \iota_{k_A+1}{}'\mathbf{H}\iota_{k_B+1} + sf_i\omega_i + b\{\omega_i\omega_f{}'\omega_f + (1-\omega_i)(f_i - \omega_f{}'\omega_f)\},$$

where ι, the **H** matrix and the other elements are the same as before.

2.A.2 Detailed specification of the expected value calculation (*Model 4*)

In Model 4, we assume that everyone who has a dominant strategy or reply of any order acts in accordance with this strategy. Those who do not have such a strategy, will base their decision on an expected value calculation that involves an estimation of the number of contributing fellow neighbors and a subjective evaluation of probabilities of possible outcomes.

The estimation can be based on the following calculation. Denote the subjective probabilities of the seven states by P_z's $\left(\sum_{z=I}^{z=VII} P_z = 1\right)$ and the estimated number of fellow neighbors for i who will contribute by $\hat{f}_{ic}$. For the sake of simplicity, let us assume that rewards are numerical and individual utility is a linear function of rewards. We obtain that contribution is a better choice, if

$$P_{III}(c-d) + P_{IV}v + P_V(v-d) + P_{VI}(v-c) + \hat{f}_{ic}(b+s) + (f_i - \hat{f}_{ic})(s-b) > g_i t + e. \tag{2.A.2.1}$$

We have to make further assumptions regarding how individuals determine critical probabilities. The calculation could be based on an approximation from the binomial distribution. People however, are unlikely to make calculations in this sophisticated way (cf. experimental results of Rapoport and Bornstein 1989), especially if it is problematic to translate rewards into utilities. On the other hand, in making decisions individuals certainly consider what the probable outcomes of the intergroup opposition will be and how their neighbors will behave. Therefore, results are aimed only at highlighting the tendencies of how the "vision of rational man" can change the predicted likelihood of intergroup conflict and the predicted relationship between segregation and conflict.

To make the model more realistic, we incorporated a certain tendency that individuals often overestimate the criticalness of their own decision (Kerr, 1989; Rapoport, Bornstein, and Erev, 1989). Precisely, we assume that if individuals do not have a dominant strategy or a dominant reply of any order, (local common knowledge is assumed about obvious reactions of neighbors), they will then use an expected value calculation based on inequality (2.A.2.1). Critical probabilities P_{III}, P_{IV}, P_V, and P_{VI} (cf. Table 2.2.2) are obtained from a binomial calculation that sums the probabilities of all possible events for the given outcome:

$$P_{III} = \sum_{j=\max(k_A^*;k_B^*)}^{\min(n_A;n_B)} \binom{n_A-1}{j-1} p_A^{*\,j-1}\left(1-p_A^*\right)^{n_A-j} \cdot \binom{n_B}{j} p_B^{*\,j}\left(1-p_B^*\right)^{n_B-j},$$

$$P_{IV} = \binom{n_A-1}{k_A^*-1} p_A^{*\,k_A^*-1}\left(1-p_A^*\right)^{n_A-k_A^*} \cdot \sum_{j=0}^{k_B^*-1} \binom{n_B}{j} p_B^{*\,j}\left(1-p_B^*\right)^{n_B-j},$$

$$P_{V} = \binom{n_A-1}{k_A^*-1} p_A^{*\,k_A^*-1}\left(1-p_A^*\right)^{n_A-k_A^*} \cdot \sum_{j=k_B^*}^{k_A^*-1} \binom{n_B}{j} p_B^{*\,j}\left(1-p_B^*\right)^{n_B-j},$$

and

$$P_{VI} = \sum_{j=\max(k_A^*;k_B^*)}^{\min(n_A-1;n_B)} \binom{n_A-1}{j} p_A^{*\,j}\left(1-p_A^*\right)^{n_A-j-1} \cdot \binom{n_B}{j} p_B^{*\,j}\left(1-p_B^*\right)^{n_B-j},$$

where p_A* and p_B* denote the *subjective probability* that a representative individual contributes to the group collective action in group *A* and *B*, respectively. We assume that people think both groups are homogenous in a sense that they order the same subjective probability to each actor’s action in the given group (cf. Rapoport and Bornstein, 1987). In the simulations, we assumed both p_A* and p_B* to be 0.25. Responsibility aversion is incorporated in the decision rule as when the procedure described does not result in contribution people are still allowed to contribute with a fixed probability. In the simulations, this probability was set to 0.25.

2.A.3 Statistical properties of network measures

In this part of the Appendix we derive some statistical properties of the network measures used in the simulation. In all simulations, a grid of a rectangle form was considered with *R* rows and *C* columns.

Proposition 2.A.3.1. The total number of dyadic connections T is equal to $2RC-R-C$, if von Neumann neighbors are considered.

Proof. Every cell in the middle (sum (R-2)(C-2)) has four adjacent cells, every cell at the edges (sum 2R+2C-8) has three adjacent cells, and the four cells in the corners have two adjacent cells. Summing this, every dyad is counted twice, thus

$$T = \frac{4(R-2)(C-2)+3(2R+2C-8)+2\cdot 4}{2} = 2RC - R - C$$

which completes the proof the proposition.

Proposition 2.A.3.2. The total number of dyadic connections T is equal to $4RC-3R-3C+2$, if Moore neighbors are considered.

Proof. Every cell in the middle (sum (R-2)(C-2)) has eight adjacent cells, every cell at the edges (sum 2R+2C-8) has five adjacent cells, and the four cells in the corners have three adjacent cells. Summing this, every dyad is counted twice, thus

$$T = \frac{8(R-2)(C-2)+5(2R+2C-8)+3\cdot 4}{2} = 4RC - 3R - 3C + 2$$

which proves the proposition.

LEMMA 2.A.3.1. Denote the grid size by S (S=RC) and the population density (the proportion of inhabited cells) by $\pi = n_{A+B}/S$. We assume no restrictions on how the grid is filled, that is every location can be filled with a probability of π. The *density* of network relations is measured by the proportion of nonempty dyads (ties) and it is denoted by δ. Irrespective of the definition of neighborhood (von Neumann or Moore neighbors), the expected density of network relations $E(\delta)$ is obtained by

$$E(\delta) = 2\pi - 1 + (1-\pi)\frac{(1-\pi)S-1}{S-1}.$$

The larger the grid size (S) is, the closer the expected value is to π^2, i.e.

$$\lim_{S\to\infty} E(\delta) = \pi^2.$$

Proof. A network relation connects two adjacent cells in the grid. We do not make specific assumptions about the definition of "adjacent" cells; thus our results are valid for any neighborhood definitions. Consider a randomly selected dyad. This dyad is empty when one of the cells it connects is empty. The probability of one cell being empty is $1-\pi$. The same holds for the adjacent cell. However, the two events are not independent given the fixed population density of the grid (the two locations are filled *without* replacement). Therefore, for the calculation of the probability of this dyad being empty the joint probability of the two events have to be subtracted, which is

$$(1-\pi)\frac{(1-\pi)S-1}{S-1}.$$

The probability of this dyad being nonempty is one minus the probability of this dyad being empty, thus irrespective of the definition of neighborhood,

$$P(\delta)=1-\left[(1-\pi)+(1-\pi)-(1-\pi)\frac{(1-\pi)S-1}{S-1}\right]=2\pi-1+(1-\pi)\frac{(1-\pi)S-1}{S-1}.$$

Since the dyad was selected randomly, by applying the expectation of the binomial distribution we get $E(\delta)=P(\delta)$ and the proof is given.

$$\lim_{S\to\infty}E(\delta)=\lim_{S\to\infty}\left[2\pi-1+(1-\pi)\frac{(1-\pi)S-1}{S-1}\right]=2\pi-1+(1-\pi)^2=\pi^2.$$

LEMMA 2.A.3.2. Denote the proportional group sizes by α ($\alpha=n_A/n_{A+B}$) and β ($\beta=1-\alpha$). We assume no restrictions on how nonempty cells are filled, that is every nonempty location can be filled with members of group *A* with a probability of α. The proportion of fellow ties ϕ (from all non-empty relations) will be used as an index of *segregation* (*clustering*). Irrespective of the definition of neighborhood, the expected proportion of fellow ties $E(\phi)$ from all non-empty relations is obtained by

$$E(\phi)=\frac{\alpha(n_A-1)+\beta(n_B-1)}{n_{A+B}-1}.$$

By enlarging the grid, the expected value gets closer to $\alpha^2+\beta^2$, that is

$$\lim_{S\to\infty}E(\phi)=\alpha^2+\beta^2.$$

The proportion of opposite ties γ (from all non-empty relations) will be used as an index of *exposure*. The index of segregation (clustering) and the index of exposure sums to one. Irrespective of the definition of neighborhood (von Neumann or Moore neighbors), the expected proportion of opposite ties $E(\gamma)$ from all non-empty relations is obtained by

$$E(\gamma) = 1 - \frac{\alpha(n_A - 1) + \beta(n_B - 1)}{n_{A+B} - 1} = \frac{2\alpha\beta n_{A+B}}{n_{A+B} - 1},$$

and

$$\lim_{S \to \infty} E(\gamma) = 2\alpha\beta.$$

For the proof of LEMMA 2.A.3.2, consider the following proposition.

Proposition 2.A.3.3. Without the loss of generality, let us consider group A, with a relative proportion of α. The expected proportion of ties (from all non-empty relations) connecting two members of group A (denoted by ρ) is

$$E(\rho) = 1 - \left[2(1-\alpha) - (1-\alpha)\frac{(1-\alpha)n_{A+B} - 1}{n_{A+B} - 1}\right].$$

Proof of the proposition. The placement of αn_{A+B} individuals to n_{A+B} locations and the determination of the proportion of ties connecting members of group A is equivalent to the placement of πS nonempty fields to S locations and the determination of the proportion of network relations. Thus using LEMMA 2.A.3.1 and substituting α to π, the proof is given.

Proof of the lemma. Fellow ties connect two individuals from the same group. The probability of a randomly selected nonempty tie being a fellow tie is

$$P(\phi) = 1 - \left[(1-\alpha) + (1-\alpha) - (1-\alpha)\frac{(1-\alpha)n_{A+B} - 1}{n_{A+B} - 1}\right] + 1 - \left[2\alpha - \alpha\frac{\alpha n_{A+B} - 1}{n_{A+B} - 1}\right] = \frac{\alpha(n_A - 1) + \beta(n_B - 1)}{n_{A+B} - 1}.$$

Since the tie was selected randomly, by applying the expectation of the binomial distribution we get $E(\phi)=P(\phi)$, which completes the proof.

$$\lim_{S \to \infty} E(\phi) = \lim_{S \to \infty}\left[\alpha\frac{\alpha n_{A+B} - 1}{n_{A+B} - 1} + \beta\frac{\beta n_{A+B} - 1}{n_{A+B} - 1}\right] = \alpha^2 + \beta^2.$$

*"…in group relations, some degree of prejudice-hostility always appears when there is the combination of **visibility** and **competition**."*

Robin M. Williams, Jr.: The Reduction of Intergroup Tensions. A Survey of Research on Problems of Ethnic, Racial, and Religious Group Relations (1947: 54)

CHAPTER 3

SOCIAL CONTROL IN INTERGROUP COMPETITION

Experimental evidence

3.1 Introduction

Single-shot social dilemma experiments often find unexpectedly high cooperation rates. Even in strictly impersonal settings, in which no communication is allowed and subjects are completely strangers, many people tend to act against their own interests and make sacrifices for the collectivity. In a competition situation with another group, it is even more likely that subjects put their direct monetary incentives aside in favor of contributing to the public good provision (Bornstein and Ben-Yossef, 1994). In intergroup competition experiments (e.g., Bornstein, Erev, and Rosen, 1990), subjects were not allowed to redistribute payoffs, to invent binding commitments, or to apply coercion. High level of cooperation was nevertheless achieved.

Why do people tend to act in favor of their groups if it is not in their interest to do so? This question is particularly challenging to answer when widespread cooperation within competing groups has lethal consequences for members of both groups. A harmful outcome provides just another incentive against contribution, but people still not refrain from making sacrifices of this kind for their group. Such action is difficult to explain assuming perfect individual rationality. In this chapter, we do not discuss limitations of rationality, but we raise doubts about whether the monetary payoff structure of experimental games fully describes the incentives of players in the laboratory. We claim that there are also other substantial utility concerns. We will concentrate on the role of incentives that stem from *structural embeddedness*. We try to demonstrate that elementary structures might be present in the laboratory and can make a significant difference to contribution decisions, even when subjects do not know each other and are not allowed to communicate.

In this way, this is the very first attempt to incorporate structural embeddedness in the experimental analysis of intergroup competition. As a model of intergroup relations, we will use the structurally embedded Intergroup Public Goods (IPG) game that was discussed in Chapter 2. We developed this model in order to explain how structural embeddedness influences the likelihood of intergroup conflict. The theoretical results we derived for this question will guide us as hypotheses in the experiments. Among effects of structural embeddedness this study is particularly interested in the effect of *segregation* on intergroup conflict. In the theoretical analysis of Chapter 2 we found that in general the effect of segregation on the likelihood of intergroup conflict can be characterized by an S-shape function. This implies that segregation is likely to promote intergroup conflict, but in certain ranges of segregation, an additional change does not result in an increase in the likelihood of conflict. These results directly lead to one of the main research questions we would like to answer in the experiments. This regards whether the likelihood of intergroup conflict is higher when group members are arranged in a segregated pattern in the experiments.

As an underlying mechanism, in Chapter 2 we analyzed the impact of *social control* on individual decisions. In previous experiments, there has been little focus on structural embeddedness and social control. There are explanations of seemingly irrational cooperation that concentrate on group identity, on fairness, or on other related concepts (social commitment, conformity), but we do not exactly know how these processes work in practice. There are some indications of the presence of social control in earlier social dilemma experiments (van de Kragt, Dawes, and Orbell, 1988; Rapoport, Bornstein, and Erev, 1989), but without clear conceptualization of these effects.

As in Chapter 2, in the experiments we focus on three fundamental forms of social control. The first is *behavioral confirmation* that expresses the subject's desire to conform to the expected behavior of other individuals. The second is the distribution of *social selective incentives,* such as prestige or respect. The third form is present in network relations between members of the opposite groups and is referred to as a *traitor reward.* It is a social selective incentive that punishes contribution and rewards defection. For the derivation of hypotheses about these social control mechanisms, results from Chapter 2 give directives. These results show that the larger the size of selective incentives compared to behavioral confirmation, the stronger the effect of segregation on intergroup conflict. In the experiments, we explore in which forms social control influences individual behavior and whether the segregation effect on intergroup conflict is indeed stronger under normative pressure than under confirmation pressure.

For the conceptualization of social control in the laboratory, we have to consider that social control might become *internalized* (Scott, 1971; Coleman, 1990: 293). The internalization of *contribution norms* creates a cognitive reward for cooperation. This motive, that can be summarized as "I do the right thing for the group", is relevant even if individual decisions are completely confidential (Opp, 1989). Similar explanations were provided for surprisingly high contribution rates in social dilemmas and for high participation rates in collective action situations from various theoretical perspectives. Such motivations have already been recognized by Smith (1976[1759]) and were as a "universal fellow feeling". The interpersonal character of these intrinsic motivations were acknowledged by Sen (1977), who called them "social commitment", and by Elster (1990), who specified these behavioral incentives as a "sense of duty". They argued that social commitment and respectively the sense of duty is obviously stronger in the direction of somehow related individuals (cf. Caporael et al., 1989). Others attributed the selective character of internalized contribution norms to internalized values of altruistic or non-selfish behavior (Kenrick, Baumann, and Cialdini, 1979; Cialdini et al., 1987).

Some concepts that have appeared in the literature as possible solutions for social dilemmas show similarity with *behavioral confirmation*. *Reciprocity* can be an efficient behavioral pattern in repeated social dilemmas (Axelrod, 1984), but it can also be present in single-shot situations in a cognitive form. For instance, individuals might believe that others are good contributors and they might believe that others have similar beliefs. Furthermore, they might also think that there is a mutual consensus about considerations of reciprocal fairness (cf. Camerer, 1997). Besides reciprocity, another mechanism that shows similarity with behavioral confirmation is *imitation*. Even if others are not able to monitor individual choice, people tend and like to behave as others or imitate behavioral samples (Asch, 1956; Dawkins, 1976; Pingle, 1995). In classical sociology, Weber (1978[1913]: Chapter I.4, 29) specified "custom" similar to internalized behavioral confirmation: "As distinguished from both 'convention' and 'law', 'custom' refers to rules devoid of any external sanction. The actor conforms with them of his own free will…" In a certain sense, fashion works in a similar way, as individuals gain enjoyment by confirming the behavior of relevant others (e.g., Richardson, 1948a; 1948b).

Traitor rewards are also likely to be internalized as a fear of local conflict and benefit for local harmony. Betraying someone else elicits an unpleasant feeling of guilt; therefore people try to avoid this (e.g., Poundstone, 1992: 223).

We believe that these forms of social control influence individual contribution decisions in intergroup related collective action. As an aggregated consequence of dyadic social control, the network structure of individual relations influences the likelihood of intergroup conflict. Dense in-group relations and scarce out-group relations are correlated with extensive distribution of social selective incentives between fellows and limited realization of traitor rewards. Hence, segregation supports the emergence of harmful intergroup clashes. The key elements of the explanatory mechanism are the fundamental forms of social control. In this chapter, we will examine what types of direct and internalized social control influence the decision of subjects to contribute or not to group collective action in controlled experimental conditions. In short, we are interested in what forms of social control back the segregation effect on intergroup conflict.

To test the presence of different forms of social control and the segregation effect on intergroup conflict, we introduced a *new experimental design*. Seating patterns were varied and visibility conditions were manipulated in order to detect forms of social control that are activated conditional on the neighborhood composition. We implemented a full segregation, a complete mixture, and a medium segregation condition. Additionally, in the control condition, subjects made decisions in isolation. As an aggregated consequence of local social control we derived a prediction that intergroup conflict will be more likely in segregated seating configurations. This prediction was tested in the experiments.

Subjects were unknown to each other. They were not able to communicate in any form and there was no way to learn of the decisions of others. Consequently, there were no real network relations between the subjects. For testing the presence of internalized social control, we introduced *minimal contact* between neighboring subjects. Minimal contact meant that subjects were able to see each other and they were able to identify the group membership of each other. We were interested in whether this minimal contact is sufficient to activate internalized forms of social control in the experiment.

Experimental evidence shows that face-to-face contact facilitates cooperation in conflict situations (Drolet and Morris, 2000). Previously, this finding was explained by the social psychological process of rapport that is conceptualized as a “state of mutual positivity and interest that arises through the convergence of nonverbal expressive behavior in an interaction” (Drolet and Morris, 2000: 27; Tickle-Degnen and Rosenthal, 1990). There is no doubt that when subjects are able to communicate with nonverbal signs or are able to send emotional signals, they influence the behavior of each other in the social dilemma task. However, we would like to emphasize that minimal contact has an additional effect that is due to the activation of internalized social control. To show this, extensive nonverbal communication was disallowed in the experiment and decision times were restricted to avoid initiations of signaling.

Our experiments provide a test whether we can partly explain a higher level of contribution in the IPG game by internalization of social control (behavioral confirmation, selective incentives, and traitor rewards) under strictly impersonal experimental conditions with only a minimal contact between subjects. In later parts of the experiments, behavioral confirmation and selective incentives were introduced as monetary side-payments. In this way we tried to model real situations, in which related individuals experience direct social control. As we deal with unrelated subjects in the experiments in artificial circumstances, it was only possible by transforming the utility of these social incentives into monetary rewards. Obviously, we predict these effects to be stronger. With their introduction, however, our main intention was to provide a meaningful comparison between the size of monetary and internalized social incentives.

A deeper question concerns why do people experience social control and why do they experience social incentives in relation to strangers. Evolutionary psychology offers an answer in the ancestral environment (Cosmides and Tooby, 1992). At a certain time of human evolution, survival success and adaptation to the environment was increased by the emergence of human sociality. There is a growing literature that helps to explain the evolution of early human sociability by using computer simulation or game theoretic tools (e.g., Axelrod, 1986; Allison, 1992; Young, 1993; Binmore, 1998; Bowles and Gintis, 1998; Bateson, 2000). Mainstream sociology, however, concentrates not on natural selection, but on culture and norms that are learnt as “rules of the game” during the socialization process. It is not our goal to be involved in this debate on nature and nurture (for a synthesis, see Gintis, 2001). We are not interested in the origins of social incentives; we just aim to test whether they operate also in an impersonal environment.

From previous experiments we know that many other factors influence cooperation rates in social dilemmas (e.g., Ledyard, 1995). These factors might be relevant also in intergroup related collective action. With the introduction of personal characteristics as control variables into the analysis, we will be able to test the significance of “traditional” research variables as gender, study direction, experience with similar experiments, risk preferences, and social orientations. We show whether prosocial individuals are more concerned about harmful outcomes and thus abstain from contribution or whether they have higher contribution propensities and are even the initiators of harmful intergroup conflict.

We have to emphasize that decision algorithms that we discussed in Section 2.4 (dominant strategy, dominant reply, local common knowledge, and expected value rules) are difficult to be traced in the experiments. Looking only at the monetary rewards, subjects have no dominant strategies in the payoff structure of the experiment. We are not able to test the hypothesis that some subjects might evaluate social control so much that contribution becomes unconditionally their best choice. Furthermore, subjects have no information about the utility function of neighboring others as it is assumed partly in the dominant reply, the local common knowledge, and in the expected value rules. Similar problems hold for subjective expected utility rules (cf. Esser, 1986) and other belief-based models (Camerer and Ho, 1999).

In this section we provided an introduction to the experimental analysis of intergroup competitions. We discussed motivations to investigate structural effects that until now have been neglected in the experimental literature. In Section 3.2 we formulate our hypotheses for the experiments. In Section 3.3 we introduce the new experimental design. The model we use to test our hypotheses is specified in Section 3.4.1. A discussion about which control variables are included in the analysis can be found in Section 3.4.2. Section 3.5 describes the method with the discussion of some major characteristics of the subjects, the experimental procedure, and some technical difficulties. Results in Section 3.6 contain descriptive statistics and the discussion of questionnaire data, followed by the multilevel analysis of contribution propensities. Conclusions are drawn in Section 3.7.

3.2 Research questions and hypotheses

The experiments were designed to answer research questions about the effect of structural embeddedness on intergroup conflict. At the aggregated level we aim to answer whether harmful intergroup outcomes more likely in sessions in which members of the two groups are seated in a segregated pattern. With regard to the underlying micro mechanisms, our main research question concerns whether social control affects individual decisions in a direct and in an internalized form. These questions will be answered by analyzing individual decisions in a series of single-shot IPG games.

The general conclusion of previous IPG experiments is that individuals often make sacrifices for their group. In our experiments we will test whether higher contribution rates can be partly explained by effects of social control. Furthermore, we are interested whether eye contact between the subjects is sufficient to trigger internalized forms of social control in impersonal experimental conditions. We will test the presence of three types of social control mechanisms: selective incentives (*s*), behavioral confirmation (*b*), and traitor rewards (*t*). When internalized, these forms of social control create non-monetary incentives for the subjects, which can be expressed as part of their utility functions. Since both monetary and these non-monetary incentives enter the game, the "real" payoff matrix is not equivalent to the "monetary" payoff matrix. The "real" game could have multiple equilibria (cf. Gächter and Fehr, 1999). Contribution, for instance, could be a dominant strategy (cf. inequality 2.4.1). Formally, in the experimental game contribution is a dominant strategy of individual *i,* if assuming a linear utility function on rewards and linear effects of network ties,

$$f_i(s_i - b_i) > g_i t_i + e_i \tag{3.2.1}$$

holds, where e_i stands for the reward for free riding, f_i denotes the number of fellow and g_i the number of opposite neighbors of *i*. The relative weight of the utility of monetary rewards and of the utilities attached to different forms of non-monetary incentives can change from person to person.[1] We do not assume a specific form of utility function that can be applied to everyone. However, we are interested only in mean tendencies and in trying to estimate the average relative size of different forms of social incentives.

Our main interest is in the analysis of social control that is experienced in dyadic contact. We predict that social control enters subjects' considerations in an internalized form when eye contact is established and in a direct form when monetary side-payments are included. With regard to different forms of social control, we can explicate the following hypotheses. These hypotheses are derived directly from the discussion in Chapter 2.

Selective incentives always reward contribution or punish defection (Olson, 1965).

Selective incentives: *Selective incentives both in an internalized and in a monetary form have a positive effect on contribution propensities.* More connections to group fellows means the distribution of selective incentives from multiple sources. *Hence, the higher the number of group fellows connected, the higher the contribution rate is.*

On the other hand, the effect of behavioral confirmation is not only dependent on the composition of the network neighborhood, but also on expected decisions of neighbors. We presume that subjects do not make qualitative differences between fellow neighbors. We can therefore derive the following hypothesis:

1 For this reason, we use the subscript *i* for non-monetary incentives in inequality (3.2.1). Later on, this subscript will be omitted.

Behavioral confirmation: *We predict that behavioral confirmation both in an internalized and in a monetary form have an effect on contribution propensities. The direction and the size of the effect depend on the number of expected contributors and on the number of expected defectors among group fellows connected. If the former is higher, the effect is positive. If the latter is higher, the effect is negative. The higher the difference, the higher the size of the effect is.*

The presence of neighbors from the opposite team triggers an internalized form of social control we called traitor rewards. Because of research limitations and for the sake of simplicity, we did not introduce these incentives in a monetary form in the experiments.

Traitor rewards: *We predict that internalized traitor rewards have a negative effect on contribution propensities. The higher the number of members of the opposite group connected, the lower the contribution rate is.*

Network connections were conceptualized as adjacency in the seating configuration in the experiment. As we believe that neighbors are the direct source of social control, different neighborhood compositions would lead to different contribution propensities. At the aggregated level, higher contribution propensities mean higher likelihood of intergroup conflict. As a consequence, different outcomes can be predicted for different neighborhood structures. Theoretical analysis in Chapter 2 has shown that from the nature of the specified social control mechanisms it follows that segregation is likely to promote intergroup conflict. On the basis of this result, we formulate the following hypothesis about the macro consequences of dyadic social control in the IPG experiments:

SEGREGATION HYPOTHESIS: *In a segregated structure, contribution rates will be higher and intergroup conflict will be more likely.*

Furthermore, the analysis in Chapter 2 also specified the impact of the relative size of social control mechanisms on intergroup conflict. Results demonstrated that the segregation effect on intergroup conflict is stronger where selective incentives are relatively important when compared to behavioral confirmation. In order to test this theoretical prediction, a *normative pressure* condition and a *confirmation pressure* condition was implemented in the experiments. In the normative pressure condition, selective incentives were introduced as monetary side-payments. In the confirmation pressure condition, monetary behavioral confirmation rewards were included. On the basis of the theoretical prediction, the hypothesis we formulate for the experiments is as follows:

The segregation effect on the likelihood of intergroup conflict will be stronger in the normative pressure condition than in the confirmation pressure condition.

In this section, we summarized our main research questions and hypotheses for the experiments. In the next section, we introduce a new experimental design that is invented to test these hypotheses.

3.3 Experimental design

In this section, the structure of the extended IPG game used in the experiments will be outlined. There were two groups: the red group and the green group consisting of five members each. Every player decided individually whether to keep a bonus of 11 points completely (in the experiments, 1 point was equivalent to 1 NLG = 0.42 USD) or to give all of it to help their group in the competition. Depending on the number of contributors in the groups, public good and "bad" rewards were distributed equally among all group members. The sizes of these rewards in the experiments are shown in Table 3.3.1.

Everyone in the groups received these rewards, regardless of the decision to keep or give away the bonus of 11 points. Table 3.3.1 does not include the bonus reward that is added to the payoff of those subjects who decided to keep it. Moreover, to ensure positive payoffs, every subject was entitled to 15 points additional payment at the end of the experiment.

Table 3.3.1 The IPG game (with clash punishment) used in the experiments

payoffs 1 point = 1 NLG		*number of contributors in the green group* 0	1	2	3	4	5
number of contributors in the red group	0	0 0	0 0	0 0	15 -15	15 -15	15 -15
	1	0 0	0 0	0 0	15 -15	15 -15	15 -15
	2	0 0	0 0	0 0	15 -15	15 -15	15 -15
	3	-15 15	-15 15	-15 15	-11 -11	15 -15	15 -15
	4	-15 15	-15 15	-15 15	-15 15	-11 -11	15 -15
	5	-15 15	-15 15	-15 15	-15 15	-15 15	-11 -11

Note: The payoffs are public good rewards distributed to everyone in the red (bottom left corner of each cell) and in the green (top right corner) group.

Due to Definition 1.2.1.1, in intergroup conflict the collective interest of a group is hurt by the collective action of the other group. In the experiment, if there were at least three contributors in one or both of the groups, negative rewards were distributed (cf.

Table 3.3.1). These are the combinations (victory of one side or clash) we call *intergroup conflict* in the experiment.

There are three pure strategy Nash equilibria in this game. One is when nobody contributes. In this situation, no player can be better off by giving the bonus away since the outcome cannot be changed by a single contribution. The other equilibria are when there are three contributors in one group and none in the other team. In this situation, contributors do not have an incentive to switch because the reward for victory outweighs the value of the bonus. Free riders are happy in the winning group and alone they cannot change the situation in the losing group. Since communication is not allowed between the players, there are strong arguments against the occurrence of the latter equilibria. First, nobody would be really interested in making the sacrifice when he or she could also free ride on the effort of other team members. Second, making such a sacrifice does not mean success immediately; there should be at least two other fellows who are willing to make this decision. This makes contribution a very risky alternative and makes the latter equilibria vulnerable (cf. Selten, 1975). On the other hand, overall abstaining from contribution is a stable equilibrium. Although keeping the bonus is not a dominant strategy in the game, it nevertheless provides, almost under every condition, a better individual payoff.[2]

In order to obtain more data in the experiments, the game was played many times in each session, but subjects received payments in a randomly selected single round only. No information was provided during the experiment about what has happened in earlier rounds. This method was applied in earlier team game experiments by Bornstein and Ben-Yossef (1994).

To have a low variation in neighborhood composition and to facilitate the testing of the segregation hypothesis, only simple neighborhood structures were applied in the experiments. Every experiment started with a control condition, in which subjects had to make their decisions in isolation and without the knowledge of their group membership. After the control condition, color labels were introduced and subjects were seated due to three structural patterns that are shown on Figure 3.3.1. These structural conditions varied in between the experiments. In all structures, subjects had at most two neighbors. The structural position of the groups was completely symmetrical in each case.

Seating patterns were varied between experimental sessions. Visibility conditions were varied and direct social incentives were introduced within every experiment. The latter was done in order to map neighborhood effects and to provide a meaningful comparison for the relative size of the effect of internalized social incentives. The importance of internalized social incentives varies across individuals. We are only

[2] The exceptions are the following situations. A member of the red group is better off with contribution, if without this player there are {2;0}, {2;1}, {2;2}, {3;3}, or {4;4} contributors in the red and in the green group, respectively.

able to demonstrate the relative importance of internalized social control on average and the extent of variation between subjects.

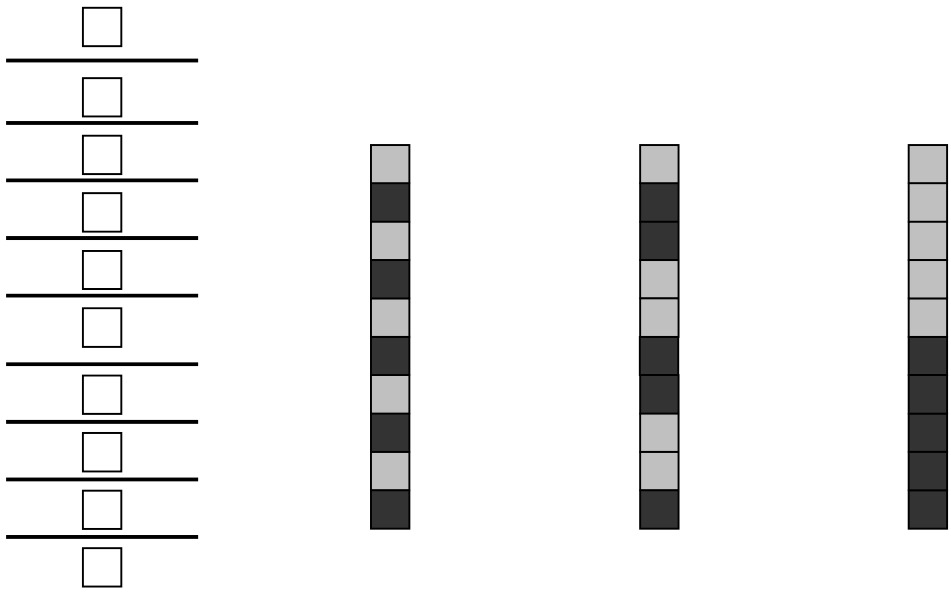


Figure 3.3.1 Structural conditions in the experiment: control condition, low, medium, and high clustering

Notes: Light and dark colors indicate seats of the members of the red and green group. In the control condition, no color labels were introduced.

Social control of the network neighborhood is predicted to be strongest in the presence of direct monetary rewards. Between experimental sessions, two conditions were applied. Next to the payoffs that were present in the beginning of the experiments (see Table 3.3.1), in the *confirmation pressure* condition direct behavioral confirmation incentives (b=5 NLG), in the *normative pressure* condition direct selective incentives (s=5 NLG) were introduced at a certain point in the experiment. Later, in both conditions the other type of incentives was also introduced. Subjects received 5 NLG behavioral confirmation reward if one of their fellow neighbors chose the same action as they did and received 10 NLG if two of their fellow neighbors acted on the same way. Selective incentives were distributed regardless of the decision of neighbors. Contributing subjects received 5 NLG for each fellow neighbor they had. In the low clustering condition there was no change due to the absence of fellow neighbors. This way we had a 2×3 block-design of experimental conditions (see Table 3.3.2).

Table 3.3.2 The number of sessions by experimental conditions

level of clustering	**low**	**medium**	**high**
confirmation pressure (*b first*)	*3*	*4*	*4*
normative pressure (*s first*)	*3*	*4*	*4*

Every experiment consisted of four parts (see Table 3.3.3). All experimental sessions started with an anonymous control condition (Part I) in which subjects were isolated from each other. They did not see each other and did not know which group they belong to. This part allows us to estimate the baseline contribution rates without eye contact between the subjects.

Table 3.3.3 Overview of experimental parts

Part I	anonymous control condition
Part II	eye contact is established
Part III	one form (*b*/*s*) of social control is introduced in a monetary form
Part IV	the other form (*s*/*b*) of social control is introduced in a monetary form

3.4 The model for explaining contribution propensities

3.4.1 Main effects: internalized social control

In this section, we describe the model we use to test our hypotheses in the IPG game experiments. In the explanatory model of individual contribution propensities we incorporate different forms of social control that are believed to be the underlying mechanisms of the segregation effect on intergroup conflict at the macro level. Besides social control being the focus of our interest, we briefly discuss the influence of personal characteristics that are handled as control variables.

For the analysis of experimental data we will use multilevel logistic regression (Bryk and Raudenbush, 1992; Goldstein, 1995). In our two-level model, single decisions are the lower level observations and characteristics of the subjects, who took these decisions, are the group level observations. The two-level model allows us to correct for the methodological problem that observations within the subjects are not independent. For the binary dependent variable, we will use the logit transformation. Formally, let the function P_{ri} denote the propensity of actor i to cooperate in the rth single-shot game. The propensity of cooperation is specified by the logit link function (Goldstein 1995: Chapter 7), which is the natural logarithm of the quotient of the probability of contribution $P_{ri}(C)$ and the probability of defection $P_{ri}(D)$.

$$P_{ri}^{I} = \ln\left(\frac{P_{ri}(C)}{P_{ri}(D)}\right) = \alpha_0 + \varepsilon_i + \xi_{ri}, \qquad (3.4.1.1)$$

where α_0 is the baseline contribution propensity. From previous experiments we know that α_0 depends primarily on the payoff parameters, on the criticalness of the decision (which is a function of group size and the threshold of collective action) and on experimental conditions of confidentiality and anonymity (whether subjects know each other, whether they are in the same room, etc.). Notation ε_i stands for a subject level error term and ξ_{ri} is intra-individual variation. The latter term represents the residual variance that is not estimated in models that include the random intercept α_0. We assume that the subject level error has a zero expected value and has a normal distribution, that is

$$\varepsilon_i \sim N(0, \sigma^2),$$

where the variance σ^2 is going to be estimated. This baseline model does not contain any explanatory variables and allows us to model behavior in the anonymous control condition (Part I).

The interindividual variation of contribution propensities depends on personal characteristics, like gender, study direction, experience in similar experiments, attitudes towards risk, or social orientations. We will include personal characteristics in the analysis as control variables. They enrich research with interesting insight and comparisons can be made with previous findings. They might also interact with internalization of social control. These effects will be discussed in the next section.

Intra-individual variation results from experimental manipulations and also because subjects might reconsider their expectations about the behavior of others. These main factors are relevant after the introduction of minimal contact in Part II. Additional reasons for such variation that can already be present in the control condition are stochastic individual decisions, consideration of mixed strategies, or simply inconsistency. In our simplest model, we assume that intra-individual variation is not correlated with the round number r and has a zero expected value. However, we have to modify this assumption and add a trend element if we find indications of learning the structure of the game through the experiment.

With the introduction of minimal contact (Part II), we predict that internalized social incentives affect individual decisions. The propensity of cooperation will also be dependent on internalized selective incentives, behavioral confirmation, and traitor rewards, as expressed by equation (3.4.1.2):

$$P_{ri}^{II} = \alpha_0 + s_0 f_i + b_0\left(\hat{f}_{cri} - \hat{f}_{dri}\right) - \left(t^+ + t^-\right)g_i + \varepsilon_i + \xi_{ri}. \qquad (3.4.1.2)$$

To recall, notation f_i stands for the number of fellow neighbors of i and g_i for the number of neighbors from the opposite group. Parameter s_0 denotes internalized selective incentives (prestige, respect, etc.) and b_0 stands for internalized behavioral confirmation rewards. We need to estimate these parameters from the experimental results. The expression after b_0 within the brackets denotes the difference between the expected number of contributing and defecting fellow neighbors in decision round r. If the latter is higher, behavioral confirmation decreases the likelihood of contribution. We used a direct way to estimate the expectations of subjects about the number of contributing and non-contributing neighbors. Subjects were asked to forecast the decision of their left and right neighbors before every decision round.

The number of neighbors from the opposite group affects contribution rates through traitor rewards. These are positive (t^+) and negative (t^-) selective incentives rewarding defection and punishing contribution.[3] For the sake of simplicity, we denote their sum by t_0. In the simplest model, we only estimate the average individual importance of internalized social control, therefore the subscript *i* is omitted for estimates s_0, b_0, and t_0. However, in part of the analysis, we will consider that there is a random variance in the size of these effects. Specifically, we will assume that the effects of internalized social control for the subjects are normally distributed around their means. This is consistent with the statement that individuals do not assign the same relative utility for social control, but the utilities are scattered normally around a certain mean evaluation. In this part of the analysis, we will estimate the variances of the influences of different forms of internalized social control, as well as their covariances.

Finally, when explicit monetary payoffs of social control (*s* and *b*) are introduced (Parts III and IV), we have to extend equation (3.4.1.2) with additional terms. The simple model in its general form can be written as

$$P_{ri} = \alpha_0 + \left(s_0 f_i + b_0\left(\hat{f}_{cri} - \hat{f}_{dri}\right) - t_0 g_i\right)p^{II} + s_1 f_i p^s + b_1\left(\hat{f}_{cri} - \hat{f}_{dri}\right)p^b + \varepsilon_i + \xi_{ri}. \tag{3.4.1.3}$$

The p^j dummies denote experimental parts: p^{II} indicates whether or not minimal contact is present, p^s denotes whether or not direct selective incentives are introduced, and p^b shows the presence of direct confirmation rewards. We have to emphasize that s_1 and b_1 in equation (3.4.1.3) are parameters we need to estimate and therefore we distinguish them from the monetary values *s* and *b*. The utility of these monetary rewards might differ between subjects, therefore, in part of the multilevel analysis we will assume a normal distribution of the estimates over the subjects.

3.4.2 Control variables and interaction effects

Instead of including individual effects as dummies, it is more meaningful to analyze what kind of personal characteristics have an influence on contribution propensities. We do not derive hypotheses about the effects of these variables, but we include them as controls because they might have a strong impact on behavior in the experiments.

[3] In Chapter 2, we considered positive traitor incentives only. The consideration of both positive and negative traitor rewards does not add complications to the model.

We know from previous social dilemma experiments that gender is an important predictor of cooperation. However, findings are contradictory about whether women or men are more cooperative (e.g., Isaac, McCue, and Plott, 1985; Mason, Phillips, and Redington, 1991; Frank, Gilovich, and Regan, 1993; Brown-Kruse and Hummels, 1993; Nowell and Tinkler, 1994; Cadsby and Maynes, 1998; Ortmann and Tichy, 1999). Most subjects participating in experiments are students at different faculties of the university. Direction of study might cause individual differences in willingness of contribution. Previous research found that economists have lower contribution rates (Marwell and Ames, 1981; Carter and Irons, 1991; Frank, Gilovich, and Regan, 1993), although there are also experiments that do not find this effect (Isaac, McCue, and Plott, 1985; for an overview, see Ledyard, 1995: 161, 179).

Besides these background variables, we control for attitude measures that indicate special forms of individual utility functions. Previous findings show that attitudes towards risk correlate with contribution propensities (Suleiman and Or-Chen, 1999). Since the contribution decision involves the possibility of a higher reward, but also involves the risk of losing the bonus completely, subjects with a risk-seeking attitude might have higher contribution rates (Budescu, Rapoport, and Suleiman, 1990). On the other hand, there are arguments that in repeated social dilemmas risk aversion increases cooperation (Raub and Snijders, 1997; van Assen and Snijders, 2002). In our experiments, we use attitudes towards risk only as control variables. For the measurement of risk preferences, questions with preference comparisons (see Farquhar, 1984) were used (see Appendix 3.A.2).

Utility functions can also include altruistic elements, which certainly influences rational decision making in social dilemma experiments (e.g., Liebrand, 1984; Doi, 1994). Subjects, who order positive utilities for the gains of others, behave differently from individualistic ones. For the approximation of such utilities, standard questions regarding social orientations were used. They consisted of a series of decomposed games with an unknown person.[4] We presumed that individuals are only *prosocial* (cooperative), *individualistic*, or *competitive*. Previous research found only these types relevant in describing human behavior (van Lange et al., 1997; van Lange, 1999; Suleiman and Or-Chen, 1999). Among each type we distinguished an egalitarian tendency (cf. van Lange, 1999). Although in a two-person PD game or in a public good experiment we should predict higher contribution rates from prosocial subjects, it is not so evident in the IPG game. One could argue that subjects who order utility weights for rewards of unknown others, would do this equally for everyone, including out-group members. Consequently, their contribution rates would not be different from individualistic subjects. A counter-argument is that prosocial (and also egalitarian) orientation is associated with high utility for *social identity*, which is obtainable in a relational comparison with the out-group. Hence prosocial orientation is primarily directed towards in-group members.

[4] The exact questions can be found in Appendix 3.A.1.

In part of the analysis, we also included interaction effects of background variables and social control, because the relative size of internalized social control in the utility function might depend on certain personal characteristics. There are contradictory findings in previous experiments about whether people are more likely to think of others of the same sex to be contributors and in general, whether men or women are more likely to be thought of as better contributors (Ortmann and Tichy, 1999; Solnick and Schweitzer, 1999). For explorative reasons, we control for interactions between gender and social control. We also include as controls interactions between social orientations and the relative weight of social control.

Experiments were combined with repeated IPG games. Repeated games followed single-shot games in all four experimental parts. We designed the experiments so as to exclude possible influences of previous decisions. Subjects were explicitly told before every part that previous parts and repeated games are completely independent from the next part. We predicted no history effects on single-shot decisions, but as a test of this hypothesis, we included previous outcomes of repeated games as control variables in part of the analysis.[5]

3.5 Method

3.5.1 Subjects

203 subjects took part in the experiments at the University of Groningen in May 2000. Subjects were recruited via e-mail and board advertisements promising monetary rewards for participation in a “team competition” experiment. All 203 subjects completed the decision tasks and only two failed to complete the post-decision questionnaire. Altogether, 21 sessions took place and subjects made 4060 decisions (20 each).

114 (56.2%) subjects were female. 187 (92.1%) subjects were university students at the time of the experiments and 16 had already graduated. Students came from all faculties of the university: 55 studied behavioral or social sciences, 47 subscribed for literary studies or art, 26 studied natural sciences, 17 studied law, 13 studied economics, 10 were students at the business faculty, there were 8 students of medical science, 8 subjects studied spatial sciences, and one subject read philosophy. Because of similarities and for the sake of simplicity, economic, business, and spatial sciences were merged in the analysis (furthermore, these faculties have the same physical location) and the student of philosophy was allocated to the category of literary studies and art. The study direction of two subjects was unknown.

Single-shot games (only the decision rounds) took approximately three minutes in each experimental part. During this time subjects had to make five decisions. Counting

[5] In the repeated games, subjects were informed about the result of the previous round.

all four parts of the experiment, every subject played in 20 single-shot games. The entire experiment was on average 80 minutes long.

The payoff for subjects was contingent on their decisions, as well as on the decisions of other participants of the session. Individual payoffs were calculated on the basis of outcomes in the single-shot and in the repeated games. The weighted average of the payoffs obtained in the decision rounds was paid to the subjects after the experiment in sealed envelopes. Payoffs from every round of the repeated games had a single weight. From the single-shot games, only one was selected randomly in each experimental part to be included in the calculation. This payoff had a weight of five rounds (the number of single-shot games in one experimental part). Payoffs varied between 14 and 32 Dutch guilders with an average of 21.1 NLG. Reserve subjects received 10 NLG for their appearance. If subjects ran out of decision time, a random decision was implemented. For all such cases, final payment was decreased by 1%. Out of 4060 decisions this happened only 26 times (0.64%). Random decisions are not included in the analysis.

3.5.2 Procedure

Experiments were conducted in the same computer laboratory.[6] Upon arrival, subjects were randomly seated at a computer. Removable walls separated the subjects to ensure their privacy. Subjects received instructions on paper and on their screen. After reading the instructions they were allowed to ask the experimenter questions. After the questions had been answered, subjects were not allowed to talk. All participants strictly adhered to the rules. After the questions, an examination of understanding followed.

Subjects understood the task quite well, on average they answered 16.5 questions in five minutes, from which 13.7 (83%) were right. The mean proportion of correct answers was 80% with a standard deviation of 18%. Only nine subjects gave more wrong than right answers, one of the nine gave answers that could be considered random. One subject did not answer any of the quiz questions.

Every experiment consisted of four parts. In each part, subjects played five rounds of single-shot IPG games, followed by a number of repeated games. Experiments began with an anonymous start (Part I). Group membership was not yet announced. In every decision round, subjects had to decide whether they would keep the 11 NLG bonus or give it to help their group to achieve success in the competition. These two options appeared in a randomized order on their screen. The bonus was represented also graphically as a bag of money. Subjects were assured of the anonymity of their decisions and that they would receive any money they earned during the experiment in sealed envelopes, after the experiments had ended. In the single-shot games, it was announced that every decision counts towards the final payment, but that only one game of each part would be chosen randomly for payment.

[6] The computer program for the experiment was written by Sicco Strampel in Delphi 5.

In the beginning of Part II, separator walls were removed and group membership was made public by the experimenter. Red and green flags were attached to the monitors and subjects also received an A-4 colored paper with the color of their group. In each condition, subjects were seated behind computers due to the neighborhood configuration of the given session. Subjects were explicitly asked to look around in the laboratory and then played five rounds of the same IPG game again. Before every decision in Part II, III, and IV, subjects had to give their expectations about the subsequent decision of their neighbors. The five single-shot games were followed by repeated games.

In Part III, monetary payoffs for social control were introduced explicitly. In 10 sessions (see Table 3.3.2) monetary confirmation rewards and in 11 sessions monetary selective incentives were incorporated. In Part IV, the other type of social control was also introduced in an explicit monetary form (see Table 3.3.3). As in the low clustering condition there were no fellow neighbors, this condition was used to control for independent learning effects (there was no change between Part II, III, and IV).

Calculation and announcement of the individual results followed the experiment. Meanwhile subjects were asked to fill in a questionnaire on their computer. Monetary payments were supplied in sealed envelopes. The first subject, who had completed the questionnaire, could go immediately to the experimenter to receive payment. Other subjects had to wait until they got a signal from the server. Hence, subjects left the laboratory individually, with a short time difference between their departure. They were informed about the aim of the study after the experiments.

3.5.3 Difficulties

The intended number of participants was ten in all the 21 experimental sessions. On average, thirteen subjects were invited to the sessions as it was anticipated that some would not come. Four sessions failed to be completely filled. In these cases, computer players were included in the teams. Subjects were told that they are programmed in a way to resemble human behavior. In fact, they were simple programs playing mixed strategies with condition-dependent probabilities of contribution. Human decisions in the incomplete experiments are also included in the analysis, but computer decisions are excluded. The inclusion of simulated participants did not have a significant influence on the behavior of subjects in the IPG games.[7]

Additional complications occurred due to a computer failure in three experimental sessions. In these cases, the given subject was seated further away from the others at another computer. No neighbors were defined for these subjects. These complications did not cause distortions in the data.

[7] This variable is not significant when added to any of the multivariate models discussed in Sections 3.6 and 5.6.

Some of the participants knew each other. As acquaintances might influence actual decisions in the experiment, we included the number of acquaintances in the experiment as a control variable in the analysis. Not only the number, but also the exact seating location of acquaintances can be a relevant factor. However, inclusion of such variables would add substantial complexity to the explanatory model without much theoretically grounded gains.

3.6 Results

3.6.1 Contribution rates under different experimental conditions

Let us first provide an overview of the general results of the experiments. Altogether 420 single-shot games were played (20 decision rounds in each of the 21 sessions). There was no tendency towards equilibria as only one outcome was {0;0}, three were {3;0}, and five were {0;3}. In the control condition, in which no minimal contact was introduced and group membership was not announced, there were no {0;0} or {3;0} outcomes and {0;3} occurred only twice (out of 105 outcomes).

Table 3.6.1.1 Outcomes by clustering conditions in all single-shot games

clustering condition in the experiment	*outcome of the decision round* peace	conflict	Total
control condition	26.97% (271)	73.03% (734)	100% (1005)
low clustering	50.23% (428)	49.77% (424)	100% (852)
medium clustering	13.75% (160)	86.25% (1004)	100% (1164)
high clustering	11.85% (120)	88.15% (893)	100% (1013)
Total N	24.27% (979)	75.73% (3055)	100% (4034)

Notes: Cases in parentheses are weighted (multiplied) by the number of human decisions in the given game. For the χ^2 –test unweighted outcomes are used, N=420.

As a consequence of dyadic social control, we predicted different outcomes by clustering conditions. The segregation hypothesis predicted that conflict will be least likely in the low clustering condition and will be most likely in the high clustering condition. Table 3.6.1.1 summarizes the experimental outcomes by clustering conditions. The hypothesis that the outcomes of the IPG game are independent of clustering conditions can be rejected ($\chi^2(3)=46.370$, $p<0.001$).

Table 3.6.1.1 shows that conflict was already quite likely the outcome in the control condition. It indicates that many subjects had a relatively high baseline contribution propensity. Conflict was much less likely in the low clustering condition, and occurred most often in the high clustering condition, which supports the segregation hypothesis. On the other hand, conflict was almost as likely in the medium clustering condition as in high clustering. Conflict occurred in 85.83% of the cases in the medium and 88.57%

of the cases in the high clustering condition (from unweighted outcomes; t=0.613, two-tailed p=0.541), which is counter to the segregation hypothesis.

Contribution rates by clustering conditions are summarized in Table 3.6.1.2. The hypothesis that contribution rates are the same across different clustering conditions can be rejected (ANOVA $F(3, 4030)=52.629$, $p<0.001$). This difference is the result of internalized *and* direct social control. In order to test whether internalized social control can alone cause such differences between clustering conditions, we compare the results from Parts I and II. The comparison reveals that eye contact made an increase in contribution rates. The difference is significant at the 5% level, but not at the 1% level (t=1.722, one-tailed p=0.043). However, in Part II, the contribution rate was highest in the medium clustering condition, which contradicts the segregation hypothesis. The hypothesis that contribution rates are the same in Part II cannot be rejected (ANOVA $F(3, 2003)=1.005$, p=0.366, not significant). Since there is an increase in contribution propensities in the medium and high clustering conditions compared to the control condition, social control might exist also between fellow neighbors in an internalized form. In order to get a better indication of whether or not internalized social control is activated when eye contact is established between subjects, we will use a multivariate analysis and estimate parameter values of the model expressed in equation (3.4.1.3).

Table 3.6.1.2 Average contribution rates (%) in different clustering conditions and parts of the experiment

	incentives introduced first	**low clustering**	**medium clustering**	**high clustering**	**Total**
*Part I**		49.64 (280)	51.81 (386)	46.61 (339)	49.45 (1005)
Part II		50.35 (282)	55.84 (385)	52.84 (335)	53.29 (1002)
P III	*b* (*confirmation pr.*)	-	58.42 (190)	47.33 (150)	53.53 (340)
	s (*normative press.*)	-	63.82 (199)	75.66 (189)	69.59 (388)
	Part III total	40.35 (285)	61.18 (389)	63.13 (339)	55.97 (1013)
P IV	*b* (*confirmation pr.*)	-	62.63 (190)	68.00 (150)	65.00 (340)
	s (*normative press.*)	-	71.00 (200)	81.48 (189)	76.09 (389)
	Part IV total	25.96 (285)	66.92 (390)	75.52 (339)	58.28 (1014)
Total (without Part I)		38.85 (852)	61.34 (1164)	63.87 (1013)	55.86 (3029)
Total		41.52 (1132)	58.97 (1550)	59.54 (1352)	54.26 (4034)

Notes: The number of cell-relevant cases is in parentheses. All human decisions are included.

* In Part I, subjects did not know their group membership and they did not see each other. Therefore their assignment to the different clustering conditions only illustrates baseline contribution rates in the different experimental sessions.

Table 3.6.1.2 also shows average contribution rates in Parts III and IV of the experiment. The hypothesis that contribution rates are the same in the different parts can be rejected (ANOVA $F(3, 4030)=5.869$, p=0.001). We also tested whether the introduction of different direct social incentives in Part III and IV made a difference in contribution rates or not. The hypothesis that contribution rates are the same in the different conditions can be rejected both in Part III (ANOVA $F(2, 1010)=30.800$, $p<0.001$) and in Part IV (ANOVA $F(2, 1011)=108.721$, $p<0.001$). Our hypothesis in Section 3.2 predicted that the introduction of monetary selective incentives would result in higher contribution rates than when behavioral confirmation is introduced in Part III ($t=4.487$, one-tailed $p<0.001$). *Results confirm this hypothesis*. Furthermore, earlier introduction of normative pressure made a difference also in Part IV ($t=3.285$, two-tailed $p=0.001$). This result indicates that history effects still play a role in determining individual decision, despite the lack of feedback regarding the results of single-shot games. We will return to this point at the multilevel analysis of results. Furthermore, figures in Table 3.6.1.2 also *support the hypothesis that under normative pressure the effect of segregation is stronger than under confirmation pressure*. In Part III, under normative pressure average contribution rates are higher in the high clustering condition (75.66%) than in medium clustering (63.82%). On the other hand, under confirmation pressure average contribution rates are higher in the medium clustering condition (58.42% vs. 47.33%).

3.6.2 Questionnaire data

Questionnaire data can only provide partial support for the results. Most subjects reported that they felt important differences between experimental parts, but we are not able to establish the mechanisms behind these differences from the answers. The analysis of actual behavior more convincingly shows the important factors that affect individual decisions.

Table 3.6.2.1 reports some descriptive statistics of relevant questions. The question on the importance of minimal contact (first column) was formulated as: “In the beginning of the experiment walls were separating you and your neighbor(s). Later the walls were removed and flags were attached to the monitors to indicate the group membership of participants. When you had to make your decisions, how important was this difference for you?” Most subjects answered that it was a rather important difference, but there were also many participants for whom this was not important.

Other questions reported in Table 3.6.2.1 were formulated similarly, but they were related to the difference between Part II and Part III and to the difference between Part III and Part IV, respectively. We display descriptive statistics about the subjective importance of these differences separately for sessions, in which direct selective incentives were introduced first (second and fourth columns) and for those, in which direct behavioral confirmation was introduced first (third and fifth columns). Only answers of those subjects who had a fellow neighbor are reported as only they received these new incentives. The hypothesis that subjects felt one form of social control more important at the first introduction can be rejected (Wilcoxon rank test $W=3721.5$, two-tailed $p=0.135$). Furthermore, the difference between these sessions is not significant with regard to changes made before Part IV (Wilcoxon $W=3918.5$, $p=0.598$). This is a somewhat surprising result, since contribution rates differed significantly between these sessions both in Part III and in Part IV (cf. Table 3.6.1.2), which implies that subjects' opinions often do not match their behavior.

On the other hand, the difference between the mean rank and the distribution of those who had these new incentives and those who had no new incentives is highly significant (Wilcoxon W=9332.5, p<0.001; Kolmogorov-Smirnov Z=5.023, p<0.001 for the first introduction and Wilcoxon W=9151.0, p<0.001; Kolmogorov-Smirnov Z=5.126, p<0.001 for the second introduction).

Table 3.6.2.1 Frequencies of answer categories of questionnaire data

questions	*importance of eye contact*	*importance of selective incentives in Part III*	*importance of behavioral confirmation in Part III*	*importance of behavioral confirmation in Part IV*	*importance of selective incentives in Part IV*
very import. (1)	28	13	15	22	22
important (2)	67	30	30	30	25
neutral (3)	42	18	12	12	9
not important (4)	40	8	2	5	3
not imp. at all (5)	24	1	2	1	2
total answers	201	70	61	70	61
mean	2.83	2.34	2.11	2.04	1.98
st. deviation	1.24	0.96	0.93	0.95	1.01
median	3	2	2	2	2
mode	2	2	2	2	2

Other questions were related to the main intentions of subjects during decision making in the experiment. These answers, however, were very much influenced by the repeated games; therefore it would not be adequate to report them here.

3.6.3 Analysis of contribution propensities: a simple model

To understand the underlying mechanisms of the segregation effect on intergroup conflict, we analyzed individual decisions. The first model in Table 3.6.3.1 reports results for the two-level model expressed in equation (3.4.1.3). For the binary dependent variable of individual contribution, the logit transformation was used. The second model assumes that estimates of social control over subjects are normally distributed around their mean. In this model the variances and covariances are estimated as random effects. All human decisions except 23 cases (0.006%) are included. In these 23 cases subjects did not present any expectations about the behavior of their neighbors. In total, 4011 decisions are included in the analysis for 203 subjects.

The two models provided similar estimates. All effects are in the predicted direction. Hypotheses about the existence of internalized behavioral confirmation and internalized traitor rewards are confirmed. The effect of the third type of internalized social control (selective incentives) was not significant. As predicted, both forms of direct social control had a significant effect.

Table 3.6.3.1.a. Results of multilevel logistic regression on contribution propensities

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of main effects	multilevel model assuming random slopes of social control effects
FIXED EFFECTS			
α baseline contribution propensity	?	-.038 (.082)	-.037 (.082)
s_0 internalized selective incentives	+	.109 (.072)	.117 (.072)
s direct selective incentives	+	.407*** (.088)	.363*** (.104)
b_0 internalized behavioral confirmation	+	.617*** (.065)	.640*** (.077)
b direct behavioral confirmation	+	.619*** (.104)	.615*** (.118)
t_0 internalized traitor rewards	-	-.175** (.055)	-.173** (.057)
RANDOM EFFECTS			
interindividual variance σ^2		.616^{+++} (.085)	.628^{+++} (.121)
σ^2_{ui} (s_0)			.000 (.000)
σ^2_{ui} (s)			.300^{++} (.139)
σ^2_{ui} (b_0)			.196^{+++} (.093)
σ^2_{ui} (b)			.326^{+++} (.226)
σ^2_{ui} (t_0)			.009 (.050)
Covariances are reported below			+
-2 Log Likelihood model		4480	4430
Improvement χ^2 (df in parentheses)		939*** (5)$^{\#}$	50*** (20)

Table 3.6.3.1.b. Random effects: estimated covariances

σ_{uxy}	ε_i	s_0	s	b_0	b
s_0	.000 (.000)				
s	-.252 (.108)	.000 (.000)			
b_0	.147 (.083)	.000 (.000)	-.194 (.085)		
b	-.359^{++} (.131)	.000 (.000)	.128 (.132)	-.079 (.116)	
t_0	-.005 (.072)	.000 (.000)	.425 (.153)	-.169 (.109)	.176 (.165)

Notes: N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors. ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).

For testing random effects it is more appropriate to use deviance tests: $^{++}$ significant at the 1% level, $^{+++}$significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates).

$^{\#}$Basis of comparison: baseline multilevel logistic regression expressed in equation (3.4.1.1); α: 0.174** (0.066); σ^2: 0.674^{+++} (0.087).

Internalized behavioral confirmation had a very strong effect, approximately as strong as 5 NLG monetary incentives for behavioral confirmation. The parameter estimate of internalized behavioral confirmation 0.617 means that an additional fellow neighbor, who is expected to contribute, increases the odds by 85.34%. For instance, in the first model, for an average subject, the predicted likelihood of contribution is 49.05% in the control condition. In the minimal contact condition, the expected likelihood of her contribution, if she were to have one contributing fellow neighbor and no opposite neighbor is 66.55%. For this increase, internalized behavioral confirmation takes the most responsibility.

For testing hypotheses about random effects it is more appropriate to use deviance tests than the *t*-test (cf. van Duijn, van Busschbach, and Snijders, 1999: 192-193). Baseline contribution rates between subjects had a high unexplained variance. The influence of behavioral confirmation and monetary selective incentives varied significantly between subjects. The hypothesis that the sizes of traitor rewards and internalized selective incentives are the same for the subjects cannot be rejected. High positive deviations from the average baseline contribution rate were correlated with negative deviations from the average importance of monetary rewards for confirmation. This is not surprising given that our general prediction was that subjects, who evaluate monetary gains less, would contribute more to the success of their group.

3.6.4 The effect of personal characteristics

To see which personal characteristics are responsible for the high inter-individual variation, in the next part of the analysis we extended the model by controlling for background variables and certain attitude measures. Again, we conducted two analyses: one assuming fixed social control effects without random variation and another assuming a random variation and covariation of these estimates (see Table 3.6.4.1). The inclusion of control variables did not rule out the significant effects of social control variables. The effect of internalized selective incentives remained insignificant.

There was no significant gender effect, although simple descriptive statistics showed that women had higher contribution propensities (55.94%) than men (52.14%). Based also on descriptive statistics, subjects who already graduated were more contributive (61.54%) than students (53.58%). This effect was not significant in the model, as it was ruled out by other variables, mainly by social orientations. The analysis of study direction did not reveal an effect of economics training. In our experiments, students of natural sciences and law had the lowest contribution rates (48.17% and 48.66%, respectively). They were followed by students of economics (51.54%) and medical sciences (53.55%; this was used as a reference category in models of Table 3.6.4.1). Subjects who studied literary studies or art and students of social sciences had the highest contribution rates (56.91% and 55.85%, respectively).

However, the effect of study direction was also ruled out by other variables. The argument that experience matters at all is questioned by the insignificant effect of participating in a similar experiment before. Again, we could probably explain the difference in descriptive statistics (56.14% vs. 51.44%) by selection on attitude measures.

We characterized subjects as strongly risk-averse, if they chose for risk-averse alternatives both in simple and complex gambles. We found that 91 subjects (45.3%) were strongly risk-averse towards gains, 92 (45.8%) were strongly risk-averse towards mixed gambles, and 83 (39.5%) were strongly risk-seeking towards losses. Although we did not try to estimate utilities more precisely, descriptive statistics confirm that most individuals have an S-shape utility function (e.g., Tversky and Kahneman, 1992). Effects of risk-aversion and loss-aversion are not significant in the models.

The only personal characteristics that we measured and found highly significant in explaining contribution propensities were *social orientations*. For questions about social orientations, 77 (37.9%) subjects gave inconsistent answers. Inconsistency was a significant predictor of contribution rates, which is probably related to the relevance of calculation abilities. Among subjects, who gave consistent answers, 76 (61.3%) were prosocial, which is higher than in previous experiments (for an overview see Schulz and May, 1989). As an exception, Liebrand (1984) found a similar high rate in his experiments conducted in Groningen. Our results clearly support the argument that prosocial (and also egalitarian) orientation is primarily directed towards in-group members and therefore increases contribution rates in the IPG game. The strong effects also indicate that social orientations are important predictors of behavior in intergroup situations. Individuals with prosocial and egalitarian attitudes seem to be responsible for the emergence of mutually harmful outcomes.

Table 3.6.4.1.a. Results of multilevel logistic regression on contribution propensities with personal characteristics

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of social control	multilevel model with random slopes of social control
FIXED EFFECTS			
Main variables			
α (constant) baseline contribution propensity	?	.568 (.402)	.736 (.389)
s_0 internalized selective incentives	+	.084 (.073)	.112 (.073)
s direct selective incentives	+	.410*** (.089)	.335** (.105)
b_0 internalized behavioral confirmation	+	.628*** (.066)	.644*** (.080)
b direct behavioral confirmation	+	.635*** (.105)	.604*** (.118)
t_0 internalized traitor rewards	-	-.159** (.056)	-.154** (.055)

cont. next page

Personal characteristics and other subject-level variables		
gender (1=male)	-.201 (.139)	-.221 (.134)
student at the university (1=yes)	-.268 (.360)	-.393 (.346)
studies at the law faculty	-.094 (.356)	-.004 (.341)
studies natural sciences	-.016 (.335)	-.030 (.320)
studies economic, business, or spatial science	-.003 (.326)	.095 (.311)
studies social sciences	.118 (.301)	.180 (.287)
student of literary studies or arts	.124 (.308)	.202 (.294)
did a similar experiment before	-.159 (.132)	-.190 (.127)
strong risk aversion towards gains	-.141 (.131)	-.154 (.126)
strong loss aversion	.117 (.130)	.153 (.125)
consistent answers on social orientation questions	-.359* (.176)	-.397* (.169)
prosocial orientation	.479** (.178)	.485** (.171)
egalitarian orientation	.382* (.172)	.386* (.165)
number of acquainted subjects in the experiment	-.089 (.086)	-.095 (.083)
delay (minutes) at the start of the experiment	.011 (.007)	.010 (.007)
quiz questions answered correctly %	-.005 (.004)	-.006 (.004)
RANDOM EFFECTS		
interindividual variance σ^2	$.537^{+++}$ (.078)	$.554^{+++}$ (.115)
σ^2_{ui} (s_0)		.000 (.000)
σ^2_{ui} (s)		$.311^{++}$ (.143)
σ^2_{ui} (b_0)		$.225^{+++}$ (.099)
σ^2_{ui} (b)		$.296^{+++}$ (.221)
σ^2_{ui} (t_0)		.002 (.048)
Covariances are reported below		+
-2 Log Likelihood model	4430	4382
Improvement χ^2 (df) for model in right column		48*** (20)
vs. previous model	50*** (16)	48*** (16)

Table 3.6.4.1.b. Random effects: estimated covariances

σ_{uxy}	ε_i	s_0	s	b_0	b
s_0	.000 (.000)				
s	-.195 (.106)	.000 (.000)			
b_0	.037 (.084)	.000 (.000)	-.213 (.089)		
b	$-.245^{+}$ (.126)	.000 (.000)	.146 (.131)	-.083 (.118)	
t_0	-.035 (.070)	.000 (.000)	.503 (.160)	-.115 (.113)	.163 (.168)

Notes: N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors. * significant at the 5% level, ** significant at the 1% level, ***significant at the 0.1% level (two-tailed).

For testing random effects it is more appropriate to use deviance tests: $^{++}$ significant at the 1% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates).

There was no significant effect of how many others were acquainted to subjects in the laboratory and delay time at the start of the experiment did not matter for contribution propensities. This result shows that these factors that are related to the experimental environment did not disturb the behavior of subjects.

3.6.5 Learning the structure of the game

Until now we assumed that intra-individual variation (ξ_{ir}) has a zero expected value and it is independent from the decision round r. We based this assumption on the randomized design of single-shot games (Bornstein and Ben-Yossef, 1994). Every decision round could be handled in an equivalent way, since only one of the rounds would be randomly chosen for payment and subjects did not receive information about the outcomes of previous rounds.

Although Bayesian learning effects cannot enter the series of single-shot games, we can probably detect a different form of learning. As experimental time passes, subjects might understand the structure of the game better and can also gather more routine in making decisions. Previous experiments of iterated PD, public good, and IPG games found that subjects approach the all-defection equilibrium over time (Isaac, McCue, and Plott, 1985; Andreoni, 1988; Andreoni and Miller, 1993; Bornstein, Winter, and Goren, 1996; Goren and Bornstein, 2000), which results in decreasing cooperation rates. Besides learning the structure of the game over time, another possible explanation is offered by Kreps et al. (1982). If subjects have an initial belief about others being irrational, it can be better for them to contribute initially, even if they are completely rational. Later on, they update their beliefs and they are then less likely to contribute. Experiments that included a "coincidental" restart of the game, however, found that contribution rates after restart went back up to virtually the same contribution rates as observed in the very first game (Andreoni, 1988). Based on these results Dawes and Thaler (1988) refuted the learning and the belief-updating hypotheses. However, by considering internalized reputation effects and image scoring (Wedekind and Milinski, 2000; Bienenstock, 2001) that can be mistakenly activated in single-shot games, we could explain decreasing contribution rates after every restart of repeated games in combination with the learning hypothesis.

If contribution propensities are not stable in the single-shot games within experimental parts, then we have to include an independent learning effect in the analysis and we have to drop our assumption that intra-individual variation (ξ_{ir}) has a zero expected value. As parts were separated by breaks, instead of checking for a single learning trend, it is better to distinguish between a within part and a between part learning trend in the analysis.

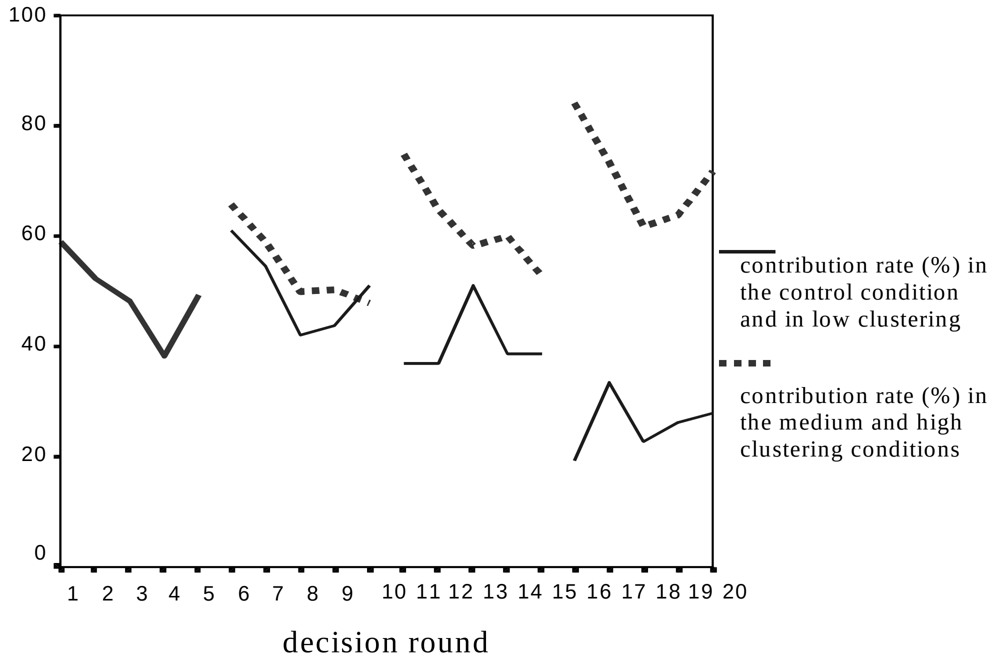


Figure 3.6.5.1 Contribution rates by decision rounds

Note: Part I: rounds 1-5; Part II: rounds 6-10; Part III: rounds 11-15; Part IV: rounds 16-20.

We found a decay of contribution for the series of single-shot games in our experiments (see Figure 3.6.5.1). Contribution rates decreased for those, who had some misunderstanding of the task before the game, but also for those, who answered quiz questions correctly. Besides the decreasing within part trend, in the last round of every part contribution rates increased significantly. It is a quite surprising result, since subjects knew that the outcome of the last round would neither be announced. This is exactly the opposite of what would be predicted on the basis of arguments of traditional game theory even if subjects had the incorrect perception that they are playing repeated games. This *endgame effect* was there independent of gender, social orientations, previous decisions, and other personal characteristics. By analyzing last rounds only, model parameters were similar to those values that were reported in Table 3.6.4.1, including the insignificant effect of internalized selective incentives. It means that higher contribution propensities in the last rounds cannot be explained by the reduction of cognitive dissonance ("in the last round I have to be nice, otherwise I cannot look at my fellow neighbors"). The resulting U-shape trend, however, has some correspondence to experimental findings in the iterated two-person PD and in collective action games (Rapoport and Chammah, 1965; Guttman, 1986). In our experiments, linear trend estimation provides the following parameters for within part learning:

$$P_{ri} = \ln\left(\frac{P_{ri}(C)}{P_{ri}(D)}\right) = 0.516(0.084) - 0.207(0.032)r + 0.376(0.112)r^5 + \varepsilon_i, \quad (3.6.5.1)$$

where r denotes the decision round number within the part, r^5 is a dummy for the last round in each part, and standard errors are in parentheses. Multilevel analysis gives an estimate of 0.693 (S.E.=0.089) for inter-individual variance. All effects are significant at the 0.1% level. The constant shows the average predicted contribution propensity for the first decision. We also tried to fit a non-linear model for within part learning, but goodness of fit did not increase significantly. Hence we included a linear trend element and an endgame effect as control variables in the model explaining contribution propensities (cf. Table 3.6.5.1). Model parameters are quite close to those in equation (3.6.5.1).

For between parts learning let us first report results from the low clustering condition (lower curve on Figure 3.6.5.1). In this condition no direct social control was introduced in between the parts, therefore Parts II, III, and IV were identical. Linear curve estimation provided a parameter of –0.5257 (S.E.=0.089, $p<0.001$) for between parts learning. It means that estimates of contribution rates in the low clustering conditions are 64.09%, 51.34%, 38.41%, and 26.94% in the four different parts, respectively.

We included both types of trend elements in the models in Table 3.6.5.1 as control variables. They were highly significant, as well as the puzzling endgame effect. But learning trends and endgame effects were not the only unexpected procedure effects. After controlling for the result of the last repeated game (if there was any), it emerged that a mutually harmful draw (clash punishment) “burns in” the memory of subjects and increases contribution propensities also in the single-shot games. This is probably an indication of that subjects use their long-term memory to estimate whether or not their decision could make a difference for the outcome in the forthcoming single-shot game. If they believe that a draw will occur, a single individual contribution can turn the outcome to victory. We did not anticipate these disturbances, but as we control for them in the analysis, results show the real effect of the main variables.

There are remarkable changes in the parameter estimates of social control. The effect of internalized selective incentives became significant and the significant effect of internalized traitor rewards has disappeared. The large increase in the estimate of baseline contribution propensity (constant) also indicates that the omission of learning trends resulted in a systematic bias in previous estimates in Table 3.6.3.1 and 3.6.4.1. Because of the negative between parts tendency, the baseline contribution rate was underestimated and the decrease between Part I and Part II was attributed to the effect of internalized traitor rewards. On the basis of the analysis reported in Table 3.6.5.1, we have to conclude that after controlling for a negative learning tendency, on average, traitor incentives in an internalized form do not influence the decision of subjects. On the other hand, we have to be careful with this interpretation and also with the confirmation of the existence of internalized selective incentives. The inclusion of a between parts trend in a linear functional form in the analysis does not stand on a firm theoretical basis. Furthermore, since the high correlation with experimental manipulations (the introduction of minimal contact and monetary forms of social control), the learning effect might include part of influence that we should attribute to other variables.

Table 3.6.5.1.a. Results of multilevel logistic regression on contribution propensities with personal characteristics and procedure effects

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of main effects	multilevel model with random slopes of main effects
FIXED EFFECTS			
Main variables			
α (constant) baseline contr. propensity	?	1.378** (.423)	1.516*** (.409)
s_0 internalized selective incentives	+	.186* (.082)	.188* (.081)
s direct selective incentives	+	.769*** (.109)	.699*** (.127)
b_0 internalized behavioral confirmation	+	.586*** (.067)	.591*** (.080)
b direct behavioral confirmation	+	.718*** (.108)	.705*** (.126)
t_0 internalized traitor rewards	-	.165 (.086)	.142 (.086)
Personal characteristics and other subject-level variables			
gender (1=male)		-.176 (.143)	-.196 (.137)
student at the university (1=yes)		-.219 (.370)	-.352 (.357)
studies at the law faculty		-.109 (.366)	-.015 (.351)
studies natural sciences		-.057 (.344)	-.065 (.330)
studies economic, business, or spatial science		-.030 (.335)	.095 (.322)
studies social sciences		.068 (.309)	.136 (.296)
student of literary studies or arts		.056 (.316)	.133 (.303)
did a similar experiment before		-.154 (.136)	-.188 (.131)
strong risk aversion towards gains		-.163 (.135)	-.180 (.129)
strong loss aversion		.115 (.134)	.132 (.128)
consistent answers on social orientation questions		-.374* (.181)	-.400* (.173)
prosocial orientation		.511** (.183)	.487** (.175)
egalitarian orientation		.388* (.176)	.392* (.169)
number of acquainted subjects in the experiment		-.079 (.088)	-.093 (.085)
delay (minutes) at the start of the experiment		.008 (.007)	.006 (.007)
quiz questions answered correctly %		-.005 (.004)	-.005 (.004)
Procedure effects			
within part trend		-.215*** (.036)	-.213*** (.036)
endgame effect		.373** (.125)	.370** (.126)
between parts trend		-.397*** (.060)	-.379*** (.061)
last iterated game was a draw		.538*** (.149)	.515*** (.152)
last iterated game was lost		.185 (.122)	.199 (.125)
last iterated game was won		.214 (.123)	.275* (.125)

cont. next page

RANDOM EFFECTS		
interindividual variance σ^2	.574^{+++} (.083)	.559^{+++} (.116)
σ^2_{ui} (s_0)		.000 (.000)
σ^2_{ui} (s)		.322^{+++} (.152)
σ^2_{ui} (b_0)		.202^{+++} (.096)
σ^2_{ui} (b)		.421^{+++} (.246)
σ^2_{ui} (t_0)		.002 (.050)
Covariances are reported below		+
-2 Log Likelihood model	4247	4198
Improvement χ^2 (df) for model in right column		49*** (20)
vs. previous model	183*** (6)	184*** (6)

Table 3.6.5.1.b. Random effects: estimated covariances

σ_{uxy}	ε_i	s_0	s	b_0	b
s_0	.000 (.000)				
s	-.163 (.109)	.000 (.000)			
b_0	.037 (.083)	.000 (.000)	-.192^{+} (.090)		
b	-.287^{+} (.133)	.000 (.000)	.063 (.143)	-.084 (.123)	
t_0	-.018 (.071)	.000 (.000)	.476 (.169)	-.054 (.117)	.152 (.180)

Notes: N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors. * significant at the 5% level, ** significant at the 1% level, ***significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{+}$ significant at the 5% level, $^{+++}$significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates).

There is another complication in relation to the difference in contribution propensities between Parts I and II. *Silent identification* (Bohnet and Frey, 1999) enters social dilemma experiments, when subjects are able to see each other. The visibility of others decreases social distance, allows for empathy and helps to conceptualize the experimental situation. We might predict this to be correlated with utility weights for social identity, consequently with higher contribution propensities. However, we cannot separate this effect from the influence of internalized social incentives that are not contingent on predictions (selective incentives and traitor rewards). If silent identification is a valid mechanism in the IPG game, our analysis overestimates the effect of internalized selective incentives. The unexpected positive sign of the t_0 estimate can also be partly explained by silent identification.

3.6.6 Interaction effects

As we have seen in Table 3.6.5.1, the significant effect of internalized traitor rewards disappeared after the inclusion of learning trends. It might be possible that we wrongly conceptualized this form of social control and traitor rewards could be there, but have another nature. They might stem from the presence of the other group as a whole or they exist only in certain dyadic relations.

The extension of the model by interaction effects helps with some clarification (see Table 3.6.6.1). It seems that internalized traitor rewards are activated in the dyadic context, but not in every neighborhood relation. Only neighbors of the opposite sex provide a significant control in the form of traitor rewards. This indicates that internalized pressure against contribution in the presence of opposite group members is activated only, if substantive distinction can be made apart from minimal group membership. Gender is possibly the most apparent characteristic that can be the source of this distinction between strangers. With respect to the interaction between gender and internalized behavioral confirmation, we did not find a significant effect on contribution propensities. However, descriptive statistics showed that subjects expected contribution more from fellow neighbors of the same sex and additionally, women were expected to contribute more.

Acquainted neighbors did not experience stronger social control than unknown ones did. Similar to the insignificant effect of the number of acquainted subjects in the experiment, this result can be probably attributed to the fact that they were not close acquaintances or to subjects' tendency to view laboratory conditions as impersonal. Prosocial and egalitarian attitudes were not correlated with higher relative weight of internalized social control. Only the interaction between traitor rewards and prosocial orientation proved to be significant. This effect indicates that subjects with prosocial orientation like to be "local heroes", who contribute even if they are surrounded by members of the other group. This is another indication of how *prosocial attitudes can be harmful in the intergroup context*.

Table 3.6.6.1.a. Results of multilevel logistic regression on contribution propensities with personal characteristics, procedure effects, and cross-level interactions

independent variable	hypothesis about the direction of effect	model with fixed slopes	model with random slopes
FIXED EFFECTS			
Main variables			
α (constant) baseline contr. propensity	?	1.346*** (.402)	1.491** (.477)
s_0 internalized selective incentives	+	.176* (.082)	.165* (.084)
s direct selective incentives	+	.769*** (.110)	.745*** (.135)
b_0 internalized behavioral confirmation	+	.589*** (.119)	.618*** (.141)
b direct behavioral confirmation	+	.703*** (.109)	.681*** (.125)
t_0 internalized traitor rewards	-	.223 (.132)	.238 (.134)
Personal characteristics and other subject-level variables			
gender (1=male)		-.089 (.146)	-.135 (.143)
student at the university (1=yes)		-.177 (.372)	-.201 (.364)
studies at the law faculty		-.162 (.368)	-.136 (.360)

cont. next page

studies natural sciences	-.101 (.349)	-.161 (.341)
studies economic, business, or spatial sciences	-.080 (.339)	-.002 (.330)
studies social sciences	-.001 (.312)	.000 (.305)
student of literary studies or arts	.045 (.317)	.066 (.309)
did a similar experiment before	-.179 (.136)	-.221 (.133)
strong risk aversion towards gains	-.172 (.134)	-.157 (.132)
strong loss aversion	.131 (.133)	.164 (.131)
consistent answers on social orientation questions	-.397* (.180)	-.404* (.176)
prosocial orientation	.330 (.206)	.353 (.202)
egalitarian orientation	.419* (.203)	.394* (.200)
number of acquainted subjects in the experiment	-.066 (.089)	-.066 (.087)
delay (minutes) at the start of the experiment	.006 (.007)	.006 (.007)
quiz questions answered correctly %	-.004 (.005)	-.005 (.005)
Procedure effects		
within part trend	-.178 (.121)	-.188 (.122)
endgame effect	.379** (.126)	.381** (.127)
between parts trend	-.397*** (.061)	-.386*** (.062)
last iterated game was a draw	.527*** (.150)	.495** (.157)
last iterated game was lost	.180 (.123)	.186 (.128)
last iterated game was won	.214 (.124)	.266* (.128)
Cross-level interactions		
$t_0 \times$ number of acquainted opposite neighbors	-.153 (.196)	-.164 (.194)
$b_0 \times$ number of acquainted fellow neighbors	.302 (.261)	.338 (.312)
$t_0 \times$ number of opposite neighbors of the other sex	-.351** (.134)	-.373** (.137)
$t_0 \times$ number of male opposite neighbors	.191 (.134)	.156 (.136)
$b_0 \times$ number of fellow neighbors of the same sex	-.038 (.084)	-.128 (.102)
$b_0 \times$ number of female fellow neighbors	.302 (.261)	.017 (.108)
$t_0 \times$ prosocial orientation	.275* (.131)	.256* (.132)
$b_0 \times$ prosocial orientation	.052 (.134)	.098 (.161)
$t_0 \times$ egalitarian orientation	-.057 (.149)	-.025 (.149)
$b_0 \times$ egalitarian orientation	.039 (.143)	.004 (.172)
within part trend $\times$ quiz questions correct %	.000 (.001)	.000 (.001)
RANDOM EFFECTS		
interindividual variance σ^2	$.563^{+++}$ (.082)	$.512^{+++}$ (.084)
σ^2_{ui} (s_0)		.000 (.000)
σ^2_{ui} (s)		$.549^{+++}$ (.187)
σ^2_{ui} (b_0)		$.143^{+++}$ (.089)
σ^2_{ui} (b)		$.379^{+++}$ (.240)
σ^2_{ui} (t_0)		.000 (.000)
Covariances are reported below		+
-2 Log Likelihood model	4211	4169
Improvement χ^2 (df) for model in right column		42** (20)
vs. previous model	36*** (11)	29** (11)

Table 3.6.6.1.b. Random effects: estimated covariances

σ_{uxy}	ε_i	s_0	s	b_0	b
s_0	.000 (.000)				
s	.037 (.107)	.000 (.000)			
b_0	.004 (.072)	.000 (.000)	-.145 (.093)		
b	-.200^{++} (.118)	.000 (.000)	.201 (.152)	-.031 (.116)	
t_0	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)

Notes: N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors. * significant at the 5% level, ** significant at the 1% level, ***significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{++}$ significant at the 1% level, $^{+++}$significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates).

Additionally, we included an interaction variable to test whether or not subjects, who did not fully understand the experimental task, have different learning tendencies. The insignificant effect indicates that learning the structure of the game during the experiment is a general tendency and does not depend on the initial stage of understanding.

3.7 Discussion

The focus of this chapter differed from the mainstream experimental tradition of social dilemmas and attempted to incorporate sociological insights into the explanation of individual contribution rates. Our main objective was to show how internalized and direct social control enter into simple experimental situations and can affect individual decisions in an intergroup competition situation that we modeled by an Intergroup Public Goods game (Rapoport and Bornstein, 1987). As an aggregated result of different forms of social control, we tried to demonstrate *why segregation might induce the emergence of conflict between groups*. To discover the underlying mechanisms, we investigated what is the exact nature of *social control* and what are the forms that are already present in a condition with only eye contact between subjects. In order to test our hypotheses we introduced a *unique experimental design* based on special seating arrangements in the laboratory. With our setup we were able to target social control of the immediate environment that is believed to be influential also in real life.

By comparing clustering conditions we found that intergroup conflict was least likely in a completely mixed setting and was most likely when members of the groups were seated according to a segregated pattern, which confirms the segregation hypothesis. Furthermore, as predicted, the segregation effect was stronger under normative pressure than in the confirmation pressure condition.

By analyzing individual decisions, we uncovered mechanisms of social control that cause the segregation effect at the aggregated level. We found that *behavioral confirmation* is the form of social control, which strongly affects individual contribution propensities, also in an internalized form. Subjects adjusted their decisions towards the expected decision of their fellow neighbors even when eye contact is established among them. Our model parameters indicate that under the chosen reward structure, internalized confirmation pressure affected contribution propensities as much as monetary confirmation incentives did. We did not find, however, strong support for the presence of other forms of internalized social control in the minimal contact condition. Internalized selective incentives continued to have a significant effect after controlling for a between parts trend. However, the net effect of these incentives might diminish, if we consider silent identification (Bohnet and Frey, 1999) as a disturbing factor. Internalized traitor rewards might be activated in a dyad with minimal contact, but it is not a general mechanism. We found its clear presence only between neighbors of the opposite sex. Direct social control that was introduced in a form of additional monetary incentives had a significant effect.

In contradiction to what we predicted in the segregation hypothesis, contribution rates in the minimal contact condition were highest in the medium clustering condition. A possible explanation is that there is a *ceiling effect* similar to what we observed in high ranges of clustering in the simulation results (cf. Figures 2.8.1.2, 2.8.2.1, 2.8.2.2). This explanation is supported by evidence of high likelihood of conflict in the medium clustering condition (cf. Table 3.6.1.1). Another reason might be that the strength of internalized social control is a nonlinear function of the number of fellow neighbors. As a consequence, there is a marginal decrease in the segregation effect on the likelihood of intergroup conflict and already medium levels of clustering are often associated with harmful outcomes.

Among personal characteristics, we found a strong predicting power of social orientations. Subjects with prosocial and egalitarian attitudes found to be more contributive and consequently were found to be also more responsible for the emergence of mutually harmful outcomes between the groups. Another indication of that prosocial orientations are correlated with more generous behavior for the in-group, but more hostile behavior towards the out-group, is the positive interaction effect of traitor rewards and prosocial orientation. This implies that subjects with prosocial orientation behave as local heroes. If they are surrounded by members of the other group, they do not surrender at all. As a macro consequence, mutually harmful outcomes can occur even in the case of complete mixing, if there are enough prosocial individuals.

Although these effects appeared to be quite strong, we have to mention a limitation of the model. Little attention was devoted to individual utility functions and we were only interested in general tendencies. A complex utility assessment, however, is usually beyond consideration because it requires a lot of experimental time and answers are often inconsistent.

The analysis of the repeated IPG games included in these experiments is a target of subsequent chapters. First, in Chapter 4 we introduce theoretical concepts about effects of temporal embeddedness that we predict to influence repeated intergroup competition.

Appendix

3.A.1 Instructions in the experiment

Instructions in the experiment were translated into Dutch. The first pages ("Instructions" and "What do you do in this experiment") were also distributed on paper.

Instructions at the arrival to the laboratory

Team competition: An experiment
Instructions

- After you have listened to the verbal introduction and read the instructions, there will be time to ask questions.
- The instructions on your screen and on the paper are exactly the same.
- Please leave the instructions next to your computer when leaving the room.
- VERY IMPORTANT: Any kind of communication with other participants is *strictly prohibited* during the experiment. We kindly ask you to keep this rule. Violation can result in money withdrawal from your earnings or your exclusion from the experiment.
- The experiment will take approximately 90 minutes and you will receive your payment at the end of the experiment. How much you earn will depend on your decisions and on the decision of other participants. This will be explained in detail on the next page.

Instructions at the start of the experiment

What do you do in this experiment?

This experiment is about competition between groups. There are 10 participants in this experiment; you are one of them. The participants are divided into ***two teams***. Both teams consist of 5 members. Your group membership and the membership of other participants will be announced later in the experiment.

The experiment will take several decision rounds. All decision rounds will count in your final payment! At the end of the experiment, the computer will calculate your AVERAGE earning in the decision rounds. This will be your real money, what you will receive in cash after the experiment has finished. If you want to earn a lot of money, try to gather as much symbolic money as possible!

At the beginning of each decision round, you will receive 11 Dutch guilders as a bonus. The bonus appears on your screen and you have to decide

- whether to give away this bonus to help your team to reach success in the team competition
- or to keep the bonus.

Other participants have to make exactly the same decisions. Decisions are kept strictly confidential. You will not know the decision of others and they will not know your decision.

The team, in which more members gave away their bonuses, wins the competition and the other team loses. For every member in the winning team (irrespective of his or her own decision) the same amount of

15 guilders are distributed. From every member of the losing team (irrespective of his or her own decision) 15 guilders are subtracted. In the case of a draw, 11 guilders are subtracted from every participant (irrespective of his or her own decision). The only exception is when in both teams there are *less, than 3 individuals* who are willing to give away their bonuses. In this case not enough interest is recorded in the team competition and therefore no money is distributed or subtracted.

For your better understanding of how much money you earn in one round please consider the following table:

Table 1 Your earnings (Dutch guilders) in one round of the experiment

	result of the competition	*not enough interest*	*your team wins*	*draw*	*your team loses*
your decision	**give away your bonus to help your team**	0	15	-11	-15
	keep your bonus	11	11+15 = 26	-11+11 = 0	-15+11 = -4

After the experiment, you will receive the average of what you have earned in the different rounds. This amount will be increased by 15 guilders to guarantee you a positive earning regardless of what you do.

Please remember that all team members are real people. The computers are only here to record your decisions. Your decisions will be correctly saved and you will always receive completely true information during the experiment.

Before the real experiment starts, we would like to be certain that you have understood the decision task correctly. You will be kindly asked to answer the quiz on your computer. In this quiz of understanding, the decision of others is generated randomly by the computer. The fictive money you gain in the quiz will not count in your real payment.

Now you can ask your questions.
Good luck!

Instructions at the start of the quiz

As a warming-up, you will receive some questions that help you to understand the decision task correctly.

We would like to be certain that you understood the decision task correctly. For this reason please consider the following possible outcomes. We ask you to select what you think your reward would be in the given situation. Now and also later during the experiment you can get help by looking at Table 1 in the instructions.

Instructions at the start of Part I

The quiz hopefully helped you to understand the decision task correctly. Once again, during the experiment you can always get help by looking at Table 1 in the instructions.

This is the real experiment. What you gain in this round, will count in your score. Be careful, if you run out of time, the computer will decide randomly for you. For each case that this happens there will be 1% decrease in your final payment.

We will also ask you to give some predictions about the outcome of the competition. Please take into consideration, that every prediction that you do not answer, decreases your final payment by 0.2%.

From the first 5 rounds, only one will count in your final payment. The computer will select this round randomly.

We kindly ask you to pay attention to the order of answers, because this changes during the experiment.

We kindly ask you to avoid any form of communication with other participants.

As soon as this information disappears from your screen, the experiment starts.

Instructions at the start of Part II

Announcement of group membership

Now we announce your team membership and the membership of others. The members of the two teams will receive red and green flags. Participants with red flags consist of one team and participants with green flags consist of the other team.

Now the experimenter distributes the flags. Please look around you and see whether your neighbors are from your team or from the other team. We kindly ask you not to stand up from your place and not to talk with other participants. Please do not remove your flag from your computer during the experiment.

You can also see who are your neighbors from your screen. Be aware that your neighbors know in which group you are and they might know what your decision was.

Now we will continue with the experiment. You will receive 11 Dutch guilders as a bonus at the beginning of each decision round. As before, you will be asked to decide

- whether to give away this bonus to help your team to reach success in the team competition
- or to keep the bonus.

Good luck!

Instructions at the start of Part III

Modifications

You are still in the same group as before and your teammates did not change either. At the beginning of the following decision rounds, you will receive 11 Dutch guilders as a bonus, as before. You again have to decide

- whether to give away this bonus to help your team to reach success in the team competition
- or to keep the bonus.

Other participants have to make exactly the same decision. Just to recall how much money you can earn in one round please consider the following table again:

Table 1 Your earnings (Dutch guilders) in one round of the experiment

	result of the competition	*not enough interest*	*your team wins*	*draw*	*your team loses*
your decision	**give away your bonus to help your team**	0	15	-11	-15
	keep your bonus	11	11+15 = 26	-11+11 = 0	-15+11 = -4

This is the same procedure as before, but now we introduce some modifications about the payments.

Please look around whether your neighbor is from your team. A neighbor is a participant, who is sitting next to you on your left and/or on your right. Your neighbor is from your team if there is a flag of the same color on your monitor and on his or her monitor. *If you do not have a neighbor from your team, then there is no modification in the method of payment.*

[for conditions in which behavioral confirmation is introduced]

You will earn 5 guilders extra in every round, if you do the same as your neighbor in the same round.

If you have two neighbors from your team, this rule applies to both of them. It means that you can earn 10 (2x5) guilders if you and both of your neighbors do the same in that round. If the decision of these two neighbors is different, then you earn 5 guilders.

[for conditions in which selective incentives are introduced]

You will earn 5 guilders extra, if you have one neighbor from your team and you decide to give away your bonus. You will earn 10 (2x5) guilders, if you have two neighbors from your team and you decide to give away your bonus.

[for conditions in which selective incentives or behavioral confirmation is introduced]

Everyone, who has neighbors from his or her team, has this opportunity to earn extra money.

Now you can ask questions.

Good luck!

Instructions before Part IV were constructed similarly.

3.A.2 Measurement of risk attitudes

Experimental evidence shows that individuals have different risk attitudes towards gains, losses, and mixed rewards, and typical risk preferences are best described by an S-shape function (Tversky and Kahnemann, 1992). Since we had to avoid a long questionnaire, we could only make a very superficial assessment of individual attitudes towards risk. We constructed six questions, enough to determine some basic characteristics of utility functions, but not enough to reconstruct them entirely.

As a measurement, we used the method of preference comparisons (see Farquhar, 1984). In three questions standard gambles and in three questions complex gambles were applied. In both cases, separate questions were related to gains, to losses, and to mixtures of gains and losses. With the standard gamble method we measured whether the subject is characterized as risk-averse or risk-seeking. The complex gamble was used to get an indication of individual behavior toward compound lotteries. Since decision in the IPG game has some correspondence to such a compound lottery, this instrument probably measures risk attitudes in this specific context more accurately. The gambles were constructed to represent the possible decision dilemmas subjects have faced during the experiment. Therefore, numbers in the gamble were close to monetary rewards that were at stake in the IPG game (see Table 3.A.2.1).

Table 3.A.2.1 Measurement of risk preferences

method	*range*	*option 1*	*option 2*
standard gamble method	gains	*f* 26	[*p* 0.5, *f* 15; *p* 0.5, *f* 37]
	losses	- *f* 26	[*p* 0.5, - *f* 15; *p* 0.5, - *f* 37]
	gains and losses	*f* 0	[*p* 0.5, - *f* 11; *p* 0.5, *f* 11]
complex gamble	gains	[*p* 0.5, *f* 26; *p* 0.5, *f* 31]	[*p* 0.5, *f* 15; *p* 0.5, *f* 37] + [*p* 0.5, *f* 0; *p* 0.5, *f* 5]
	losses	[*p* 0.5, - *f* 26; *p* 0.5, - *f* 31]	[*p* 0.5, - *f* 15; *p* 0.5, - *f* 37] + [*p* 0.5, *f* 0; *p* 0.5, - *f* 5]
	gains and losses	[*p* 0.5, *f* 0; *p* 0.5, *f* 5]	[*p* 0.5, - *f* 11; *p* 0.5, *f* 11] + [*p* 0.5, *f* 0; *p* 0.5, *f* 5]

We classified subjects as strong risk-averse towards gains, if they chose the risk-averse option (option 1 in Table 3.A.2.1) both in the standard and in the complex gamble with positive rewards. They were classified as strong loss-averse, if they chose the risk-seeking option (option 2 in Table 3.A.2.1) both in the standard and in the complex gamble with negative rewards. These variables were used as predictors in the analysis of contribution propensities (Sections 3.6.4 till 3.6.6).

3.A.3 Measurement of social orientations

Although in theory many types of social values are identified (cf. Liebrand, 1986), experimental evidence shows that the vast majority of people can be characterized as prosocial (cooperative), individualistic, or competitive (Suleiman and Or-Chen, 1999). For the classification of individuals into these categories two techniques are used. The *triple-dominance method* consists of a series of decomposed games with an unknown, randomly assigned person (see van Lange et al., 1997). Individuals have to make decisions between alternatives of paired monetary rewards for themselves and for the other. The triple-dominance measure of social values has the clear advantage that from few (6 or 9) questions social orientations can be detected in a reliable way. In comparison, the *ring measure* of social values (McClintock and Liebrand, 1988; van

Lange, 1999) consists of choices between 24 pairs of self-other outcome combinations. The ring measure gives a better approximation of the weights assigned to the outcome for the self and for the other. In the self-other outcome space, these weights can be conceptualized in terms of the angle θ between the horizontal axis and the line connecting the individual's ideal point with the origo (Doi, 1994):

$$SU_i = x\cos\theta_i + y\sin\theta_i, \tag{3.A.3.1}$$

where x denotes the outcome for the self and y denotes the outcome for the other. A graphical representation in Figure 3.A.3.1 indicates ranges of prosocial (cooperative), individualistic, and competitive orientations. Dashed lines separate these ranges.

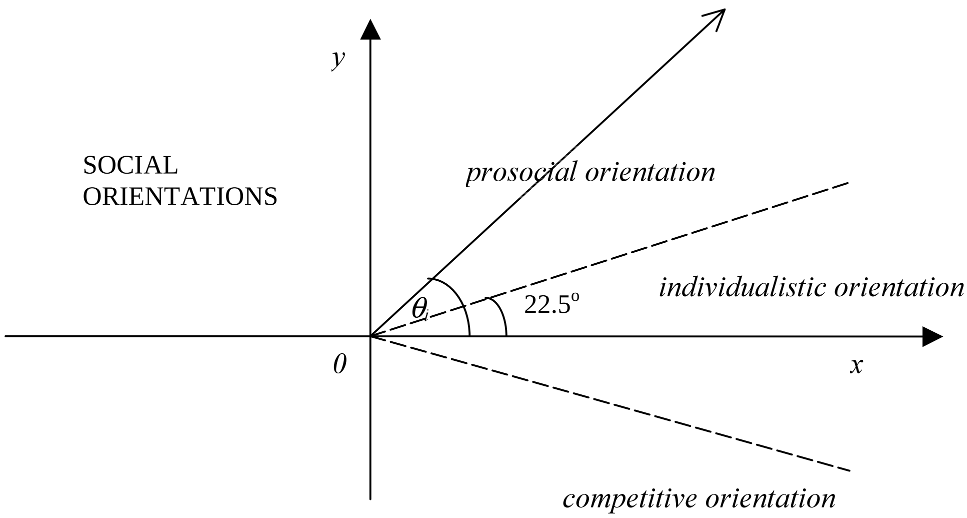


Figure 3.A.3.1. Graphical representation of social orientations with an example of prosocial (cooperative) individual values

Due to this model of social values, individuals have linear indifference curves in the self-other outcome space, with a slope of

$$RS_i = \frac{\cos\theta_i}{-\sin\theta_i}, \tag{3.A.3.2}$$

indicating the constant rate of substitution for individual *i.* For individuals, who are exactly at the border of individualism and competition, the rate of substitution is 1+√2. It means that they enjoy the same social utility, when one unit decrease in their own outcome is compensated by 1+√2 unit decrease in the outcome of the other. For competitive individuals, the substitution rate is higher than this value. Individuals at the border of cooperation and individualism have a -1-√2 substitution rate. They have the same social utility; when one unit decrease in their own outcome is compensated by

$1+\sqrt{2}$ unit increase in the outcome of the other. For individuals with cooperative social orientation, less compensation is satisfactory.

The triple-dominance measure of social orientations revealed good internal consistency and test-retest reliability over a considerably long period (e.g., Kuhlman, Camac, and Cunha, 1986; van Lange et al., 1997). However, the method is constructed only to a specific range of possible outcomes. All outcomes used by the measurement are expressed as gains, both for the self and for the other (see van Lange et al., 1997). Furthermore, the outcome for the self is never smaller than the outcome for the other.

Because of these restrictions, some paradoxes might arise showing inconsistencies and incompleteness of the social utility function in equation (3.A.3.1). Since individual behavior is significantly different toward gains and losses for the self (Tversky and Kahneman, 1986), we expect violations of the consistency of social orientations, where subjects have to decide over self-other pairs of losses or mixtures of gains and losses. Violations could also occur due to the decreasing marginal utility of money. Such a phenomenon might also exist for the outcome of the other. Furthermore, individual decisions can differ when a higher outcome is offered for the other from when it is offered to self. It is well known in social psychology that relational interpersonal comparisons affect psychological well-being (utility).

In a recent study, van Lange (1999) proposed an integrative model of social orientations. He claimed that the social utility function of cooperative individuals includes a weight assigned for the equality of the outcomes. His experimental findings provided support for this hypothesis. Previously, Knight and Dubro (1984) incorporated egalitarianism into the social utility function. In the analysis of individual behavior in the ultimatum game, which has a close correspondence with the measurement of social orientations, fairness was a major motivational concern for non-selfish behavior (Bethwaite and Tompkinson, 1996). Fairness was conceptualized as a utility element for the distance between individual payoffs. The smaller the difference is the higher the social utility of the individual.

The difference between the outcomes for the self and for the other may also matter for individuals with competitive or individualistic orientations. Kuhlman, Camac, and Cunha (1986) discussed that competitive individuals are either appetitive or aversive. Appetitive individuals have a positive desire to surpass others. On the other hand, aversive competitiveness results from a fear of being surpassed. Bethwaite and Tompkinson (1996) introduced this concept into the utility function as envy. Their experimental results showed that it was less important than fairness, but still it had an influence on behavior in ultimatum games (Bethwaite and Tompkinson, 1996).

For the problems with social orientations that occur as anomalies Schulz and May (1989) provided an elegant solution. There are individuals, who are individualistic when the outcome for self is smaller than it is for other and they are cooperative when the outcome for self is larger than the outcome for other. For them, the parameter of social orientations, which is represented as the angle θ_i (see Figure 3.A.3.1) is different for the ranges $x<y$ and $x>y$ (let us denote these by θ_i^A and by θ_i^B). Assuming that the utility function is continuous at the point, where $x=y$, this special case of conditional utility fits in the more general description of social utility (Schulz and May, 1989), given as

$$SU_i = a_i x + b_i y - c_i |x - y| \tag{3.A.3.3}$$

where a_i, b_i are the individual weights assigned to the outcome for self and for other, and c_i is an individual parameter assigned to egalitarian and other relational interpersonal tendencies. In our special case,

$$a_i = \frac{\cos\theta_i^A + \cos\theta_i^B}{2};$$

$$b_i = \frac{\sin\theta_i^A + \sin\theta_i^B}{2};$$

and $$c_i = \frac{\cos\theta_i^A - \cos\theta_i^B}{2} = \frac{\sin\theta_i^B - \sin\theta_i^A}{2}.$$

For the sake of interpersonal comparisons and graphical representation, Schulz and May (1989) introduced the restriction $a_i^2+b_i^2+c_i^2=1$ for the weights. In this way, social orientations can be represented as points on the surface of a sphere. Simple social orientations that can be described properly by the utility function in equation (3.A.3.1) are located on the equator plane. There is a third dimension that represents the importance of interpersonal comparisons. Poles of the sphere mark extremely high importance (positive or negative) for interpersonal comparisons. Schulz and May (1989) detected that social orientations of many people fall outside of the equatorial, and thus cannot be described simply as individualistic, cooperative, or competitive.

Our experiments are conducted on computers, which made it possible to ask subjects to choose self-other outcome combinations, conditional on their previous response. A new measurement is proposed here, which uses only four questions, can classify subjects into subcategories of social orientations, can detect egalitarianism, and includes a consistency check. This method, similar to the triple-dominance measure of van Lange et al. (1997), is only related to behavior towards positive outcomes in the $x>y$ range. However, the method could be used in the positive $y>x$ range (4 additional questions), towards negative outcomes (4+4 questions), and also towards mixed incentive structures (gains and losses).

Table 3.A.3.1 Measurement of social orientations for $x>y$

question 1				question 2				question 3				question 4			
ty	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*
	24	2	*c*		25	10	c^{+}		25	25	*e*		29	7	*c*
all	29	14	*i*	*c*	29	14	c^{0}	c^{+}	30	27		*all*	34	19	*i*
	26	21	*p*		33	20	c^{-}		29	28			31	27	*p*
					27	4	i^{+}		25	25	*e*				
				i	29	14	i^{0}	c^{0}	23	22					
					28	19	i^{-}		21	20					
					27	12	p^{-}		25	25	*e*				
				p	25	15	p^{0}	c^{-}	22	18					
					23	17	p^{+}		20	15					
									25	25	*e*				
								i^{+}	24	22					
									21	20					
									25	25	*e*				
								i^{0}	26	20					
									25	15					
									25	25	*e*				
								i^{-}	27	20					
									28	15					
									25	25	*e*				
								p^{-}	27	22					
									30	15					
									25	25	*e*				
								p^{0}	27	23					
									29	20					
									25	25	*e*				
								p^{+}	27	24					
									29	22					

Notes: c=competitive, *i*=individualistic, *p*=prosocial (cooperative), *e*=egalitarian orientations, *ty*=type, *slf*=reward for the self, *oth*=reward for the other.

In the first step, subjects are classified into the three main categories (cooperation, individualism, and competition) in the classical way. They have to choose between three alternatives of outcome pairs. Options are constructed in order to represent the borders between the categories (dashed lines on Figure 3.A.3.1). In the second step, the θ_i parameter is approximated, assuming that c_i is close to zero. The third step is constructed to detect egalitarianism. If subjects choose for the egalitarian choice in this question, the assumption of $c_i \approx 0$ will be violated (assuming that θ_i was approximated correctly). The fourth question is a consistency check. Numbers that are close to the rewards of the experiment were applied (cf. Tables 3.A.3.1 and 3.A.3.2).

Table 3.A.3.2 Measurement of social orientations for $x<y$

question 5				question 6				question 7				question 8			
ty	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*	*ty*	*slf*	*oth*	*ty*
	12	14	*c*		13	22	c^{+}		25	25	*e*		17	19	*c*
all	17	26	*i*	*c*	17	26	c^{0}	c^{+}	26	28		*all*	22	31	*i*
	14	33	*p*		21	32	c^{-}		28	30			19	38	*p*
					15	16	i^{+}		25	25	*e*				
				i	17	26	i^{0}	c^{0}	30	30.5					
					16	31	i^{-}		32	33					
					15	24	p^{-}		25	25	*e*				
				p	13	27	p^{0}	c^{-}	27	28					
					11	29	p^{+}		28	30					
									25	25	*e*				
								i^{+}	27	30					
									28	35					
									25	25	*e*				
								i^{0}	26	30					
									25	35					
									25	25	*e*				
								i^{-}	24	30					
									22	35					
									25	25	*e*				
								p^{-}	23	30					
									21	33					
									25	25	*e*				
								p^{0}	23	28					
									21	30					
									25	25	*e*				
								p^{+}	23	27					
									20	29					

Notes: c=competitive, *i*=individualistic, *p*=prosocial (cooperative), *e*=egalitarian orientations, *ty*=type, *slf*=reward for the self, *oth*=reward for the other.

"Social action ... may be oriented to the past, present, or expected future behavior of others. Thus it may be motivated by revenge for a past attack, defense against present, or measures of defense against future aggression. The 'others' may be individual persons, and may be known to the actor as such, or may constitute an indefinite plurality..."

Max Weber: Economy and Society, vol. I, Ch. I.1.B (1978: 22)

CHAPTER 4

DETERMINANTS OF PARTICIPATION IN DURABLE CONFLICTS

Derivation of hypotheses about effects of temporal embeddedness

4.1 Introduction

The major theoretical contribution of Chapter 2 was the incorporation of structural embeddedness and social control in the team games model of intergroup relations. In Chapter 3, we showed that social control in different forms influences individual decisions in single-shot game experiments, even if only eye contact is established between the subjects. A new experimental design was applied to test which forms of social control influence decisions and to analyze how can social control lead to higher likelihood of intergroup conflict in segregated settings. As an aggregated consequence of micro mechanisms, harmful group conflict was least likely in a setting with low segregation.

In this chapter, we construct a model about individual behavior in *repeated intergroup competitions*. In repeated interactions, temporal embeddedness together with structural embeddedness constrains individual behavior (Granovetter, 1985; Buskens, 1999). While there is no doubt about the validity of this statement, there is less consensus on *how* embeddedness constrains individual action.

With regard to *structural embeddedness*, we are interested in finding answers for the same questions as in the single-shot games. Individual actions are embedded structurally also in repeated intergroup relations; it is therefore natural to investigate the role of structural embeddedness in repeated IPG games. In particular, we would like to answer whether there is a segregation effect on the likelihood of intergroup conflict. Furthermore, we attempt to uncover the underlying mechanisms of the segregation effect. Principally, we are interested in how different forms of social control from the close environment influence individual decisions and consequently the outcome of intergroup competition. We do not go into details about research questions concerning structural embeddedness, since there are no differences in the major factors we predict to influence individual behavior in single-shot and in repeated IPG games. These major factors are different forms of social control, namely social selective incentives, behavioral confirmation, and traitor rewards. A theoretical discussion of these effects can be found in Chapter 2 and the hypotheses about their effect in the experimental IPG games in Section 3.2.

With regard to *temporal embeddedness*, we are interested in the question of *how* do the past and the future govern present decisions. Furthermore, we are interested if there are typical scenarios of intergroup conflict and peace in the experiments that are comparable with histories of intergroup relations in real life. In pursuit of underlying mechanisms of macro dynamics, we examine regularities of behavior at the individual level. Elementary behavioral rules, particularly *criticalness*, *reinforcement learning* and *reciprocity* were suggested by related research as relevant mechanisms of human behavior in similar situations (see Section 1.7). Following these recommendations, in this study we attempt to trace these behavioral mechanisms in the laboratory. As a new development compared to previous research, we test the existence of different behavioral mechanisms simultaneously.

With regard to *interactions of structural and temporal embeddedness*, we are interested whether observed scenarios of intergroup conflict and peace differ in structural conditions. At the individual level, we test whether behavioral mechanisms are applied conditional on the structural environment and we examine the impact of local reciprocity that is a heuristic triggered by previous actions in the neighborhood.

These questions are addressed in the repeated IPG game experiments. In the next section, we specify them more concretely. Subsequently, in Section 4.3.1 we discuss how we build on the foundations of the model we used for explaining individual behavior in the single-shot games. In Section 4.3.2 we explicate our hypotheses about the behavioral mechanisms of criticalness, reinforcement learning, and intergroup reciprocity. Section 4.3.3 demonstrates that structural and temporal embeddedness might interact with each other. Particularly, local reciprocal strategies receive focus here. At the end of this section, we give a summary of the main hypotheses. In Section 4.3.4 we discuss effects of other variables and their interactions that might be important for individual decisions and consequently for the outcome of repeated IPG games. In Section 4.4 we provide arguments for the multilevel methodology we use to analyze experimental data.

4.2 Research questions

In Section 1.7 we discussed how the question of whether the past *or* the future governs decisions is related to debates over individual *rationality*. We argued that neither the clearly forward-looking perspective, nor the clearly backward-looking model is appropriate to describe human behavior. In the forward-looking perspective, individuals always choose for the alternatives that offer the best future consequences, irrespective of the past. In the extreme backward-looking model, individuals are not influenced by their expectations about the future; they only base their decision on past experience. We believe that reality is somewhere in between, as both the past and the future influence individual decisions.

In order to answer the question of *how* the past and the future affect present decisions, based on arguments of bounded rationality we consider *simple behavioral heuristics*. In Section 1.7 it was argued that individuals can neither perfectly calculate all possible future consequences nor can they remember and use all information from the past. Another practical advantage of considering simple behavioral rules is a parsimonious model of individual action. This is required, because of the limited number of observations. Even if relevant, it would be difficult to test the existence of more complex decision rules.

As we put it forward in the previous section, when formulating hypotheses, we concentrate on three simple heuristics: *criticalness, reinforcement learning*, and *reciprocity.* Arguments on their relevance are driven by previous research that were discussed in Section 1.7. Here we only characterize their meaning in the specific context of repeated IPG games.

Criticalness is a behavioral rule that prescribes contribution when a single individual decision is expected to change the aggregated outcome. Otherwise, criticalness allows free riding.

For a forward-looking but shortsighted individual, who looks one encounter ahead, contribution is a beneficial action when a single contribution would change the outcome of the game. In any other situation, defection brings higher rewards. Criticalness in this *objective* sense provides the main rationale for the influence of *perceived* criticalness on individual decisions in the IPG game.

In repeated games, subjects receive information about the previous outcome. This is a major difference compared to single-shot games. This information is highly valuable for decisions in the subsequent round. In our hypotheses about how subjects use this information we focus on reinforcement learning and reciprocal mechanisms. *Reinforcement learning describes that when a particular action resulted in a satisfactory outcome, the chance of staying at this action increases. If the given action leads to an unsatisfactory outcome, the probability of choosing this option again in the next round decreases*.

Reciprocity prescribes contribution as a response to contribution of others and a choice to defect as a reaction to defection of others. It is a guideline to react nicely to nice action of others and to give a nasty response to nasty behavior. Reciprocal behavior might relate to global (the intergroup context) and also to local (neighbors) stimuli (Whatley et al., 1999). The intergroup context provokes vengeance in *interpersonal* relations between members of the opponent groups, but actions within the group might also be reciprocated. *Local reciprocity* is a similar mechanism to behavioral confirmation with the exception that it works as a reciprocation of past action and not as confirmation to expected future actions. We discuss the relation between local reciprocity and behavioral confirmation in more detail in Section 4.3.3.

Additionally, we need to provide arguments for handling the behavioral rules of criticalness, reinforcement learning, and reciprocity *simultaneously*. Previous research focuses most often on the evaluation of only one, without controlling for other mechanisms. Exceptions that made use of both reinforcement learning and reciprocal strategies are Macy (1996) and Flache (1996) for repeated PD games and Goren and Bornstein (1999) for repeated IPD games. Other attempts tried to unify belief-based models including criticalness with choice reinforcement (Ho and Weigelt, 1996; Camerer and Ho, 1999).

Handling all three mechanisms as important might help to overcome predicting failures of previous research, which originated in the oversimplification of human decision processes (cf. van Assen, 2001). *First*, an integrative view of behavior is necessary, because different individuals play different strategies. As long as we are interested in explaining the behavior of a representative subject, we have to control for the presence of relevant strategy rules among the subjects. *Second*, we are also convinced that the very same subject can also follow different behavioral rules. In repeated games, where individuals have opportunities to change their heuristics based on experience, behavior is best described by softwired and not by hardwired strategies (Macy, 1996). Subjects might switch from one rule to another as a result of experimentation, previous failures, pattern recognition, or simply by inconsistency. Individuals also follow different heuristics in different real life situations, based on the framing of the situation (Lindenberg, 1993; Lindenberg and Frey, 1993; Vanberg, 2002). Although it is difficult to observe and explain the timing of switches, at least we should not exclude the possibility of strategy changes for the very same individual.

These basic heuristics, however, are not supplementary and at some points they provide contradictory predictions. At these points, we leave the question of which effect is stronger open for an explorative analysis of the data. The exact hypotheses about these behavioral principles and about their interrelations are formulated in Section 4.3.2.

The primary goal of this study is to explain certain types of conflicts between groups. For the explanation of repeated conflicts, the role of behavioral heuristics is essential. They are, together with structural effects, the focus of our explanation. Therefore it is necessary to consider what *macro consequences* these individual behavioral rules imply in the repeated IPG game and what are the outcome scenarios that follow as aggregates from individual behavior.

Criticalness reinforces old contributors and provokes free riders to contribute after a clash situation. In a hypothetical situation, in which everyone follows the principle of criticalness, once groups are in a clash situation, clash is repeated over and over again and full contribution is established. On the other hand, peace and unconditional defeat decreases the chance of contribution, consequently a scenario of *stable peace* would follow. Criticalness might contribute to long lasting peaceful scenarios and to the *spiral of conflict* after a clash situation, but also supports a tendency towards a peaceful resolution of one-sided conflicts.

Reinforcement learning also stabilizes peace. However, after a victory of one group, reinforcement learning will not decrease contribution rates in the winning group. When everyone sticks to this principle, members of the losing side will continue to change their decisions in the hope of a better outcome (assuming that the evaluation of monetary payoffs as satisfying or not does not change over time). Conflict can only be stopped when contribution rates are equal in the two groups. After a clash, only those who were previously free riders have higher propensities to contribute. In a uniform world of reinforcement learners, such actors are too few to sustain a scenario of *durable conflict*.

Intergroup reciprocity also supports *stable peace*. Contrary to reinforcement learning, it decreases contribution rates in the winning group after a clear victory. When everyone reciprocates the action of the other group overall retaliations might result in the alternation of victory and defeat in the two groups. However, in a uniform world of reciprocators when both groups establish collective action conflict is stabilized and a *spiral of conflict* reaches full contribution. Step by step *local reciprocity* drives in the direction of uniform action. Consequently, it supports either a *spiral of peace* or a *spiral of conflict* scenario.

It is quite unrealistic to consider that everyone follows the criticalness principle or there are only reinforcement learners or reciprocators. Subjects not only differ in their propensities toward baseline contribution, but they differ also in their reactions to certain outcomes, in their expectations, and consequently in the behavioral mechanisms they follow. Furthermore, they do not apply certain schemes mechanically, but they might experiment with their decisions. They might switch to different strategies and they might also give inconsistent choices. Subject behavior has a probabilistic feature (cf. Rapoport and Chammah, 1965: 109). Therefore we should consider a *combination* of the stochastic versions of these behavioral rules in our explanation. However, in this way, macro level predictions become far more complex. Without knowing the proportion of certain types in the population, it is impossible to make exact predictions about macro outcomes and testing can only take place at the individual level.

Still, there are scenarios that are generally supported by the behavioral mechanisms and therefore we could predict their occurrence in the experiments. After an initially low contribution level, all mechanisms predict a further deflation of contribution rates. The *spiral of peace* leads to a stabilization of the peaceful outcome (*stable peace*). On the other hand, all behavioral principles provide some support for *durable conflict*. Furthermore, with the exception of reinforcement learning, all mechanisms predict a *spiral of conflict* after a clash situation. These are the scenarios of intergroup competition that are most likely also in empirical situations (Fearon and Laitin, 1996). We are interested in whether we can also find them in our repeated IPG game experiments.

At the end of this introduction, Table 4.2.1 summarizes the main research questions we address in the repeated IPG game experiments. The first group of questions is related to *aggregated outcomes*. The question of how intergroup conflicts are affected by factors of structural embeddedness, especially by *segregation*, was the primary focus of previous chapters. Here we continue our investigation of structural effects on intergroup conflict, since they are also highly relevant in repeated IPG games. Besides, our main research questions are centered around effects of temporal embeddedness and interaction effects of structure and time on intergroup conflict.

Table 4.2.1 Summary of main research questions

	Structural embeddedness	*Temporal embeddedness*	*Interactions of structure and time*
Macro effects on intergroup conflict	Is there a segregation effect? Is the segregation effect stronger under normative pressure?	How do intergroup outcomes change over time? Are there typical scenarios?	Are there different scenarios in different structures?
Micro effects on individual decisions	Are subjects influenced by forms of social control?	What are the simple heuristics that guide individual choice?	Are subjects influenced by previous decisions of their neighbors? Do they follow different heuristics in different structures?

Segregation effect on intergroup conflict:

- *Is there a segregation effect in the repeated IPG games? Are contribution rates and consequently the likelihood of conflict lower in the low clustering condition and higher in the high clustering condition? Is the segregation effect stronger under normative pressure than under confirmation pressure?*

Dynamics of intergroup conflict and peace:

- *How do contribution rates and consequently intergroup conflict change over time in repeated IPG games? Are there typical patterns of the dynamics? Does conflict elicit further conflict and a spiral of harmful outcomes? Is peace likely followed by a sequence of peaceful outcomes?*

Interactions of structure and time:

- *Are there different scenarios in different structures? Is stable peace more frequent in the low clustering condition and durable conflict in the high clustering condition?*

Outcomes are the aggregated consequences of *individual behavior*. Our next group of research questions is related to the explanatory mechanisms that connect processes at the individual level with macro outcomes. These questions are centered on the general problem of how structural and temporal embeddedness and their interaction affects choice and its consequences in intergroup relations.

Social control mechanisms operative:

- *Are subjects influenced by different forms of social control, namely by selective incentives, behavioral confirmation, and traitor rewards? Do these forms of social control also affect decisions in an internalized form?*

Behavioral heuristics:

- *What are the simple heuristics that guide individual choice? Do subjects contribute more in the repeated IPG game, if they perceive their decisions to be critical? Are their decisions reinforced by recent experience? Do their reciprocate collective actions of the other group and the behavior of their neighbors?*

Interactions of structure and time:

- *Do subjects follow different behavioral principles in different experimental conditions?*

Besides these main research questions, we would like to know, what are the most important other influences we should consider as control variables in the investigation of contribution propensities. The motivation for the inclusion of control variables stems from findings of previous research and from the need to disentangle undesired effects of the experimental procedure. These control variables are discussed briefly in Section 4.3.4.

4.3 Model of explaining contribution propensities in repeated situations

4.3.1 Relation to the model used for single-shot games

The explanation of choices in single encounters focused on the effects of structural embeddedness. For analyzing repeated intergroup situations, we choose the same modeling framework of intergroup interdependence as in the single-shot interactions, except that the IPG game is repeated over time. The payoff structure in every repetition is identical with the one in the single-shot games (cf. Section 3.3). The two (red and the green) teams consisted of the same five-five players in an unchanged seating arrangement. This also means that we could anticipate similar effects of structural embeddedness as in the single-shot games. The major difference is that subjects in repeated games are affected by previous outcomes and decisions, but future interactions might also influence their choices. Hence, we also have to count on effects of temporal embeddedness.

Our research questions about changes in aggregated outcomes can be answered by looking at elementary statistics or by analysis of variance. On the other hand, testing for underlying mechanisms behind the aggregated outcomes is only possible by analyzing decision-level data. In this way, we can test which behavioral rules are effective, what are the structural constraints, and what other variables are interfering in individual contribution decisions in repeated IPG games. In the analysis of contribution choices, we use model foundations of the single-shot games. In this section we briefly summarize the model and describe the main effects that should be included also in the investigation of repeated games.

The analysis of repeated games is far more complicated with regard to the interdependence between observations within subjects and within sessions. The major difference in comparison to single-shot games is that within session independency cannot be assumed. The outcome of the intergroup game is the same for all observations within a session in a given round. Because of the hierarchical structure of observations (decisions – individuals - sessions) multilevel logistic regression seems to be the best tool for the analysis of contribution decisions (Bryk and Raudenbush, 1992; Goldstein, 1995). In the multilevel analysis, decisions (*r*) are the first level observations, individuals (*i*) are at the second level, and experimental sessions (*x*) are at the third level.

The basic elements of the model are discussed in Section 3.4.1. The baseline model is identical to what is expressed in Equation 3.4.1.1 except that we consider three levels instead of two, and therefore a session level error term is also included (see Equation 4.A.1 in the Appendix). Our predictions about social control effects are identical to hypotheses formulated in Section 3.2 for the single-shot games. Social control effects are incorporated in the model similarly as in the single-shot games (cf. Equations 3.4.1.3 and 4.A.2).

Besides social control effects, in repeated games temporal embeddedness also influences individual decisions. Some time effects might also interact with variables of social control. Hence, the really essential formulation of an explanatory model for repeated games only starts in the next section.

4.3.2 Criticalness, reinforcement learning, and intergroup reciprocity

For the repeated games, we have to extend the explanatory model of the single-shot games by effects of *temporal embeddedness*. In this section, we derive exact hypotheses about some main effects of temporal embeddedness. Specification is necessary because the theoretical concepts of criticalness, reinforcement learning, and reciprocity do not provide us with punctual directives about how subjects would behave in the laboratory. Despite the limitations, our main hypotheses for subject behavior in repeated games are derived from the main theoretical principles. As we emphasized before, explanatory mechanisms at the micro level, including social control, criticalness, reinforcement learning, and reciprocity are all very important for understanding and explaining how and why embeddedness influences the outcomes of repeated intergroup competitions at the macro level.

A forward-looking heuristic that can be particularly important in the experiments with repeated IPG games is *criticalness* (Caporael et al., 1989; Rapoport, Bornstein, and Erev, 1989) or perceived efficacy (Kerr, 1989). Experiments with step-level public goods showed that criticalness is an important predictor of contribution (Rapoport, 1987; Erev and Rapoport, 1990; Chen, Au, and Komorita, 1996; Au, Chen, and

Komorita, 1998). Subjects might expect that their choice would be decisive on the basis of three rational arguments. First, the experimental game is a step-level public good game, therefore an individual contribution might change the outcome, if the number of contributors is just under the minimal contributing set (MCS). Second, the payoff structure includes a punishment payoff for clash (equal level of collective action). In the case of clash, one additional contribution is sufficient to change the harmful outcome into victory. Third, groups in the experiments are small (five members), therefore the likelihood that contributions are at a critical level is not extremely small. Still, there are much more situations, in which a single contribution does not make any difference and it is just a waste of bonus rewards. However, there is evidence that people overestimate their criticalness in the game and have illusions of efficacy (Kerr, 1989). We also do not require judgements of criticalness to be precise. It is more important that we believe that subjects act in accordance with these judgements.

To test whether or not criticalness plays an important role in determining individual decisions, the best is to rely on subjective expectations about the forthcoming outcome. Subjects were asked to forecast the result of the next game at the same time when they had to make their decisions. When they made these forecasts, they could already incorporate in their calculations their own subsequent decisions. They had to choose only one of the following options: peace, defeat, victory, or clash. These outcomes were formulated using the terminology “there will be not enough interest,” “my team will lose,” “my team will win,” and “there will be a draw,” respectively. On the basis of the criticalness principle, if someone anticipated a clash, then contribution is his consistent choice. Contribution is consistent in the sense of criticalness, but it is not supported by economic benefits, since the bonus for free riding is larger than the difference between the payoffs of defeat and clash. In case the subject anticipated defeat or peace, then his decision cannot make a difference for the outcome (if the subject believed that it could, then his expectations were inconsistent). Hence, not to contribute should be the appropriate action. If the subject expected a victory of his team, then we do not have a clear indication what his decision would be considering the principle of criticalness. By consulting the outcome of the previous round, we might get some hints for deriving a hypothesis. If his team did not win the previous round and he anticipated victory for the following round, then we can assume that he thought his choice might be the decisive one, therefore our prediction is that he will contribute. If his team won the previous round and he anticipated victory again, then we do not have enough information to derive clear predictions about what criticalness dictates.

By focusing on the expectations of subjects, we could separate perceived criticalness from the other principles. In general terms, we formulated the following hypothesis about criticalness:

CRITICALNESS HYPOTHESIS: *If a subject perceives his or her decision in the forthcoming round as decisive, his or her contribution to the group collective action will be more likely.*

Table 4.3.2.1 Predicted effect of subjects' perceived criticalness on their contribution propensities

expectation of subject	*previous outcome*	*predicted sign of effect*
peace (*p*)	any	-
defeat (*d*)	any	-
victory (*v*)	not victory	+
victory (*v*)	victory (*v*)	? (reference category)
clash (*c*)	any	+

We captured perceived criticalness through expectations of subjects about the subsequent round. In operationalized terms the *criticalness hypothesis* is formulated as the effects of four expectation dummies on contribution propensities (see Table 4.3.2.1). The case, in which the previous round was victory and the subject anticipated a victory for the subsequent round, is used as a reference category in the analysis. This is necessary to avoid multicollinearity, but also handy because criticalness did not provide clear predictions for this case.

For the discussion of the effects of *intergroup reciprocity and reinforcement learning* we have to rely on the previous outcome and decision. Hypotheses are formulated in stochastic terms, because of the presence of other effects and because subjects sometimes might experiment with their decisions or might act inconsistently. *Reinforcement learning* prescribes to stick to the same decision, if the previous round resulted in a satisfactory outcome, and to change the decision, if the previous outcome was unsatisfactory. Similar to the sucker's payoff and the punishment reward in the two-person Prisoner's Dilemma (e.g., Macy, 1996), defeat and clash can be considered as unsatisfactory outcomes that evoke a shift in the individual decision. Respectively, peace and victory are satisfactory outcomes that reinforce the choice of the subject. We assume a very simple reinforcement mechanism that is a stochastic version of the Win-Stay, Lose-Change (WSLC) strategy and has its reference point at the zero payoff. The zero payoff is a natural division line that separates gains and losses, therefore choosing this as a reference point is not an arbitrary assumption. Unlike in more complex reinforcement models (for an overview see Fudenberg and Levine, 1998; Macy and Flache, 2002) we assume that this reference point is fixed over time and not adjusted, if success or failure is repeated continuously. For instance, in the control and in the minimal condition of the experiment, the fixed interior reference point equals the reward for peace, if the subject contributed his or her bonus in that round. We assume a fixed interior reference point for the sake of simplicity and for the opportunity to integrate the reinforcement mechanism with reciprocity and criticalness in a transparent model. We do not make any additional assumption about the speed and accuracy of learning. In Section 4.3.4 we discuss to what extent our predictions are different, if we consider the effect of the introduction of new incentives.

The strategy that resembles the principle of *intergroup reciprocity* is based on the same mechanism as the TFT strategy in the two-person Prisoner's Dilemma (Axelrod, 1984). For the purpose of this study, it is more relevant to consider a probabilistic version of the intergroup TFT strategy. A stochastic version of TFT is more forgiving, which has proved to be an additional advantage compared to the deterministic version of TFT in two-person PD simulations (Nowak and Sigmund, 1992; Kollock, 1993). An individual, directed by a probabilistic version of the intergroup TFT strategy, would decrease his or her contribution propensity, if the other group behaved peacefully in the previous round. Respectively, his or her contribution propensity would increase, if there was a collective action in the other group. However, subjects were not always able to recognize whether or not collective action was achieved in the other group, since they were only informed about the outcome of the previous round (peace, defeat, clash, victory). This is important in the case of a previous victory. Victory does not exclude the possibility that collective action was established in the other group. Therefore, subjects cannot be completely sure that there is nothing to retaliate. On the other hand, their group was certainly more effective in mobilizing its members, which provides reasons to decrease contribution propensities based on reciprocal arguments. Intergroup reciprocity in this sense is an adjustment of contribution levels in comparison to the efforts of the other group. Hence, we could predict a negative tendency in contribution rates after a victory of the team in the IPG games, but because of the uncertainty, we handle this case as a reference category of intergroup reciprocity.

In general terms, we formulated the following hypotheses about reinforcement learning and intergroup reciprocity:

REINFORCEMENT LEARNING HYPOTHESIS: *Subjects will be more likely to stick to their decisions if they won money in the previous round and they will be more likely to change their actions if they lost.*

INTERGROUP RECIPROCITY HYPOTHESIS: *Subjects will be likely to reciprocate the observed collective behavior of the other group. Peace decreases their willingness to contribute to the collective action. Competitive behavior of the other group that results in defeat or clash increases the probability of contribution.*

For the operationalization of the hypotheses, we summarized predictions based on perceived criticalness, intergroup reciprocity, and reinforcement learning in Table 4.3.2.2. The table provides an overview of the cases in which theoretical concepts coincide and under which circumstances they give contradictory predictions. In each cell, there are three characters. The first sign indicates predictions derived from the criticalness principle, the second stands for predictions of intergroup reciprocity, and the third is for predictions of reinforcement learning in the absence of monetary selective incentives and confirmation rewards. A positive sign means that based on the corresponding decision heuristic the subject would contribute more likely in the forthcoming round than what his or her baseline contribution propensity would prescribe. A negative sign implies a decrease in the probability of contribution. Question marks denote situations in which the direction of effect cannot be derived from the theoretical principle. This is a consequence of limited feedback for the subjects in the experiment.

Table 4.3.2.2 Summary of predictions derived from criticalness, intergroup reciprocity, and reinforcement learning in the absence of additional monetary incentives.

expectation of subject		peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)	
previous decision		C	D	C	D	C	D	C	D
	peace (*p*)	- - ?	- - -	- - ?	- - -	+ - ?	+ - -	+ - ?	+ - -
previous	*defeat* (*d*)	- + -	- + +	- + -	- + +	+ + -	+ + +	+ + -	+ + +
outcome	*victory* (*v*)	- ? +	- ? -	- ? +	- ? -	? ? +	? ? -	+ ? +	+ ? -
	clash (*c*)	- + -	- + ?	- + -	- + ?	+ + -	+ + ?	+ + -	+ + ?

Note: Signs indicate predictions of these mechanisms, in this order.

If the outcome was a victory, subjects did not know how many contributors were in the other group and whether it was above the minimal contributing set or not. In other words, the origin of uncertainty is that subjects did not have information about whether the losing group had a competitive or peaceful attitude, i.e., whether they established collective action in the previous round or not. Hence, we handle this case as a reference category of intergroup reciprocity.

There are also question marks at reinforcement learning. This is when the previous outcome was peace and the subject gave away his or her bonus. It means that we are not sure whether this situation is evaluated as a gain ("Our group did not lose and there is peace.") or as a loss ("I gave away my bonus and our group did not win in the competition."). In this case the received zero reward equals the zero reference point. There are question marks also in cases, if the previous outcome was a clash and the subject kept his or her bonus. In these situations, the 11 NLG bonus compensates for the 11 NLG deduction for the clash outcome.

Since the intergroup reciprocity and reinforcement learning hypotheses are completely independent from expectations of the subject, the effect of perceived criticalness could be smoothly separated. We have problems with contradictory predictions only with regard to intergroup reciprocity and reinforcement learning. Table 4.3.2.2 shows that contradictory predictions occur in cases where the subject contributed in the previous round and the outcome of the previous round was a clash or defeat. For the sake of clarity, we report our exact predictions about *intergroup reciprocity* and *reinforcement learning* in the absence of monetary selective incentives and confirmation rewards again in Table 4.3.2.3. The first sign indicates predictions derived from intergroup reciprocity and the second is for predictions of reinforcement learning in the absence of monetary selective incentives and confirmation rewards.

These predictions are operationalized as effects of dummy variables concerning the previous outcome and decision on contribution propensities. The exact specification of how these dummies were built in the multilevel model can be found in the appendix to this chapter. Since we found that previous repeated games affected contribution propensities in the single-shot games (cf. Chapter 3), these dummy variables include the outcome of the last round of the previous experimental part for single-shot games and for the first round of repeated games.

Table 4.3.2.3 Predictions of intergroup reciprocity and reinforcement learning (in this order) on contribution propensities in the absence of additional monetary incentives

previous outcome	*previous decision*	*predicted sign of effect*
peace (*p*)	contribution (C)	- ?
	defection (D)	- -
defeat (*d*)	contribution (C)	+ -
	defection (D)	+ +
victory (*v*)	contribution (C)	? +
	defection (D)	? -
clash (*c*)	contribution (C)	+ -
	defection (D)	+ ?

Note: Signs indicate predictions of these mechanisms, in this order.

In Section 4.2 we discussed the *implications* of uniform applications of individual behavioral mechanisms *for the dynamics of intergroup relations*. It was mentioned that *stable peace* is a scenario at the macro level that is supported by all three micro mechanisms (criticalness, reinforcement learning, and intergroup reciprocity).

STABLE PEACE HYPOTHESIS: We predict that peace is a stable outcome of intergroup competition and this outcome is repeated in subsequent rounds.

Stable peace can be reached after a relatively low initial contribution rate. In such cases, all three behavioral mechanisms imply a further deflation of contribution rates.

SPIRAL OF PEACE HYPOTHESIS: If the outcome of intergroup competition is peace, we predict a gradual decrease in contribution rates.

The transformation from behavioral mechanisms at the individual level to macro scenarios is less transparent in other cases. Criticalness increases the chance that clash situations are repeated after each other. If everyone follows this principle, there is no way out from overall conflict. Intergroup reciprocity would also support the spiral of clashes. Reinforcement learning, however, would lead to a drop in contribution propensities. As reinforcement works in the opposite direction, the prediction about a spiral of conflict does not stand on as firm micro foundations as the previous hypotheses.

SPIRAL OF CONFLICT HYPOTHESIS: After a close result of the intergroup competition with conflict as an outcome, we predict that contribution rates are likely to increase gradually.

Macro predictions are also contradictory about what is likely to happen after conflict with victory of one side. In such cases, criticalness supports a tendency towards a peaceful resolution. Exceptions would be cases, in which victory has been reached by a minimal margin. More exactly, since subjects do not know the difference in the number of contributors, exceptions are cases in which subjects believe that victory has been reached by a minimal margin. As a result of reinforcement learning, contribution rates would generally *increase*. In the winning group, high contribution rates would stabilize. In the losing side, there would be more alternation of choices, which would establish a one-sided regime with repeated exploitation. On the other hand, it is more likely that intergroup reciprocity *decreases* the contribution rates of the winning group and definitely increases the contribution rates in the losing group. This might lead to a clash or even to an alternation of victory and defeat. As we assume the presence of the *combination* of these principles, observed macro dynamics are predicted to be somewhere in the middleway. Conflict is likely to be the permanent outcome, but it might occasionally change which side is winning and which is losing.

DURABLE CONFLICT HYPOTHESIS: We predict that conflict is likely to be repeated in subsequent rounds.

These scenarios are the macro consequences of individual behavioral mechanisms that might vary between structural conditions. Hence, there could be differences in which scenarios occur in different structures. Hypotheses about these interactions of structural and temporal embeddedness are discussed in Section 4.3.4.

We also have to clarify *how the predictions derived from the main micro hypotheses change during the experiment as a result of certain manipulations*. After the control condition, eye contact was established between subjects. In general, the principles of criticalness and intergroup reciprocity are not hurt by this change, or by the introduction of direct social control in Parts III and IV. The introduction of social control affects only *predictions concerning reinforcement learning*. We assumed that the reinforcement process has an interior reference point at the zero payoff. Gains are interpreted as sources of positive reinforcement and losses as sources of negative reinforcement. In Part II, internalized social incentives might change the evaluation of outcomes of the team competition. These incentives have no monetary value, but subjects order utilities to them and therefore they are substitutable with monetary payoffs. There is a definite shift in the evaluation of outcomes in Parts III and IV because of the introduction of additional monetary incentives.

In which direction subjects are influenced depends on the composition and previous decisions of the neighborhood and it depends on the relative weight of social control in the utility function of the subject. There are no major complications in the low clustering condition in which subjects have *only neighbors from the opposite group* and therefore only internalized traitor rewards can be present. Traitor rewards suppress contribution under all circumstances. Therefore defection choices are reinforced more and contribution choices are reinforced less than before. Hence all signs concerning reinforcement learning should be shifted in the negative direction in Table 4.3.2.2. The relative differences between the predictions in the cells do not change. Internalized traitor rewards cause a similar change also in the medium clustering condition.

We have a more complicated situation with respect to changes in the reinforcement learning hypotheses in the presence of social control of *fellow neighbors*. *Selective incentives* increase contribution rates in any case and therefore all prediction signs are shifted in the positive direction in Table 4.3.2.2. The effect of *behavioral confirmation*, however, is dependent on the previous decision of every fellow neighbor and on the previous choice of the subject. There are no complications in the high clustering condition, if the decisions of two fellow neighbors were *different* in the previous round. Behavioral confirmation rewards make either choice equally beneficial. All prediction signs in Table 4.3.2.2 are shifted in the positive direction, because of selective incentives.

In case fellow neighbors made *uniform* decisions (or the subject had only one fellow neighbor) prediction signs also change relative to each other. If the decision of the fellow neighbor(s) *coincides* with the decision of the subject, then this decision gains positive reinforcement (predictions are shifted in the positive direction in the case of contribution and predictions are shifted in the negative direction in the case of defection). If the decisions *do not coincide*, then there is no additional monetary gain for the subject and there is no direct loss, either. However, there were opportunity costs for the subject: if he had chosen otherwise, he would have gained extra rewards. Hence the outcome is evaluated as a loss and the subject's willingness to change his action increases.

Consequently, it is worthwhile to examine the effect of reinforcement learning separately in the experimental parts. Besides, if additional monetary incentives are introduced, predictions of reinforcement learning have to be controlled for the previous decisions of fellow neighbors. Exact hypotheses about interaction effects with reinforcement learning are formulated in Section 4.3.4, after the discussion of main neighborhood effects.

4.3.3 Neighborhood effects

In this section, we extend our model to include effects of previous neighbor decisions. Personal experiences and the experiences of friends and neighbors can be important in determining individual action in the intergroup context. Such influences might also enter the repeated IPG game experiments.

Additional to behavioral confirmation, we assume that confirmation exists in a more direct form, which can be called *local reciprocity* (Whatley et al., 1999) or *imitation* (Pingle, 1995). Subjects might be motivated to imitate their neighbors, because they think this is an easy way to find an optimal choice (Pingle, 1995). On the other hand, they might reciprocate the previous defection of neighbors to give them a lesson to cooperate (e.g., Poundstone, 1992). We can interpret the assurance process (Chong, 1991; Oberschall, 1994) in repeated collective action dilemmas similar to local reciprocity. If a friend or a neighbor participates in the collective action, the next time I might also join. If he or she does not contribute, why should I care at the following occasion? Simulation research also provides evidence for the success of local reciprocal strategies (Watanabe and Yamagishi, 1999).

These reasons suggest that subjects learn and adopt behavior from their neighbors. Local reciprocity is different from behavioral confirmation, because *it is defined as an imitation of previous action* and behavioral confirmation is specified as an imitation of expected future action. Based on evidence of *in-group reciprocity* (Brewer, 1981) our prediction is that a contribution choice of a fellow neighbor in the previous round will increase the chance of contribution and a defection choice will decrease this chance.

Equation 4.3.1.3 already includes an element that is related to the effect of neighbors from the *opposite group*. We called this form of social control traitor rewards. When subjects play the repeated game and they receive information about the behavior of their neighbors, evidence of treachery becomes clear. Subjects then can punish or give indication to their neighbors that they have chosen inappropriate actions. On the other hand, they can also reward their well-behaved (defecting) neighbors. This signaling can only be done by choosing the adequate decision in the forthcoming round (defection or contribution). Such a strategy can be derived from the reciprocity principle. As a result of local reciprocity, we predict that a contribution choice of a neighbor from the opposite team in the previous round will increase the chance of contribution and a defection choice will decrease this chance.

In general terms with regard to previous behavior of neighbors, we formulate the following hypothesis:

LOCAL RECIPROCITY HYPOTHESIS: *Subjects will be more likely to contribute, in case their neighbors contributed in the previous round. They will be more likely to keep their bonuses, in case their neighbors kept their bonuses in the previous round.*

How this hypothesis is built in the general model of individual decisions in the experiment can be found in the appendix to this chapter. At the aggregated level, local reciprocity drives stepwise in the direction of uniform behavior. When every subject would follow a local reciprocal strategy, either a *spiral of peace* or a *spiral of conflict* would emerge, depending on the initial distributions. This supports the hypotheses that were formulated about macro scenarios in the previous section.

In Parts III and IV, because of the introduction of direct social control in the medium and high clustering conditions, the effect of local fellow reciprocity is predicted to *increase*. For this reason, in the explanatory model we include two parameters for local fellow reciprocity, one is when there are no monetary confirmation rewards and another when this direct form of social control is introduced. Just as in the case of behavioral confirmation, we assume that subjects reciprocate the behavior of all fellow neighbors equally. The effect of an additional (second) fellow neighbor is as strong as the influence of the first fellow neighbor.

Another research question concerned the *interaction effect* of behavioral mechanisms and structural conditions. We were interested whether subjects follow different behavioral principles in different experimental conditions or not. We do not have strong arguments to believe that decision heuristics differ according to structural conditions. Hence, our hypothesis is that *structural conditions have no direct effect on which strategies are played by the players*. On the other hand, we still predict differences regarding the extent to which these behavioral rules can be traced in the different clustering conditions. As an aggregated consequence, observed scenarios in the IPG game would be also different. This is because we expect that contribution propensities vary between structural positions, due to different social control effects. These initial differences evoke different responses, because there are opportunities to imply only certain intergroup and local conditional strategies.

In the *low clustering condition* subjects have neighbors only from the other group, therefore there is only a pressure that decreases contribution propensities in the form of internalized traitor rewards. Peaceful behavior of others leads to lower contribution rates as a result of intergroup and local reciprocity. Consequently, we predict the occurrence of a *spiral of peace* scenario, in case contribution rates were not exceptionally low originally. After the decrease, peace will be the likely outcome of the game and not much change can be predicted during the experiment (*stable peace* hypothesis). Therefore, there will not be many cases of reciprocation of collective action and not many situations in which an individual action would be critical.

In the *high clustering condition* subjects are surrounded by fellow neighbors. They therefore experience pressure towards contribution in the form of internalized and monetary selective incentives. Besides, they will reciprocate higher contribution rates, in case they are influenced by intergroup and local reciprocity. We can therefore predict a *spiral of conflict* scenario until a stable regime of conflict is established with high contribution rates (*durable conflict* hypothesis).

After the discussion of interaction effects of structure and time, we can summarize our main hypotheses (see Table 4.3.3.1). Our major research questions were centered on the effect of structural and temporal embeddedness in repeated IPG game experiments. Structural effects have been discussed in detail and analyzed in single-shot games in Chapter 3. In this chapter, by turning towards repeated interactions, we formulated hypotheses about the effects of temporal embeddedness and their interactions with structural factors. Part of our research questions concerned the outcomes of the game through time and by different experimental conditions. These questions are related to the *aggregated consequences* of individual decisions. On the basis of the principle of methodological individualism, hypotheses for these macro questions were derived from micro hypotheses about individual behavior. The upper row in Table 4.3.3.1 is a wrap-up of the main macro hypotheses and the lower row includes the corresponding micro hypotheses. As key mechanisms at the individual level, we predicted that structural effects are mediated by different forms of social control, and temporal embeddedness influences through certain behavioral mechanisms.

Table 4.3.3.1 Summary of main hypotheses

	Structural embeddedness	*Temporal embeddedness*	*Interactions of structure and time*
Macro hypotheses *dependent variable:* outcome(s) of the game	- Conflict is more likely in segregated structures (segregation hypothesis). - The segregation effect is stronger under normative pressure.	- Stable peace. - Durable conflict. - Spiral of peace. - Spiral of conflict.	Stable peace and the spiral of peace is more likely, if segregation is low and durable conflict and the spiral of conflict is more likely in segregated structures.
Micro hypotheses *dependent variable:* individual decision(s)	Subjects are influenced by internalized and direct forms of social control namely selective incentives, behavioral confirmation, and traitor rewards.	Subjects follow (a combination of) simple behavioral rules, namely criticalness, reinforcement learning, and intergroup reciprocity.	- Subjects reciprocate previous behavior of their neighbors (local reciprocity hypothesis). - There is no direct effect of structure on which behavioral rules are followed.

4.3.4 Control variables and interaction effects

In this section, we discuss effects of other variables and their interactions that might be important for individual decisions and consequently for the outcome of repeated IPG games. These variables include *personality characteristics* and *time effects* that are not covered by the main explanatory factors.

One time effect that might cause differences in baseline contribution rates between sessions is *delay* time at the start of the experiment. Sessions were expected to start punctually, but some subjects arrived to the laboratory late, therefore causing others to wait. Meanwhile they were waiting, they might have gained some silent identification due to a minimal contact between them. However, this sort of identification might increase as well as decrease contribution rates.

Changes in individual behavior can be due to *experience* in addition to modifications in the task. Intra-individual variation of contribution rates might be time-dependent. Results of previous experiments with iterated games support the argument that subjects learn the basic characteristics of the game during the experiment, consequently outcomes get closer to overall defection over time (cf. Isaac, McCue, and Plott, 1985; Andreoni, 1988; Andreoni and Miller, 1993; Bornstein, Winter, and Goren, 1996; Goren and Bornstein, 2000). On the other hand, based on a similar argument, learning the structure of the game would mean an increase in contribution rates in segregated structures where new incentives make contribution more attractive. The most accurate analysis would take it into account that learning works differently in different structural positions and in different incentive structures. Instead of testing all trend elements in one model, which would create a huge amount of variables, we group similar structural positions and conditions and control for learning trends in these groups.

To summarize, we include the following trend variables as controls in the analysis: within part trend for no additional incentives, trend for session parts in which selective incentives are introduced, trend for parts in which behavioral confirmation is introduced, and trend for parts in which both selective incentives and behavioral confirmation are introduced. With the exception of the trend for no additional incentives, we distinguish between medium and high clustering conditions.

Detection of independent trends is only possible if the main effects of criticalness, reciprocity, and reinforcement learning are already included in the analysis. In previous research, both linear and exponential learning trends have been found. Since we do not have strong theoretical support for any of them, our analysis for this control variable is of an explorative kind.

Another control variable is the *endgame effect* of the single-shot games. Single-shot games were played in each experimental part five times and subjects knew this in advance. There should not be an endgame effect in the repeated games, because the number of decision rounds was determined randomly. Subjects did not know how many rounds they were playing. However, subject might have thought of some possibilities (e.g., five or ten repeated games) in association with the number of single-shot games played. There could be effects that are related also to the total number of repetitions in earlier parts. Subjects might get bored, feel fatigue, or might handle the experimental tasks mechanically as time passes. For this reason, we control in the analysis for the number of rounds played before in the given session.

In the repeated games, subjects were informed about the result of the game (victory, clash, defeat, or peace). Since subjects knew in advance that they receive this information, in the beginning of repeated games, they might try to build their reputations towards group fellows. A good reputation or image is expected to evoke contributions from group fellows, which is beneficial for the self in the long run (Raub and Weesie, 1990; Nowak and Sigmund, 1998; Wedekind and Milinski, 2000; Bienenstock, 2001). However, what is beneficial within the group, might lead to harmful consequences in the intergroup context. Therefore we control for a reputation effect, but we do not formulate predictions about its direction. From another perspective, the introduced reputation variable indicates the net (otherwise unexplained) difference between single-shot and repeated games.

Among our main explanatory variables, we discussed the effect of criticalness on contribution rates in the subsequent round. *Criticalness might have also a long-term effect*. Stability of the outcome might signal the subjects that their contribution does not make a difference. Therefore a series of the same result (except clash) might decrease contribution rates. This prediction assumes that subjects use long-term experiences to determine their decisions. Since this is not really obvious, we handle the long-term effect of criticalness only as a control variable.

Instead of leaving inter-individual variation unspecified, it is worthwhile to control for certain *personal characteristics* that might partly explain this variation. These control variables include social orientations, experience with similar experiments, direction of study, and number of acquaintances in the experiment. Their effect is not likely to be different from the single shot games as personality and sociological background variables are more likely to influence baseline contribution rates and less likely to influence changes in contribution propensities. However, previous experimental research shows that certain background and attitude variables have more complex or even reversed effects in repeated encounters.

With regard to *gender*, there is experimental evidence that initial differences might disappear after repetitions (Mason, Phillips, and Redington, 1991). One interpretation of this result is that payoff incentives drive subject behavior towards equilibrium outcomes regardless of initial predispositions (Mason, Phillips, and Redington, 1991: 232). Other experiments, however, found that gender differences in contribution rates do not diminish over time (Nowell and Tinkler, 1994). Because of these contradictory findings we do not explicate a hypothesis about the interaction effect of gender and time, but we include this in the analysis as a control variable.

Arguments about the effect of *risk preferences* can also be different in repeated interactions and in single encounters. There are arguments that risk averse individuals are more likely to contribute in repeated social dilemma situations than risk seeking people (Raub and Snijders, 1997; van Assen and Snijders, 2002). In the two-person PD, if the continuation probability is high, the equilibrium pair of conditionally cooperative strategies ensures higher payoffs than the equilibrium in which both players defect. Risk averse individuals are less likely to deviate from this equilibrium path and are not as much motivated by temptation incentives. However, this argument cannot be directly applied to repeated IPG games. Since overall defection is mutually harmful, risk aversion is not likely to be associated with higher contribution rates.

Besides these control variables, we consider some interactions of the main behavioral mechanisms. Predictions about criticalness and reciprocity do not change during the experiment. On the other hand, *reinforcement learning can get different operationalizations depending on the structural location of the subject and on the experimental part*. In clustered structures, where selective incentives and behavioral confirmation are distributed, all choices are more likely to be reinforced because of higher payoffs. Comparing the relative attractiveness of defection and contribution, contribution receives additional gains due to the introduction of monetary selective incentives.

In clustered structures in Parts III and IV, some of the predictions regarding reinforcement learning change. In comparison with Table 4.3.2.2, differences are as follows. *First*, if there was peace in the previous outcome and the subject contributed, this decision is certainly reinforced because of additional monetary incentives (there was a question mark at this point in Table 4.3.2.2, first row). This has an implication for the overall effect of the previous outcome on the subsequent decision (cf. Table 4.3.2.3, first row): the clearly negative predicted sign disappears. *Second*, if the previous game was lost, shifts in decision under certain conditions cannot be predicted. The previous decision is reinforced in Part IV in the high clustering condition for all subjects, who had two fellow neighbors both with identical action to theirs in the previous round. In this case, the subject earned money, though his or her team lost the competition. It is a very special case that is not likely to happen often during the experiment. If only one of the fellow neighbors had an identical action, then the subject receives a zero payoff (or 1 NLG in the case of defecting choice, which we consider as negligible). Since we assume that this is the reference point, we have no predictions whether his or her action is reinforced or not. Otherwise predictions are as in Table 4.3.2.2 and 4.3.2.3. *Third*, if the previous round ended in a clash, two fellow neighbors and at least one identical action is sufficient to elicit reinforcement of the previous decision. Otherwise predictions are as in Table 4.3.2.2 (negative sign after contribution and question mark after defection).

All these changes in our predictions are due to the introduction of additional monetary incentives in Parts III and IV. The analysis of these effects certainly raises some complications also concerning their interpretation. Basically, they are interaction effects of reinforcement learning, the experimental part, the clustering condition, and local reciprocity. However, for the sake of simplicity, we can interpret them simply as operationalizations of reinforcement learning in the new payoff structure.

We have already discussed in Section 4.2 that *local reciprocity* and *behavioral confirmation* can be interpreted quite similarly. The difference is that local reciprocity is an imitation of previous action and behavioral confirmation is an imitation of expected future action. In the single-shot games only behavioral confirmation was relevant, but in the repeated games both might play important roles. Complications arise because these two are not independent and expectations are highly influenced by previous action. However, these complications are easily solved when we use regression analysis and include both previous action and expectations among the predictors. If we did not include previous action, the coefficient for behavioral confirmation would also include the indirect effect of previous decision. In the model that includes both predictors, the indirect effect will appear as part of the local reciprocity effect. Hence, there is no need to include an interaction variable at this point.

We know from everyday experience that people differ, according to temperament, in the extent to which they are vengeful or forgiving. This implies that reciprocal intentions and *the application of intergroup and local reciprocity rules might correlate with certain social and personal characteristics*. In their classic book on the repeated two-person PD, Rapoport and Chammah (1965) found that men in general were more cooperative than women, but not at the beginning of the experiments. For the explanation of this change, they looked at conditional responses to previous decisions. They found that on one hand men were more likely to respond cooperatively to a cooperative choice than to retaliate defection and on the other hand women were much more likely to retaliate defection than to give a cooperative response for cooperation. Besides, men in general were more inclined to play TFT than women (Rapoport and Chammah, 1965: 192). In the repeated IPG game experiments we might encounter similar gender differences regarding reciprocal behavior, therefore we include such interaction variables in the analysis.

4.4 Method of data analysis

Finally, we have to be more specific about the *methodology* we use for analyzing the experimental data. We consider individual decisions as separate observations. These observations are definitely not independent. First, decisions made by the same subject cannot be handled in the same way as decisions made by different subjects. There are personal characteristics that directly influence the choice. Moreover, the effects of main explanatory variables might vary between subjects. In these cases, where we have nested sources of variability, we should use multilevel analysis (cf. Section 3.4.1).

Second, decisions made within one experimental session cannot be handled in the same way as decisions in different sessions. In the single-shot games there was no feedback after decision rounds, hence we could completely control for the effects of experimental manipulations. However, in the repeated games, where information is provided to the subjects after each decision round, complete control for session effects is impossible. Therefore, we need to introduce a third level to the analysis. An example of a variable at the session level that might influence individual decisions is delay time at the start of the experiment.

To summarize, we conduct a multilevel analysis in which single decisions are the first level observations, subjects are at the second level, and experimental sessions are at the third. Since our dependent variable (single decision) is a binary variable, we use multilevel logistic regression (Goldstein, 1995: Chapter 7). Exact specification of our model used to explain individual contribution propensities can be found in the Appendix to this chapter.

In this chapter, first we discussed our objectives for analyzing repeated experimental situations and introduced the theoretical concepts on which we built our integrated model. We formulated hypotheses about effects of structural and temporal embeddedness in the repeated IPG game. We explicated predictions of the main mechanisms that determine individual decisions as well as of possible control variables and interactions. We also discussed that what scenarios can be predicted at the intergroup level as a result of aggregation of individual decisions. At the end of this chapter, we described the research methods we use to analyze the repeated games. After this we can start reporting our results. Because of the complexity and multitude of hypotheses, we devote a separate chapter for this purpose.

Appendix

Exact specification of the multilevel model

Here we specify the multilevel models we used for explaining individual contribution propensities in repeated IPG games. As we discussed in Section 4.3.1, we use the logit function as the core of the model, since the dependent variable (individual decision) is discrete. Using the same notations as there, the baseline three-level model is expressed as

$$P_{rix} = \ln\left(\frac{P_{rix}(C)}{P_{rix}(D)}\right) = \alpha_0 + \varphi_x + \varepsilon_{ix} + \xi_{rix}, \tag{4.A.1}$$

where the *propensity* of cooperation P_{rix} in decision round r (level one) of actor i (level two) in experimental session x (level three). The propensity of cooperation is specified by the logit link function (Goldstein 1995: Chapter 7). The baseline model contains an intercept α_0 that is interpreted as the baseline contribution propensity, a session level error term φ_x, a subject level error term ε_{ix}, and an intra-individual variation ξ_{rix}. The latter term represents the residual variance that is not estimated in models that include the random intercept. We assume that

$$\varphi_x \sim N(0, \rho^2) \text{ and}$$
$$\varepsilon_{ix} \sim N(0, \sigma^2),$$

where the variances ρ^2 and σ^2 are estimated.

This baseline model is extended by the predictors of structural embeddedness (see Equation 4.3.1.3). These include internalized social control effects, namely selective incentives, behavioral confirmation, and traitor rewards, and effects of monetary rewards for social control, namely selective incentives and behavioral confirmation. Parameter estimates of these effects were denoted by s_0, b_0, t_0, s_1, and b_1, respectively. For the sake of simplicity, let us denote the vector of these parameter estimates by β_S:

$$\beta'_S = [s_0, b_0, t_0, s_1, b_1].$$

The vector β_S estimates the effect of vector S_{rix} on contribution propensities, where S_{rix} is:

$$S'_{rix} = [f_i p^{II}, (\hat{f}_{cri} - \hat{f}_{dri})p^{II}, g_i p^{II}, f_i p^s, (\hat{f}_{cri} - \hat{f}_{dri})p^b], \tag{4.A.2}$$

where f_i denotes the number of fellow neighbors and g_i indicates the number of neighbors from the other group. The expression within the parentheses denotes the difference between the expected number of contributing and defecting fellow neighbors of player *i* in round *r*. The dummy p^{II} has a value of one from the first round of Part II, the dummy p^s equals one, if monetary selective incentives are introduced, and the dummy p^b indicates the introduction of monetary confirmation rewards. Using these vectors, the three-level model that includes the structural effects can be expressed as:

$$P_{rix} = \alpha_0 + \beta_S' S_{rix} + \varphi_x + \varepsilon_{ix} + \xi_{rix}. \tag{4.A.3}$$

After this, we specify the effects of temporal embeddedness, namely perceived criticalness, intergroup reciprocity, and reinforcement learning are incorporated in the explanatory model. In operationalized terms *criticalness* is formulated as the effect of subjects' expectations about the outcome of the subsequent round. Let us denote the vector that contains the values of four dummy variables about expectations (anticipation of peace, defeat, clash, and victory, if the previous round was not victory) by C_{rix} (see Table 4.3.2.1) and the corresponding vector of parameter estimates by β_C. For tracing *intergroup reciprocity*, we use three dummy variables about the outcome of the previous round (peace, defeat, and clash) and consider a previous victory as the reference category. We denote the vector that contains their values by G_{rix} and the corresponding vector of parameter estimates by β_G. The predictions of *reinforcement learning* change during the experiment with the introduction of new monetary incentives. Therefore, for the sake of simplicity, we created two dummy variables, one that indicates when reinforcement drives towards contribution and another that indicates when reinforcement prescribes defection. We used the case in which the direction of the effect of reinforcement learning is uncertain as a reference category (see Section 4.3.4). The values of these two dummies are included in the vector R_{rix} and the corresponding parameter estimates in β_R. The model that includes all these effects of temporal embeddedness can be written as:

$$P_{rix} = \alpha_0 + \beta_S' S_{rix} + \beta_C' C_{rix} + \beta_G' G_{rix} + \beta_R' R_{rix} + \varphi_x + \varepsilon_{ix} + \xi_{rix}, \tag{4.A.4}$$

where vector entries of G_{rix} and R_{rix} are zeros for the rounds without information about the previous rounds ($r=1,\ldots,6$). On the other hand, these vectors include the outcome of the last repeated game before single-shot games in Part II, III, and IV of the experiment.

Regarding local reciprocity, let us denote the parameter estimates of reciprocating actions of neighbors from the other group by t_1, of fellow reciprocity by b_2, and of additional reciprocation, if behavioral confirmation is introduced in a monetary form by b_3. The introduction of the latter term is necessary, because monetary side payments for local coordination provide strong incentives to reciprocate the decision of fellow neighbors. For the sake of simplicity, let us denote the vector of these parameter estimates by β_L:

$$\beta'_L=[t_1, b_2, b_3].$$

The vector β_L estimates the effect of vector L_{rix} on contribution propensities, where L_{rix} is:

$$L'_{rix} = [\left(g_{c(r-1)i} - g_{d(r-1)i}\right)p^{II}, \left(f_{c(r-1)i} - f_{d(r-1)i}\right)p^{II}, \left(f_{c(r-1)i} - f_{d(r-1)i}\right)p^{b}], \quad r>6 \text{ if Part II}, \qquad (4.A.5)$$

where $g_{c(r-1)i}$ denotes the number of neighbors of i, who belonged to the opposite group and who contributed in the previous round. Similarly, $f_{d(r-1)i}$ denotes the number of fellow neighbors of i, who defected in the previous round. Vector entries are zero in the control condition (Part I) and in the first six rounds of Part II, since subjects had no information about decisions of neighbors in these rounds. On the other hand, for single-shot games in Parts III and IV, information from the last repeated game is used.

After including local reciprocity the three-level model can be written as:

$$P_{rix} = \alpha_0 + \beta'_S S_{rix} + \beta'_C C_{rix} + \beta'_G G_{rix} + \beta'_R R_{rix} + \beta'_L L_{rix} + \varphi_x + \varepsilon_{ix} + \xi_{rix}. \qquad (4.A.6)$$

Equation 4.A.6 contains all of our main explanatory variables. Some of them might interact with each other and there are also important personality characteristics that cause significantly different behavior. In the extended versions of the model, these are simply added as additional predictors.

Until now we described a model in which we assumed that parameter estimates do not vary between individuals. However, it is possible that our main predictors influence decisions differently subject by subject. For instance, one subject may reciprocate more the actions of his or her neighbors while another only considers whether his or her decision resulted in a satisfactory outcome. Hence, we analyze also a model in which we allow for a random variation of the parameter estimates of our main explanatory variables around their mean. We assume that this variation follows a normal distribution. If we denote the vector of parameters by B ($B'=[\beta'_S, \beta'_C, \beta'_G, \beta'_R, \beta'_L]$) and the vector of observations by capital X ($X'=[S', C', G', R', L']$), then it means that under this model equation (4.A.6) can be written as

$$P_{rix} = \alpha_0 + B'_i X_{rix} + \varphi_x + \varepsilon_{ix} + \xi_{rix}, \qquad (4.A.7)$$

and

$$B_i = B + u_{ix}, \tag{4.A.8}$$

where u_{ix} is an error vector that follows a multivariate normal distribution with

$$u_{ix} \sim \mathbf{N}(\mathbf{0}, \Omega),$$

where the elements of the covariance matrix Ω are going to be estimated. To keep the analysis simple and parsimonious, we restrict the covariance estimates to zero. In general, fixing covariances to zero should be based on deviance tests that compare models that include and exclude them (van Duijn, van Busschbach, and Snijders, 1999). These covariances would not be meaningless, but with their estimation our model would go far beyond a comfortable size.

"There is another motivation which some would class as narrowly self-interested and some as moral. ... It is the desire to be there, to take part in history, to have oneself develop through participation…"

Russell Hardin: Collective Action (1982: 108)

CHAPTER 5

CRITICALNESS, REINFORCEMENT AND RECIPROCITY

Experimental evidence of behavior in repeated situations

5.1 Introduction

In this chapter, we present results of the analysis of repeated Intergroup Public Goods (IPG) game experiments. We describe characteristics of the subject pool and the experimental procedure in Section 5.2. The structure of the rest of the chapter follows the classification of our main hypotheses (see Table 4.3.3.1) as it is summarized in Table 5.1.1.

Table 5.1.1 The structure of this chapter following the outline of hypotheses

	Structural embeddedness	*Temporal embeddedness*	*Interactions of structure and time*
Macro hypotheses *dependent variable:* outcome(s) of the game	Section 5.3.1	Sections 5.3.2, 5.3.3	Section 5.3.3
Micro hypotheses *dependent variable:* individual decision(s)	Section 5.6	Sections 5.4.1, 5.5, 5.6, 5.7	Sections 5.4.2, 5.4.3, 5.6

First dynamics of intergroup conflict and peace are investigated under different structural conditions in Section 5.3. This investigation is related to macro hypotheses about the effect of structural embeddedness, of temporal embeddedness, and of their interaction on intergroup conflict. In particular, we test whether conflict is more likely in segregated structures, we try to trace typical scenarios of intergroup relations and we discuss differences in scenarios by structural conditions.

We turn to the analysis of individual behavior in Section 5.4. We discuss how decisions are conditional on subjects' expectations, on their previous decisions, on previous outcomes, and on previous decisions of neighbors. This investigation is related to micro hypotheses about the effect of temporal embeddedness and of interactions of structure and time.

In Section 5.5, we analyze questionnaire data that provide an indirect test for the micro hypotheses about the use of certain decision heuristics, including criticalness, reinforcement learning, intergroup and local reciprocity. These main micro hypotheses are tested directly in Section 5.6. In this section, the results of multilevel analyses are presented, explaining individual contribution propensities.

Additionally, in Section 5.7 we take a closer look behind the *criticalness hypothesis* by analyzing the origins of different subjective expectations. We summarize our results and conclude the chapter in Section 5.8.

5.2 Design of the repeated IPG game experiments

5.2.1 Subjects

Repeated games followed the single-shot games in each experimental part. The subject pool was exactly the same. The major characteristics of the subjects were discussed in Section 3.5.1. As in the single-shot games, in case a subject ran out of decision time, a random decision was implemented. For all such cases, final payment was decreased by 1%. Altogether, it has happened only 31 times from 14275 decisions (0.22%). 26 out of the 31 cases were during the single-shot games. Random decisions are not included in the analysis.

5.2.2 Procedure

Structural conditions were the same in the repeated games as in the single-shot games. Details about the design were discussed in Section 3.3. Repeated games were embedded in the procedure that is described in Section 3.5.2. In each experimental part, five single-shot games were played first. After the fifth round subjects were notified by a short message "From now on every decision round will count in your final payment." This message was on their screens for ten seconds.

From the sixth round on, subjects received the information of what was the outcome of the previous round. At the beginning of Part II, separating walls were removed and flags were attached to the monitors to indicate group membership. Additionally, subjects also received colored A-4 size papers with the color of their group. After the sixth round in Parts II, III, and IV subjects also received information about the decision of their neighbors. This information appeared on their screens for ten seconds together with a flag indicating the group membership of their neighbors. No information was provided about the decision of other players.

Subjects had the same amount of time for every decision in the repeated games. They had to make their decision whether they would keep the 11 NLG bonus or give it to help their group to achieve success in the competition. They also had to indicate which outcome they anticipate and what kind of actions they expect from their neighbors in the subsequent round. Subjects had 22 seconds to make a decision and to provide information about their expectations. A clock was indicating how much time they had, counting backwards from 20 seconds. After 15 seconds a warning message appeared on the screen ("Decide now!"). Decisions made with some delay (two seconds) were still accepted. Subjects could reconsider their decision until they clicked on the confirmation button. If there was no click on the confirmation button, the last selection was interpreted as final choice.

The number of decision rounds in the repeated games was determined randomly. Subjects did not know how many rounds they play. Within experimental parts, the number of rounds varied between 13 and 22, with one outlier of 29 rounds. On average, in one experimental part subjects played 17.51 rounds (of which five were single-shot games) with a standard deviation of 2.46. Altogether 1471 rounds were played in the 21 experimental sessions.

5.3 Results of the intergroup competition

5.3.1 Results by structural conditions

In Section 5.3, we provide statistics that are related to our research questions about aggregated outcomes of the repeated IPG games. First, in Section 5.3.1, we compare outcomes and contribution rates between the different structural conditions of the experiments. The main dependent variable at the macro level is the outcome of the game. As before, the category *conflict* is used for all outcomes, in which at least in one of the teams the number of contributors exceeded the minimal contributing set (victory, defeat, and clash). Including single-shot and repeated rounds, altogether 1471 IPG games were played in the experiments. From these, 1081 (73.5%) ended in conflict. In 861 (58.5%) cases, one of the teams achieved victory and in 220 (15.0%) cases there was a clash with equal numbers of contributors in the teams.

Table 5.3.1.1. Outcomes by clustering conditions in the experiments

clustering condition in the experiment	*outcome of the decision round* peace	conflict	Total
control condition	29.7% (1041)	70.3% (2463)	100% (3504)
low clustering	59.9% (1830)	40.1% (1223)	100% (3053)
medium clustering	11.4% (454)	88.6% (3516)	100% (3970)
high clustering	13.1% (490)	86.9% (3258)	100% (3748)
Total N	26.7% (3815)	73.3% (10460)	100% (14275)

Notes: Cases in parentheses are weighted (multiplied) by the number of human decisions in the given game. For the χ^2–test unweighted outcomes are used, N=1471.

The *segregation hypothesis* (Section 3.2) predicted that conflict would be least likely in the low clustering condition and would be most likely in the high clustering condition. Table 5.3.1.1 summarizes the frequency of intergroup conflict and peace by clustering conditions in the repeated IPG games. The hypothesis that outcomes are independent of clustering conditions can be rejected ($\chi^2(3)=251.143$, $p<0.001$). As predicted, peace was most likely in the low clustering condition. Conflict, however, occurred slightly more often in the medium than in the high clustering condition, which contradicts our structural hypothesis.

Similar to the overall figures (cf. Table 5.3.1.1), in Part II conflict occurred more frequently in the medium clustering condition (84.6%) than in the high clustering condition (78.3%). This could be explained by a *ceiling effect*, according to which internalized forms of social control evoke sufficiently high number of contributions already in the medium clustering condition. In relation to our theoretical analysis, this corresponds to results from the normative pressure condition under which the likelihood of conflict was already high in the middle ranges of clustering (cf. Figures 2.8.1.2, 2.8.2.1, 2.8.2.2). Conflict occurred frequently in the medium clustering condition in Part III, regardless of whether behavioral confirmation (87.7%) or selective incentives were introduced (89.3%). On the other hand, there was a clear difference (t=4.215, p<0.001) between these two conditions in the high clustering condition (70.5% and 96.9% respectively). The relative frequency of sequences of conflict is similar to these figures.

Table 5.3.1.2. Average contribution rates (%) in different clustering conditions and different parts of the experiment

	incentives introduced first	**low clustering**	**medium clustering**	**high clustering**	**Total**
Part I (*control condition*)*		45.57 (937)	50.31 (1270)	46.89 (1286)	47.48 (3493)
Part II (*minimal contact*)		43.93 (972)	58.50 (1323)	53.23 (1255)	52.65 (3550)
Part III	*b* (*confirmation pr.*)	-	59.51 (699)	51.15 (610)	55.61 (1309)
	s (*normative press.*)	-	67.68 (659)	75.53 (617)	71.47 (1276)
	Part III total	33.08 (925)	63.48 (1358)	63.41 (1227)	55.44 (3510)
Part IV	*b* (*confirmation pr.*)	-	61.55 (632)	71.65 (589)	66.42 (1221)
	s (*normative press.*)	-	73.08 (650)	88.62 (668)	80.96 (1318)
	Part IV total	27.86 (1152)	67.39 (1282)	80.67 (1257)	59.58 (3691)
Total (without Part I)		34.57 (3049)	63.08 (3963)	65.79 (3739)	55.86 (10751)
Total		37.16 (3986)	59.98 (5233)	60.96 (5025)	53.94 (14244)

Notes: The numbers of cell-relevant cases are in parentheses. All human decisions of single-shot and repeated games are included.

* In Part I subjects did not know their group membership and they did not see each other. Therefore their partition into the different clustering conditions only illustrates baseline contribution rates in the different experimental sessions.

We predicted a segregation effect on intergroup conflict because we assumed that contribution rates differ in these conditions. Table 5.3.1.2 gives a summary of contribution rates by clustering conditions. The hypothesis that contribution rates are the same in the different conditions can be rejected (ANOVA F(3, 14240)=304.482, p<0.001). There is an interesting contrast between this table and Table 5.3.1.1. In the high clustering condition conflict occurred less often, but the total contribution rate was higher (65.79%) than in the medium clustering condition (63.08%; t=2.484, one-tailed p=0.007). Additional to the ceiling effect, Table 5.3.1.2 provides an indication why it is the case. Subjects in the medium clustering condition by chance were more likely to contribute than subjects who were assigned to the high clustering condition (contribution rates in the control condition were 50.31% and 46.89%; t=1.733, one-tailed p=0.042). We could observe this already in the single-shot games of the first experimental part (see Section 3.6.1). In Part II, the difference in favor of the medium clustering condition became higher, which provides indirect support for the intergroup reciprocity hypothesis.

Contribution propensities were highest in the medium clustering condition in all experimental parts except when monetary selective incentives were introduced. It is not surprising that the introduction of monetary selective incentives made a radical turn in this relationship. This supports our theoretical prediction that *under normative pressure the segregation effect on intergroup conflict will be stronger than under confirmation pressure*. Subjects recognized the change in the payoff structure and were more motivated to contribute if they received selective incentives twice. The high frequency of conflict in Part III increased contribution rates also in Part IV. This explains the difference in contribution rates in Part IV between conditions where behavioral confirmation and where selective incentives were introduced first as additional rewards.

The hypothesis that contribution rates are the same in Part II can be rejected (ANOVA $F(2, 3547)$=24.307, p<0.001). Contribution rates are significantly highest in the medium clustering condition (compared to the high clustering condition t=1.931, one-tailed p=0.028). This indicates that history effects in the repeated games strengthened group composition effects. A comparison of Part I and II reveals that minimal contact made an increase in contribution rates compared to the control condition (t=4.088, p<0.001).

By considering the change in contribution rates through experimental parts, the hypothesis that contribution rates are the same in the different parts can be rejected (ANOVA $F(3, 14240)$=35.625, p<0.001). Test of linearity shows that there is no significant deviation from a linear trend through experimental parts (F=0.513, p=0.599). This partly supports that, similar to the single-shot games, analysis of individual level data should be controlled for a linear between-parts trend.

We also tested whether the introduction of different monetary incentives in Parts III and IV made a difference in contribution rates (see Table 5.3.1.2). The hypothesis that contribution rates are the same in the different conditions can be rejected both in Part III (ANOVA $F(2, 3507)$=175.890, p<0.001) and in Part IV (ANOVA $F(2, 3688)$=474.154, p<0.001). As predicted, the introduction of monetary selective incentives resulted in higher contribution rates than did the introduction of behavioral confirmation (t=8.493, p<0.001). Which incentives were introduced first made a significant difference also in Part IV (t=8.452, p<0.001). This result indicates that history effects play a significant role in between experimental parts.

This section concerned our macro hypotheses about structural embeddedness. We compared outcomes and contribution rates between different conditions of the experiment. The *segregation hypothesis* is partly supported by the data. As predicted, conflict was least likely in the low clustering condition. Intergroup conflict, however, was not more frequent in the high clustering condition than in medium clustering. This result was partly the consequence of a ceiling effect and was partly caused by high baseline contribution propensities in the medium clustering condition. Furthermore, we also found support for the hypothesis that the segregation effect would be stronger under normative pressure than under confirmation pressure.

5.3.2 Conflict over time

This section reports experimental results that are related to our macro hypotheses about effects of temporal embeddedness. Particularly, we show how conflict emerged and changed over time in the experiments.

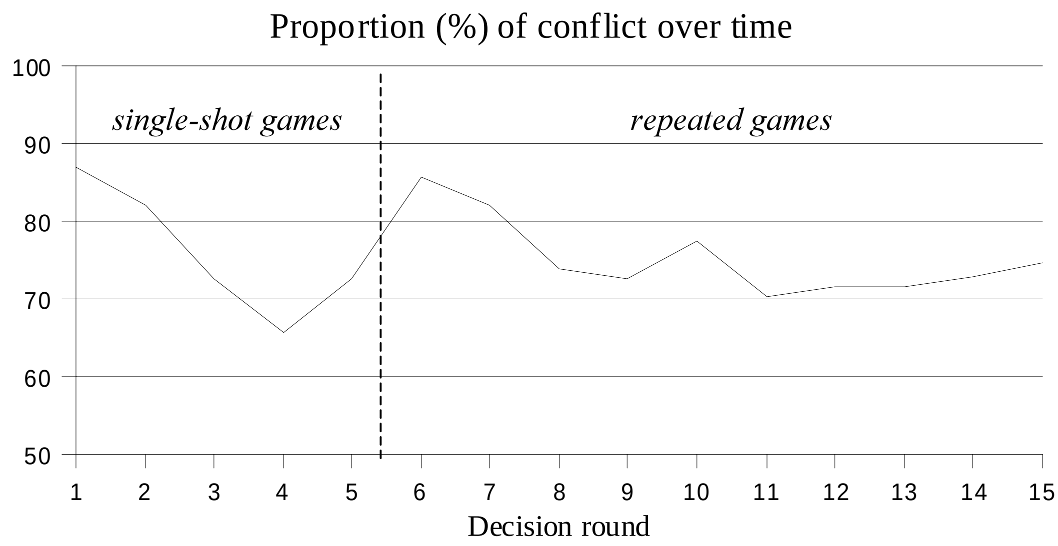


Figure 5.3.2.1 The likelihood of intergroup conflict over time aggregating all experimental parts and all experimental conditions

Note: Rounds 1-5 are single-shot games. N=84 for rounds 1-13, N=81 for r=14, and N=71 for r=15.

Figure 5.3.2.1 demonstrates how the likelihood of intergroup conflict changed within experimental sessions. There was a clear decreasing tendency in the first four rounds of the single-shot games (cf. Section 3.6.5). In the first repeated game (round 6) the frequency of intergroup conflict increased basically to the same level as it was in the first single-shot game. This was followed by a similar decrease as in the single-shot games. The gradual decrease was interrupted in round 10 with a sudden positive shift. From round 12, there was a slight increase in the frequency of conflict again.

Real reasons for these changes can only be found after the deeper analysis of micro processes that is going to follow in the subsequent sections. If we do not find convincing support for these macro outcomes relying only on the assumed micro mechanisms, we could consider the following *ex post* explanation. The surprising shift could be explained by a *hypothetical endgame effect* that is an indirect consequence of single-shot games. Subjects played five single-shot rounds and they might have (falsely) believed that there would be also five rounds of repeated games. Similar to the peculiar end-behavior in the single-shot games, subjects increased their contribution propensities at the hypothetical end. Since we did not ask subjects about such expectations, we cannot test this alternative hypothesis in this data.

In the control condition (Part I), there are no neighborhood effects and consequently no structural effects, either. Therefore, we gain an important insight by looking at data from this condition separately. The frequency of conflict in this part is reported in Figure 5.3.2.2. Conflict was the most frequent in the very first single-shot game. After a gradual decrease, conflict occurred more often in the last single-shot game. In the first repeated game, conflict was even more likely, but still not as likely as in round 1. In later rounds, there was less conflict, with a nadir in round 12. In the last rounds, the proportion of conflict increased again.

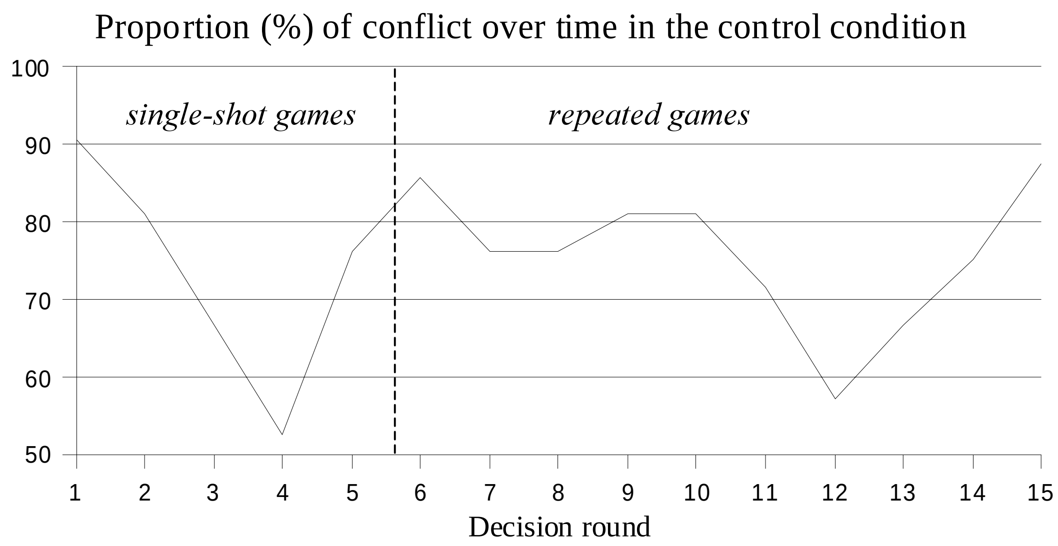


Figure 5.3.2.2 The likelihood of intergroup conflict in the control condition (Part I)
Note: Rounds 1-5 are single-shot games. N=21 for rounds 1-13, N=20 for r=14, and N=16 for r=15.

Figures 5.3.2.1 and 5.3.2.2 provided aggregated statistics about how conflict occurred over time within experimental parts. In order to get a somewhat closer view on scenarios of intergroup conflict and peace, we report how different outcomes followed each other in the experiment. This brings us closer to the testing of *stable peace* and *durable conflict* hypotheses. Table 5.3.2.1 contains information on transition probabilities between different outcomes.[1] For example, the fourth cell in the first row indicates that clash occurred after peace in 17 cases and the transition probability from peace to clash is 4.4%. A peaceful outcome was repeated in 57.6% (N=223) of the cases and conflict elicited further conflict in 85.8% (N=822) of the cases, which supports the

[1] The total number of cases in columns and in rows are different, since previous outcomes do not include results of the last round and subsequent outcomes do not include results of the first repeated game in Part I. Results from the single-shot games of Part I are not included.

stable peace and *durable conflict* hypotheses. If we assume some random variation in the decisions, it is not surprising that a clash outcome was not so frequently repeated. Instead, clash was more likely followed by victory of one team.

Table 5.3.2.1 Frequency of outcomes conditional on previous outcomes in the *entire* experiment

previous outcome outcome	peace (*p*)	victory of the red team	victory of the green team	clash (*c*)	*conflict total*	Total
peace (*p*)	223 (57.6%)	69 (17.8%)	78 (20.2%)	17 (4.4%)	164 (42.4%)	387 (100%)
victory of the red team	54 (14.1%)	177 (46.1%)	83 (21.6%)	70 (18.2%)	330 (85.9%)	384 (100%)
victory of the green team	59 (15.7%)	89 (23.7%)	156 (41.6%)	71 (18.9%)	316 (84.3%)	375 (100%)
clash (*c*)	23 (11.6%)	60 (30.2%)	70 (35.2%)	46 (23.1%)	176 (88.4%)	199 (100%)
conflict total	136 (14.2%)	326 (34.0%)	309 (32.3%)	187 (19.5%)	822 (85.8%)	958 (100%)
Total	359 (26.7%)	395 (29.4%)	387 (28.8%)	204 (15.2%)	986 (73.3%)	1345 (100%)

Notes: Number of cases. Color labels have not been introduced yet in the control condition. For single-shot games and for the first repeated game the outcome of the last round in the previous part is considered as a previous outcome.

Table 5.3.2.1 reported transition probabilities between outcomes in the entire experiment. To demonstrate the *net* effect of temporal embeddedness, we report in Table 5.3.2.2 how outcomes followed each other in the control condition (Part I). In this part of the experiment, there were no effects of structural embeddedness and social control. Results show clear support for the *stable peace* and *durable conflict* hypotheses. Peace followed peace in 62.3% (N=43) of the cases and conflict was repeated in the subsequent round in 80.7% (N=134) of the cases. The frequent repetition of intergroup conflict is a consequence of high baseline contribution rates.

Since color labels had not been introduced yet in the control condition, all differences between the red and the green groups are simply coincidental. It seems from Table 5.3.2.2 that subjects, who by chance were assigned to the green group, were less consistent in their decisions. They moved their group more often out of peace, but they were also more likely to leave the bandwagon after victory.

Table 5.3.2.2 Frequency of outcomes conditional on previous outcomes in the repeated games of the *control condition*

previous outcome **outcome**	peace (*p*)	victory of the red team	victory of the green team	clash (*c*)	*conflict total*	Total
peace (*p*)	43 (62.3%)	9 (13.1%)	14 (20.3%)	3 (4.3%)	26 (37.7%)	69 (100%)
victory of the red team	6 (9.0%)	31 (46.3%)	21 (31.3%)	9 (13.4%)	61 (91.0%)	67 (100%)
victory of the green team	20 (28.2%)	20 (28.2%)	21 (29.6%)	10 (14.1%)	51 (71.8%)	71 (100%)
clash (*c*)	6 (21.4%)	6 (21.4%)	13 (46.4%)	3 (10.7%)	22 (78.6%)	28 (100%)
conflict total	32 (19.3%)	57 (34.3%)	55 (33.1%)	22 (13.3%)	134 (80.7%)	166 (100%)
Total	75 (31.9%)	66 (28.1%)	69 (29.4%)	25 (10.6%)	160 (68.1%)	235 (100%)

Note: Number of cases. Color labels have not yet been introduced in Part I.

As we are interested in tracing typical scenarios, we should not look merely at subsequent rounds. If we consider all *sequences* of four subsequent outcomes, we find that out of 349 cases, in 140 (40.1%) cases, peace was repeated in the last three rounds before a peaceful outcome. This supports the *stable peace* hypothesis. Sequences of intergroup conflict were even more likely. Out of 954 cases ended in conflict, 629 (65.9%) were preceded by three conflicting outcomes. However, sequences of conflict in most cases did not mean that the same outcome was repeated. There were 68 (7.1%) cases, in which the red team won the competition four times in a row. The same happened 60 (6.3%) times with the green team. Clash almost never took place four times in a row (3 cases). The latter means that the *spiral of conflict* hypothesis that could be derived from the criticalness and reciprocity micro principles did not gain support from the data. On the other hand, groups hardly found the way out of *durable conflict*.

Looking at the same figures in the control condition, in which there were no effects of structural embeddedness, we find that peace occurred three times in a row before a peaceful outcome in 20 (30.8%) cases. A sequence of three conflict outcomes preceded conflict in 70 (54.7%) cases. These figures are lower relative to the overall case, indicating that the introduction of structural embeddedness increased the likelihood of *stable peace* and *durable conflict*.

Our hypotheses about typical scenarios imply that the number of contributors in the red and in the green group should correlate. Indeed, by looking at all experimental parts, we find that there is a significant correlation (Pearson r=0.391, p<0.001). On the other hand, we should not find significant correlation in the single-shot games of the control condition, because there is no feedback information about previous outcomes. For these rounds, there was no significant correlation between the number of contributors in the red and in the green group (Pearson r= -0.063, p=0.526, N=105).

When subjects learned the outcome of previous rounds in the repeated games of the control condition, the correlation between the number of contributors in the red and in the green group became significant (Pearson r=0.192, p=0.002, N=256). This correlation increased in repeated games in later parts of the experiment (Pearson r=0.255, p<0.001, N=262 in Part II; Pearson r=0.465, p<0.001, N=257 in Part III; and Pearson r=0.557, p<0.001, N=276 in Part IV). This implies that the likelihood of a clash is also increased and it was more likely that single individual decisions were decisive. These correlations support the *spiral of conflict* and *spiral of peace* hypotheses.

Although subjects were informed only about the outcome of the previous round and not about the exact number of contributors in the two groups, it is interesting to look at these figures. Before the introduction of minimal contact, there were no other incentives at stake than the monetary payoffs. Under these conditions, subjects are almost always better off by keeping their bonus. However there were only 5 cases out of 361 outcomes in the control condition without any contributors. Three of the five {0;0} outcomes occurred in the last repeated games of the same session. This was the only session in which we could trace an emergence of the overall defection equilibrium.

In this section, in relation to our macro hypotheses about effects of temporal embeddedness, we analyzed how conflict and other outcomes of the IPG games changed over time in the experiment. We predicted the occurrence of certain typical scenarios, such as *stable peace* and *durable conflict*. We found evidence of both. On the other hand, *durable conflict* did not mean that groups engaged in a series of clashes. In this section, we did not discuss differences of observed scenarios between structural conditions. These differences together with the tests of the *spiral of peace* and *spiral of conflict* hypotheses will be discussed in the following section.

5.3.3 Contribution rates over time

This section is still related to research questions that regard aggregated consequences of individual decisions over time. However, the focus now is closer to the individual; instead of examining the outcomes of the repeated IPG game, we summarize statistics about individual decisions.

Figure 5.3.3.1 reveals that after the radical decrease in the single-shot rounds, contribution rates remained at a medium level in the repeated games. This does not correspond with any of the typical scenarios we predicted. However, an average oscillation at a medium level can also be an aggregation of decreasing and increasing trends in different experimental sessions. While in some experiments intergroup reciprocity leads to a *spiral of conflict*, in other sessions it leads to a *spiral of peace*. In order to see whether that was indeed the case, we examined the dynamics of 21 experimental sessions separately. We analyzed contribution rates in the repeated games of the control condition, because in this part of the experiment there were no effects of structural embeddedness.

The experimental data do not support the hypothesis that the average oscillation at a medium level is an aggregation of counterpolar trends. We conducted logistic regression analyses on data from each experimental session, assuming a linear trend over time. Results showed no significant change in 18 sessions. There was a significant decrease over time only in three sessions ($p<0.05$, in two cases $p<0.01$). In one additional case, contribution rates increased significantly until round 10 and then dropped again. From the remaining sessions, there were huge fluctuations in contributions in three experiments. In most of the experimental sessions (14) however, *contribution rates were relatively stable over time*. These findings support the *stable peace* and *stable conflict* hypotheses, but show no correspondence with the *spiral of peace* and *spiral of conflict* hypotheses. To find more convincing explanations for the changes in average contribution rates, we will turn to the analysis of individual behavior in the following sections.

Our macro hypotheses about interactions of structure and time predicted that there are different scenarios under different structural conditions. Results show clear support for the hypothesis that *peace is a highly stable outcome in the low clustering condition*. Considering all sequences of four subsequent outcomes in the low clustering condition, we find that out of 187 cases, peace was preceded 122 (65.2%) times by a sequence of three peaceful outcomes. This is a higher proportion than what we have experienced in Part I (30.8%) and also higher than the overall figure (40.1%).

On the other extreme, *conflict was basically unavoidable* in Part IV *in the high clustering condition*. All 129 games ended in conflict except one. Still, there was only a single case in which there were four clashes after each other. Sequences of four victories of the red team occurred 15 times (11.7%) and series of four victories of the green team took place in 4 cases (3.1%). Thus, in most cases, *durable conflict* meant changing fortune for the two sides. In Part IV, there was not much hope for peace in the medium clustering condition, either; the outcome of the game was conflict in 120 of the 131 cases (91.6%). Peaceful outcomes concentrated in few sessions (five of them occurred in one session). Sequences of four conflict outcomes were therefore frequent (N=100, 83.3%).

For a closer investigation of our main hypotheses about the macro dynamics of the repeated games, let us consider average contribution rates over time. To search for any evidence of a *spiral of conflict* or a *spiral of peace* in any of the experimental conditions, we will also analyze trends in different structural conditions separately.

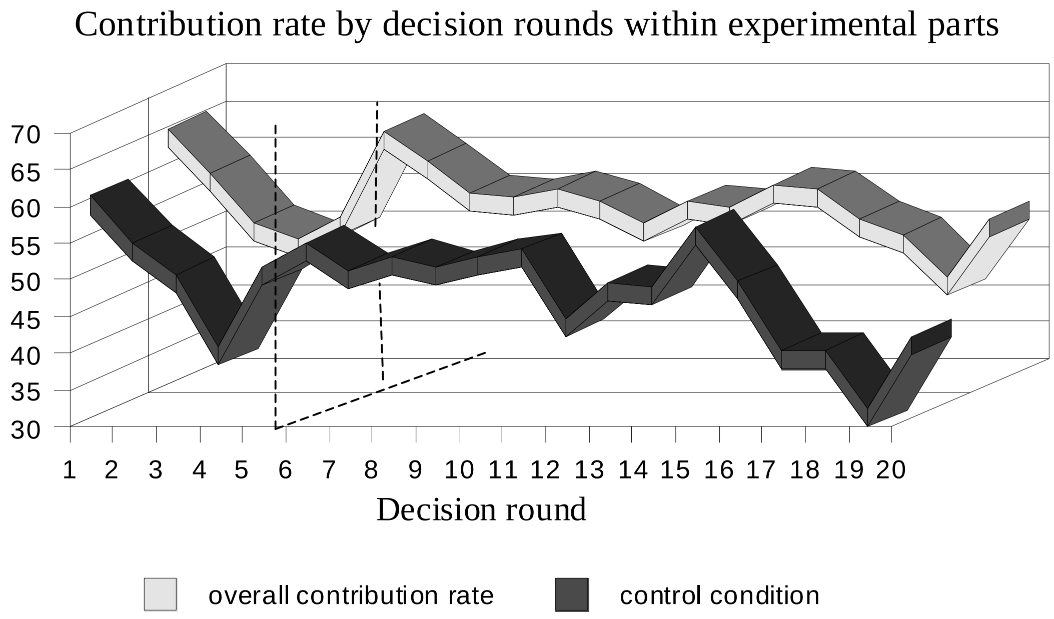


Figure 5.3.3.1 Contribution rates by decision rounds within parts
Notes: Single-shot games are rounds 1-5. Results are shown for the first 15 repeated games.

As Figure 5.3.3.1 shows, without controlling for other variables, contribution rates decrease over time, but with surprising fluctuations. When the positive endgame effect of the single-shot games is handled as an outlier, results show a sharper decrease in contribution rates in the single-shot games than in the repeated games. Contribution rates in the control condition (lower curve in Figure 5.3.3.1) are most closely comparable to anonymously played repeated PD, IPD, and IPG game experiments. Typically, these experiments find a quick decrease in contribution rates in the beginning of the experiment and a smooth decrease in later rounds (cf. Bornstein, Winter, and Goren, 1996; Goren and Bornstein, 1999). Contribution rates over time in our experiment show quite a different pattern with a smooth decrease throughout the repeated games with an exception of a larger drop in late rounds.

Similar to the upward shift that was observed in the dynamics of intergroup conflict in Figures 5.3.1.1 and 5.3.1.2, we can detect a rise in contribution rates in round 10 (cf. Figure 5.3.3.1). This local maximum is followed by a decreasing pattern and an even more characteristic increase around rounds 14 and 15. In Section 5.3.2 we discussed that surprising shifts could probably be explained by hypothetical endgame effects. This alternative hypothesis, however, cannot be directly tested in this data, as we did not ask subjects about endgame expectations. Instead, later in the multilevel analysis of individual decisions, we will handle these shifts as outliers from the general tendencies.

The light line in Figure 5.3.3.2 represents the overall mean rates of contribution and the dark line shows contribution rates in the low clustering condition. The latter could be the basis of comparisons with similar experiments, because in the low clustering condition, no new incentives were introduced. Because of structural breaks between parts, however, the comparison with other experiments is quite difficult.

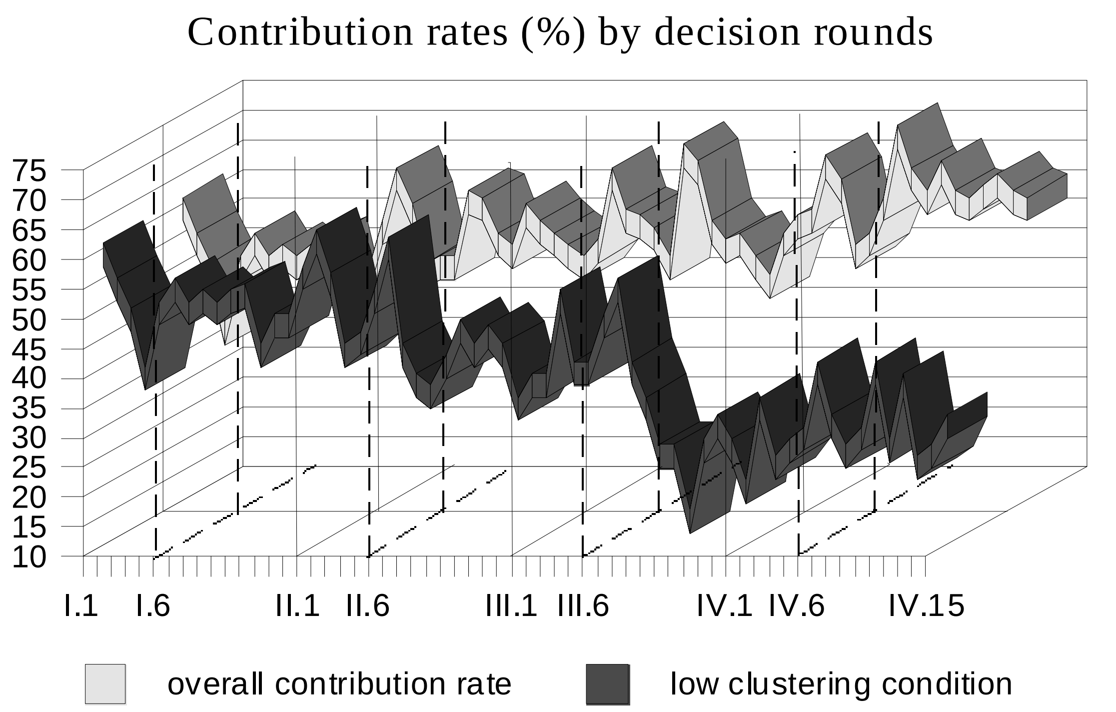


Figure 5.3.3.2 Contribution rates by decision rounds

Notes: Single-shot games are rounds 1-5 in each part. Results are shown only for the first 10 repeated games in each part, total N=14244 and N=6542 in the low clustering condition.

Figures 5.3.3.1 and 5.3.3.2 also show a significant rise in contributions in every part in the first repeated game. As we discussed in Section 4.3.4, this is possibly a result of reputation effects or image scoring (Nowak and Sigmund, 1998; Wedekind and Milinski, 2000; Bienenstock, 2001). It seems that subjects, in the hope of long term benefits, signal unselfish behavior towards group members in the beginning of the repeated games.

An average decreasing trend does not say much about whether there are *different scenarios in different parts and in different structural conditions* of the experiment. With regard to the *spiral of peace* hypothesis, we predicted that contribution rates would decrease over time in the low clustering condition until *stable peace* prevails. On the other hand, we anticipated *a spiral of conflict* in the medium and especially in the high clustering condition that would later lead to *durable conflict*. Looking at the data session by session we hoped to distinguish these typical scenarios in the experiment.

We found evidence for the *spiral of peace* in sessions of the *low clustering* condition. In all the six sessions there was a significant downward linear trend at least in one of the experimental parts ($p<0.05$). However, there are differences with respect to when did this occur. In one session, there was a significant decrease both in Part II ($p<0.05$) and in Part III ($p<0.001$) and the number of contributors remained stable at a low level in the last part (around two contributors). In another session, in which there was a significant decrease in Part I ($p<0.01$), the contribution rate was relatively low in Part II, but with some sudden peaks. Later, in Part III the contribution rate decreased again significantly ($p<0.01$) and it remained stable and low in Part IV. There were two sessions that produced an unexpected significant increase in Part II ($p<0.01$). In one, contribution rates dropped significantly in Part I ($p<0.01$) and remained stable and low in the first repeated rounds of Part II. At the end of this part, more and more subjects contributed, but the contribution rate fell back to a low level again in Part III and remained low with some sudden but short peaks. In the other case, there was no significant decrease in Part I, and the rise in Part II was followed by extreme fluctuations in Part III, until the contribution rate decreased significantly in Part IV ($p<0.05$). In the remaining two sessions no clear pattern could be recognized in the first two parts, but there was a significant downward trend in Part III ($p<0.05$ and $p<0.01$, respectively). In both cases, only a few people contributed in most of the rounds in Part IV, but there were some sudden and short peaks.

To summarize, it is hard to say that these scenarios look the same, but to a certain extent all sessions gave some support for the *spiral of peace* hypothesis. Common in all sessions of the low clustering condition is that in Part IV the contribution rate was already at a low level, below which there was no way to go. Some diehard contributors were not willing to change their decisions and eventually they provoked some fluctuations in the contribution rate.

Probably we have the clearest picture from the *medium clustering* condition. In this condition, there are no indications of any recognizable trends. There was neither a significant decrease nor increase in any part of the eight sessions. Contribution rates were oscillating around a certain contribution rate, which is probably an indication of a *ceiling effect*. To tell more about what is behind the stable or oscillating contribution rates, we will rely on the analysis of individual data.

There were some surprises in the results from the *high clustering* condition. There was no significant increase of contribution rates in any parts of the seven sessions. Consequently, we do not find support for the *spiral of conflict* hypothesis. *Contribution rates have risen between and not within the experimental part*s. In five of the seven sessions, as predicted, they stabilized at a high level (eight or nine contributors) in Part IV, which is a unique characteristic of the high clustering condition. In three sessions, this high level had already been reached in Part III. All these sessions belonged to the condition in which monetary selective incentives were introduced in Part III, which confirms the hypothesis that the *segregation effect is stronger under normative pressure than under confirmation pressure*. On the other hand, we found a significant decrease in Part II (twice $p<0.05$, once $p<0.01$) in three sessions, which is contrary to our predictions. In this part, minimal contact was introduced between the subjects, but no new incentives were distributed. A possible explanation for why contribution propensities dropped in some sessions is that subjects might have been highly influenced by internalized social control in the beginning of the experimental part, but they became disappointed in later rounds, as their nice behavior did not bring positive results.

In this section, we summarized the observed dynamics of the repeated games and related the results to our macro hypotheses about effects of temporal embeddedness and interactions of structural and temporal factors. We found evidence of scenarios of *stable peace* and of the *spiral of peace*. As we predicted, these scenarios appeared most likely in the *low clustering* condition. We also found evidence of *durable conflict*, especially in *segregated* structures. It did not mean, however, that one particular outcome was repeated. Moreover, contribution rates increased *between* experimental parts and therefore we could not trace the *spiral of conflict* scenario. The test about the mechanisms behind these macro processes should be based on the investigation of micro mechanisms that follows in the subsequent sections.

5.4 Conditional responses

5.4.1 Criticalness, reinforcement learning, and intergroup reciprocity

One of our major research questions concerned simple behavioral heuristics that guide individual decisions in the repeated IPG experiments. We predicted that *criticalness*, *reinforcement learning*, and *reciprocity* would play a crucial role here. In this section, we first look at how decisions were conditional on expectations and previous outcomes, which is a first step towards direct testing of our micro hypotheses.

Table 5.4.1.1. Average rates of contribution conditional on subjects' expectations about the outcome

expectation of subject	*previous outcome*	*criticalness hypothesis*	*contribution rate (from N cases)*
peace (*p*)	any	-	9.52% (2919)
defeat (*d*)	any	-	25.55% (1926)
victory (*v*)	*p, d, c*	+	77.03% (4166)
	victory (*v*)	?	78.38% (2433)
	no previous round	?	70.85% (590)
v total			*76.98% (7189)*
clash (*c*)	any	+	62.72% (2114)
total			53.93% (14148)

With respect to *criticalness*, we must analyze the relationship between expectations and decisions of subjects. Table 5.4.1.1 summarizes these statistics. As predicted, contribution rates are much lower than the overall mean, when the subject anticipated peace (9.52%) or defeat (25.55%) and they are higher, when the subject expected a clash (62.72%) or victory (76.98%). *This supports our criticalness hypothesis*. The hypothesis that contribution rates are the same after different expectations can be rejected (ANOVA $F(3, 14144)=2230.847$, $p<0.001$).

If victory is anticipated, individual decision is obviously not critical. We use the outcome of the previous round as an additional source of information to see whether the criticalness argument is appropriate to explain high contribution propensities. We believe that there is a higher chance that an individual perceives his or her decision as critical if the previous round was not victory (and he or she still believes in the success of the group). In this case subjective expectations are likely to be wrong, but actual decisions are not in contradiction with subjective expectations. Self-efficacy is illusionary (cf. Kerr, 1989), but decision is rational in expected utility terms. Subjects fell into this illusion quite often (N=4166).

If victory is anticipated, the criticalness argument is not supported by the data. Contribution rates are larger in case the previous round was victory (78.38% versus 77.03%; $t=1.277$, one-tailed $p=0.101$, not significant). It is still possible that subjects believed that their contribution was crucial to reach victory in the previous round and it would be exactly so in the next round. They did not have complete information about the outcome (number of contributors in the teams) and therefore on the basis of responsibility aversion (see Section 2.4.4) they might continue to contribute. This does not fully explain, however, the very high contribution rates in case victory was anticipated. These rates are even higher than when clash was expected ($t=12.254$, $p<0.001$). There should be something else besides criticalness that is responsible for this result, given that in the latter case the decision of the subject is certainly critical.

A possible explanation could be that people like to be part of the winning side and they experience this feeling more intensely in case they also made sacrifices for the team success. For some people, *the utility of such constructionism might outweigh the utility for free riding benefits*. A similar *bandwagon mechanism* was traced in the empirical analysis of mass legal political protests (Finkel, Muller, and Opp, 1989) and of voter turnout (Hong and Konrad, 1998). The crucial relevance of the bandwagon effect in our experiment has origins in the fact that subjects expected the victory of their team more than any other outcome (in 50.81% of all cases). This blind optimism together with a bandwagon effect substantially increased contribution rates.

We did not predict the influence of a bandwagon mechanism in advance, but we can formulate an *ex post hypothesis* about this effect.

(FORWARD-LOOKING) BANDWAGON HYPOTHESIS: *Subjects are more likely to contribute to the collective action of their group if they anticipate the victory of their group.*

A direct test of this hypothesis is left for subsequent research. In this study, we can only test bandwagon effects implicitly. If subjects value active participation to the winning side highly, we could predict that after victory, on average, contribution rates would increase in the winning team. In this team, both previous contributors and previous defectors are likely to have had high contribution rates. A problem with this implicit testing is that it assumes that forward-looking actors with bandwagon tendencies use past experience in the same way. To emphasize this difference, here we formulate a backward-looking hypothesis for which we could obtain independent support from the data.

(BACKWARD-LOOKING) BANDWAGON HYPOTHESIS: *Subjects are more likely to contribute to the collective action of their group if their group achieved victory in the previous round.*

Average contribution rates conditional on the previous outcome do not support this hypothesis (see Table 5.4.1.2). There is no clear evidence for backward-looking bandwagon effects. Victory in the previous round provoked contributions of relatively low proportion of defectors (43.8%). This proportion deviates significantly from the overall mean (t=3012.7, p<0.001).

Table 5.4.1.2. Effect of intergroup reciprocity and reinforcement learning on contribution rates in the entire experiment

		hypotheses		contribution rates		N		total (C+D)	
prev. decision		C	D	C	D	C	D	C rate	N
previous outcome	*p*	- ?	- -	51.15%	29.57%	954	2831	35.01%	3785
	d	+ -	+ +	61.62%	53.79%	1597	2043	57.23%	3640
	v	? +	? -	71.64%	43.84%	2870	771	65.75%	3641
	c	+ -	+ ?	69.24%	52.54%	1453	512	64.89%	1965
total				65.96%	41.30%	6874	6157	54.31%	13031

Notes: Among hypotheses, the first sign indicates hypothesis based on intergroup reciprocity and the second is based on reinforcement learning. For single-shot games and for the first repeated game the outcome of the last round in the previous part is considered as a previous outcome.

Defeat and clash were more efficient mobilizers of defectors, which is a sign that *supports our main hypotheses of intergroup reciprocity and reinforcement learning*. As Table 5.4.1.2 shows, contribution rates were always higher after a contribution choice than after a defection choice, which shows some stability of individual contribution propensities (the overall means 65.96% and 41.30% differ significantly; t=29.109, p<0.001). Contribution rates were highest if the previous round was won by the group (65.75%). In more than half of the cases subjects immediately retaliated the defeat or clash punishment of the group by contributing in the next round. Victory and clash reinforced the most contributing decisions. Defection choice was reinforced by peace. Altogether, mean contribution rates show support for the combined presence of intergroup reciprocity and reinforcement learning and to a less extent for bandwagon effects. A more sophisticated test in the form of multilevel logistic regression will follow in Section 5.6.

To provide a general overview of the relative strength of the criticalness, intergroup reciprocity, and reinforcement learning mechanisms on conditional contribution probabilities, we report mean contribution rates in Table 5.4.1.3 in the same structure as our main hypotheses were summarized in Table 4.3.2.2. For example, the first cell in the first row (expected peace and previous contribution column, *p*C) indicates that subjects contributed 52 times (16.72%) in those decision situations (N=311), in which they expected peace in the forthcoming round, they contributed in the previous round, and the previous outcome was peace.

Table 5.4.1.3. Contribution rates (%) in the *entire* experiment by subjects' expectations about the next round, previous decisions, and previous outcomes[2]

		subject's expectation								total	
prev.		peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)			
dec.		C	D	C	D	C	D	C	D	C	D
previous outcome	*p*	--? 16.7 (311)	--- 4.6 (1435)	--? 18.6 (70)	--- 15.7 (236)	+-? 77.7 (471)	+-- 69.4 (908)	+-? 56.8 (95)	+-- 43.3 (238)	51.2 (947)	29.6 (2817)
	d	-+- 12.6 (167)	-++ 13.0 (261)	-+- 38.2 (432)	-++ 22.6 (483)	++- 80.8 (730)	+++ 78.3 (969)	++- 78.7 (258)	+++ 59.5 (316)	61.7 (1587)	53.7 (2029)
	v	-?+ 15.9 (208)	-?- 7.9 (114)	-?+ 27.3 (205)	-?- 9.6 (94)	??+ 82.4 (2000)	??- 60.3 (431)	+?+ 69.5 (442)	+?- 45.8 (120)	71.6 (2855)	43.9 (759)
	c	-+- 19.1 (131)	-+? 15.0 (60)	-+- 32.0 (200)	-+? 22.6 (62)	++- 83.5 (783)	++? 68.9 (305)	++- 78.3 (327)	++? 41.7 (84)	69.3 (1441)	52.5 (511)
total		16.0 (817)	6.2 (1870)	32.9 (907)	19.3 (875)	81.8 (3984)	71.1 (2613)	73.1 (1122)	50.3 (758)	66.0 (6830)	41.3 (6116)
		9.2 (2687)		26.2 (1782)		77.6 (6597)		63.9 (1880)		54.3 (12946)	

Notes: Numbers of cases are in parentheses. In each cell, signs indicate predictions derived from criticalness, intergroup reciprocity, and reinforcement learning, in this order. For single-shot games and for the first repeated game the outcome of the last round in the previous part is considered as a previous outcome.

[2] The total number of cases differs from Table 5.4.1.2 because some expectations of subjects are missing.

A first conclusion we can draw from Table 5.4.1.3 is that none of the behavioral rules would predict these figures *alone*. The integration of different mechanisms in the explanation was certainly necessary to get a better understanding of subject behavior in the repeated IPG games. Based on these descriptive statistics, *it seems that the strongest effect is of the criticalness principle*. Contribution rates are immensely higher, if victory or clash is anticipated than if peace or defeat is expected, irrespective of the previous outcome. There is a significant difference after a previous contribution (t=46.452, p<0.001) and also after a previous defection (t=56.049, p<0.001). High differences in contribution probabilities after a contributing and after a defecting choice (t=29.047, p<0.001) show that contribution propensities are relatively stable over time.

Statistics in Table 5.4.1.3 also provide support for the *intergroup reciprocity* hypothesis. If we compare cells that have positive prediction signs for intergroup reciprocity in Table 4.3.2.2 with cells that have negative prediction signs (see also superscripts in Table 5.4.1.3), we find significantly higher contribution rates in cells with positive prediction signs (t=24.557, p<0.001). This difference is significant after a previous contribution (t=7.682, p<0.001) and also after a previous defection (t=18.336, p<0.001), thus we have some indication that reciprocity works in both directions. However, three columns of Table 5.4.1.3 show that there are some conditions, in which conditional contribution rates do not support the intergroup reciprocity hypothesis. First, when the subject anticipated peace and contributed in the previous round (*p*C column in Table 5.4.1.3), we predicted higher contribution rates after a defeat than after peace. However, we did not find this in the data. The second irregularity occurs, if the subject defected in the previous round and he or she anticipates victory for the subsequent round (*v*D column in Table 5.4.1.3). In this case, we predicted that contribution rates are higher after a clash than after peace. Probably subjects feel their contribution more substantial in the latter case, which outweighs the reciprocal intentions and might explain why our prediction failed. Finally, after a defection and an anticipated clash, we predicted a subject to contribute more likely, if the previous outcome was also clash than if it was peace (see *c*D column in Table 5.4.1.3). Mean contribution rates reveal just the opposite.

We have some support for the *reinforcement learning* hypothesis from the descriptive statistics. If we compare cells that have positive prediction signs for reinforcement learning in Table 4.3.2.2 with cells that have negative prediction signs (see also superscripts in Table 5.4.1.3), we find significantly higher contribution rates in cells with positive signs (t=20.247, p<0.001). This difference is significant after a previous contribution (t=5.305, p<0.001) and also after a previous defection (t=6.824, p<0.001). From this we could conclude that reinforcement works in both directions. However, what we have here is a classic case of *aggregation fallacy*. After a previous contribution choice, we predicted that reinforcement learning elicits contribution after victory and impedes contribution after a harmful clash. Considering overall statistics for these cases, we find that indeed, contribution rates were higher after victory

(71.6% versus 69.3%), although this difference was not significant (t=1.628, one-tailed p=0.052). Meanwhile, in all of the subcategories the relationship was in the other way around (see columns *p*C, *d*C, *v*C, and *c*C in Table 5.4.1.3)! After a clash, contribution rates were always higher than after victory. The misleading aggregated relationship was caused simply by the huge number of cases in the victory/victory/contribution cell (cf. Table 5.4.1.3). This means that *we do not find support for reinforcement after contribution choices, only for reinforcement after defection choices*. It seems that reinforcement learning does not work in a symmetric way.

In this section, we summarized mean contribution rates conditional on expectations of subjects about the next round, on previous decisions, and on previous outcomes of the IPG game. This way, we provided some insight in the underlying micro mechanisms that influence individual behavior in the experiment. Mean contribution rates provided support for our criticalness hypothesis, but with a necessary adjustment for forward-looking bandwagon tendencies. We found some support for the intergroup reciprocity hypothesis, but we found evidence of reinforcement only after defection choices.

5.4.2 Local reciprocity

This section focuses on the *local reciprocity* hypothesis that predicts that *subjects reciprocate the previous actions of their neighbors*. Hence, this hypothesis relates to a micro mechanism that depends on the local structure as well as on the shadow of the past. For this hypothesis, we have to consider contribution rates conditional on the previous decisions of neighbors. Since we are interested in the conditional choices and not in the previous neighbor decisions *per se*, choices that are implemented by the computer instead of the neighbor are also included in the analysis. If a subject did not make a decision in time, his or her choice was randomly determined. Neighbors did not know, in which cases decisions were real and in which cases they were implemented by the computer, therefore we can handle them equivalently.

We predicted that a previous contribution of a neighbor elicits contribution and a previous defection induces defection, irrespective of the group membership of the neighbor. Table 5.4.2.1 summarizes contribution rates depending on the number of contributing and defecting neighbors in the previous round.[3] The first cell in the first row is a special case. It indicates the mean contribution rate of three subjects, who had no neighbors because of some technical difficulties during the experiment (cf. Section 3.5.3). These subjects contributed 46 times (38.33%) out of 120 decision situations.

[3] Since no information was provided about the behavior of neighbors before round 7 in Part II, local reciprocity might have played a role only in decisions afterwards. We included only the latter cases in Tables 5.4.2.1, 5.4.2.2, and 5.4.2.3.

From mean contribution rates in other cells of the table it seems that *contributing choices of neighbors are reciprocated*. On the other hand, there was no convincing support that the presence of an additional defecting neighbor decreased contribution rates.

Table 5.4.2.1 Contribution rates (%) by number of *all neighbors*, who contributed and who defected in the previous round

		Number of defecting neighbors in r-1			
		0	*1*	*2*	*Total*
Number of	*0*	38.33 (120)	37.03 (721)	37.72 (1853)	37.56 (2694)
contributing	*1*	60.65 (958)	56.27 (3160)	-	57.29 (4118)
neighbors in *r-1*	*2*	72.68 (2734)	-	-	72.68 (2734)
Total		68.57 (3812)	52.69 (3881)	37.72 (1853)	56.13 (9546)

Notes: Numbers of cases are in parentheses. All cases, in which the previous decision of the neighbors was known, are included.

Table 5.4.2.2 Contribution rates (%) by number of *fellow neighbors*, who contributed and who defected in the previous round

		Number of defecting fellow nbs in r-1			
		0	*1*	*2*	*Total*
Number of	*0*	35.05 (3347)	55.52 (1403)	42.73 (344)	41.21 (5094)
contributing	*1*	71.65 (2744)	64.91 (778)	-	70.16 (3522)
fellow nbs in *r-1*	*2*	84.73 (930)	-	-	84.73 (930)
Total		55.93 (7021)	58.87 (2181)	42.73 (344)	56.13 (9546)

Notes: Numbers of cases are in parentheses. All cases, in which the previous decision of the neighbors was known, are included.

Table 5.4.2.3 Contribution rates (%) by number of *neighbors from the other group*, who contributed and who defected in the previous round

		N of defecting other group neighbors in r-1			
		0	*1*	*2*	*Total*
N of contributing	*0*	66.98 (2895)	56.30 (1746)	28.98 (1049)	56.70 (5690)
neighbs from the	*1*	64.69 (2634)	31.35 (874)	-	56.39 (3508)
other group in *r-1*	*2*	44.25 (348)	-	-	44.25 (348)
Total		64.61 (5877)	47.98 (2620)	28.98 (1049)	56.13 (9546)

Notes: Numbers of cases are in parentheses. All cases, in which the previous decision of the neighbors was known, are included.

To distinguish between *reciprocal effects towards fellow neighbors and neighbors from the other group*, Tables 5.4.2.2 and 5.4.2.3 show the same statistics separately for these two groups. As predicted, contributing fellow neighbors elicited contribution and defecting fellows decreased contribution in the subsequent round. However, contribution rates were lowest in the case of no fellow neighbors (35.05%), due to stable peace in the low clustering condition and shows that structural effects were stronger than local reciprocity. With regard to reciprocal effects towards neighbors from the other group, descriptive statistics show that contribution decisions were not reciprocated, but defection choices decreased contribution rates. For a more accurate test of local reciprocity, we will have to rely on the multivariate analysis of the data that follows in the subsequent sections.

In this section, we summarized mean contribution rates conditional on the previous decision of neighbors. In this way, we provided statistical evidence about the nature of local reciprocal mechanisms. We found that local reciprocity works only under certain conditions. Subjects on average reciprocated contribution decisions of fellow neighbors and defecting choices of neighbors from the opposite group.

5.4.3 Interactions of structure and time

After the discussion of results related to the local reciprocity hypothesis, let us return to the reinforcement learning hypothesis. Predictions derived from the principles of criticalness and intergroup reciprocity do not change during the experiment. On the other hand, *predictions about reinforcement learning change because of the introduction of new monetary incentives* (cf. Section 4.3.2). This might partly explain why we did not find convincing support for the reinforcement learning hypothesis from overall descriptive statistics (cf. Table 5.4.1.3). Below we will have a look at these statistics under different experimental conditions separately. In this way, we can also see whether the effect of criticalness and intergroup reciprocity changed in experimental parts.

Table 5.4.3.1 displays contribution rates in Part I of the repeated IPG game conditional on expectations of the subject, on previous outcome, and on previous decision. In the first part of the experiment there were no additional monetary rewards and the effects of internalized social control can also be excluded. Hence, these data can be used to test predictions of reinforcement learning as they were expressed in the corresponding cells of Table 4.3.2.2.

Mean contribution rates show support for the *reinforcement learning hypothesis*, if the subject *defected* in the previous round. In this case, as predicted, a defeat elicited a relatively higher willingness to contribute (51.42%) and peace and victory reinforced previous defection (30.00% and 41.24% respectively). However, the data strongly contradict predictions derived from reinforcement learning after a *contribution* choice. As in the overall figures (cf. Table 5.4.1.3), contribution rates are higher after a clash than after victory (see columns *d*C, *v*C, and *c*C), which has nothing to do with any kind of reinforcement principle. We cannot simply blame the introduction of new monetary incentives, because predictions fail even in the absence of new monetary incentives.

This conditional support for the reinforcement mechanism is fairly similar to Flache's findings (1996: 188-189). In his experiments, subjects who shirked at some point discovered the advantages of this backward-looking rule, unlike contributors, who tended to take it easy after their compliance decision.

Table 5.4.3.1. Contribution rates (%) in the *control condition* by subjects' expectations about the next round, previous decisions, and previous outcomes.

		subject's expectation									
	prev.	peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)		total	
	dec.	C	D	C	D	C	D	C	D	C	D
previous outcome	*p*	18.75	4.15	13.33	10.53	71.60	68.51	57.89	30.00	46.37	30.00
		(64)	(241)	(15)	(38)	(81)	(181)	(19)	(40)	(179)	(500)
	d	0.00	18.60	27.03	15.07	75.22	67.70	63.33	56.10	51.67	51.42
		(23)	(43)	(74)	(73)	(113)	(226)	(30)	(82)	(240)	(424)
	v	17.78	4.76	11.36	6.67	69.14	53.21	51.43	40.63	56.52	41.24
		(45)	(21)	(44)	(15)	(324)	(109)	(70)	(32)	(483)	(177)
	c	8.70	20.00	25.71	0.00	72.45	59.32	62.50	20.00	53.89	43.01
		(23)	(10)	(35)	(9)	(98)	(59)	(24)	(15)	(180)	(93)
total		14.19	6.67	21.43	11.85	71.10	64.35	56.64	43.79	53.33	40.28
		(155)	(315)	(168)	(135)	(616)	(575)	(143)	(169)	(1082)	(1194)
		9.15		17.16		67.84		49.68		46.49	
		(470)		(303)		(1191)		(312)		(2276)	

Notes: Numbers of cases are in parentheses. Only repeated games are included.

Table 5.4.3.2. Contribution rates (%) in the *low clustering condition* by subjects' expectations about the next round, previous decisions, and previous outcomes.

		subject's expectation									
	prev.	peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)		total	
	dec.	C	D	C	D	C	D	C	D	C	D
previous outcome	*p*	11.83	3.32	11.54	11.11	72.46	69.06	44.00	40.66	37.43	21.33
		(169)	(905)	(26)	(72)	(138)	(320)	(25)	(91)	(358)	(1388)
	d	5.41	5.97	26.00	16.28	56.52	76.24	63.16	52.70	37.71	47.79
		(37)	(67)	(50)	(86)	(69)	(181)	(19)	(74)	(175)	(408)
	v	7.46	2.78	30.43	6.67	71.91	69.74	65.43	43.48	56.18	43.33
		(67)	(36)	(46)	(15)	(235)	(76)	(81)	(23)	(429)	(150)
	c	0.00	0.00	18.18	0.00	65.12	68.18	36.36	57.14	42.50	44.19
		(15)	(10)	(11)	(4)	(43)	(22)	(11)	(7)	(80)	(43)
total		9.38	3.44	24.06	12.99	69.28	71.29	58.82	46.15	45.59	28.91
		(288)	(1018)	(133)	(177)	(485)	(599)	(136)	(195)	(1042)	(1989)
		4.75		17.74		70.39		51.36		34.64	
		(1306)		(310)		(1084)		(331)		(3031)	

Notes: Numbers of cases are in parentheses. Data is from Parts II, III, and IV. For single-shot games and for the first repeated game the outcome of the last round in the previous part is considered as a previous outcome.

With regard to the other behavioral mechanisms, we find overwhelming support for *criticalness* and slight support for *intergroup reciprocity* in Part I. For intergroup reciprocity there are also some contradicting cases. This might partly be a consequence of relatively unstable contribution propensities (contribution rates are sometimes higher after a defecting choice than after a previous contribution) and partly of low identification with the group.

In order to see the effect of minimal contact on conditional responses, consider the *low clustering* condition. In this condition, there were no changes between Part II and Parts III and IV. Hence, we can examine the effect of minimal contact on conditional response strategies in a longer perspective. As Table 5.4.3.2 shows, intergroup reciprocity was not strengthened over time in this condition. There is no support for the hypothesis that subjects reciprocated the previous collective action of the other group (cf. relatively low contribution rates in rows *d* and *c* of Table 5.4.3.2). On the other hand, expectations still have a huge impact on actual decisions (see totals in columns in Table 5.4.3.2). High contribution rates after an anticipated clash and victory show support for the combined relevance of criticalness and forward-looking bandwagon effects.

As in Part I, *data from the low clustering condition supports the existence of reinforcement after a defection choice. Contrary to the control condition and to the overall figures, there is indication of reinforcement also after contribution*. We could be satisfied with the confirmation of our hypothesis, but a difference between Part I and the low clustering condition is puzzling. Do subjects "learn to learn" only at later stages of the repeated games?

In this respect, it is especially interesting to look what happens in the medium and high clustering conditions in Parts II, III, and IV. In order to have a sufficient number of cases in the cells, we handle the medium and high clustering conditions together. Table 5.4.3.3 displays average conditional responses in Part II, when minimal contact was introduced without new monetary incentives. Table 5.4.3.4 shows conditional contribution rates in Parts III and IV after the introduction of monetary selective incentives and behavioral confirmation rewards. In clustered structures in these experimental parts, the predictions of reinforcement learning change (see Sections 4.3.2 and 4.3.4): they become conditional on the previous decision of neighbors. In general, because of additional monetary incentives, they shift towards the stabilization of the previous choice.

As average contribution rates show, *decisions stabilize over time* (see differences between C and D columns in Table 5.4.3.4). In the last parts, both contribution and defection choices are repeated more often than in Part II and much more often than in Part I. Similar to the low clustering condition, stabilization goes together with a process of "learning to learn," at least after a previous defection. While in Part II there is only slight support for reinforcement after defection, in later parts this support is more convincing. On the other hand, it seems that reinforcement does not work after contribution, even if we consider adjustments in the predictions because of new incentives.

Table 5.4.3.3. Contribution rates (%) in *Part II in the medium and high clustering conditions* by subjects' expectations about the next round, previous decisions, and previous outcomes.

		subject's expectation									
prev.		peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)		total	
dec.		C	D	C	D	C	D	C	D	C	D
previous outcome	*p*	28.26	5.48	41.67	22.22	79.56	62.79	57.14	49.15	64.13	38.21
		(46)	(146)	(12)	(72)	(137)	(215)	(28)	(59)	(223)	(492)
	d	7.50	15.00	43.90	23.36	85.62	79.90	68.75	44.00	62.34	52.35
		(40)	(60)	(82)	(107)	(146)	(209)	(48)	(50)	(316)	(426)
	v	20.00	13.33	16.67	9.38	76.87	65.66	60.71	72.41	65.34	48.95
		(45)	(30)	(48)	(32)	(402)	(99)	(56)	(29)	(551)	(190)
	c	18.18	16.67	17.14	37.50	79.61	78.57	68.29	61.54	63.60	63.96
		(22)	(12)	(35)	(16)	(152)	(70)	(41)	(13)	(250)	(111)
total		18.95	9.27	31.07	22.03	79.33	71.16	64.16	52.98	64.10	47.17
		(153)	(248)	(177)	(227)	(837)	(593)	(173)	(151)	(1340)	(1219)
		12.97		25.99		75.94		58.95		56.04	
		(401)		(404)		(1430)		(324)		(2559)	

Notes: Numbers of cases are in parentheses. For single-shot games and for the first repeated game the outcome of the last round in Part I is considered as a previous outcome.

Table 5.4.3.4. Contribution rates (%) in *Parts III and IV in the medium and high clustering conditions* by subjects' expectations about the next round, previous decisions, and previous outcomes.

		subject's expectation									
prev.		peace (*p*)		defeat (*d*)		victory (*v*)		clash (*c*)		total	
dec.		C	D	C	D	C	D	C	D	C	D
previous outcome	*p*	21.87	11.19	17.65	16.67	86.09	78.13	69.57	52.08	66.84	45.77
		(32)	(143)	(17)	(54)	(115)	(192)	(23)	(48)	(187)	(437)
	d	23.88	14.29	42.48	27.19	84.83	85.27	86.34	73.64	69.16	58.88
		(67)	(91)	(226)	(217)	(402)	(353)	(161)	(110)	(856)	(771)
	v	21.57	11.11	43.28	12.50	90.95	57.14	78.30	30.56	83.98	42.15
		(51)	(27)	(67)	(32)	(1039)	(147)	(235)	(36)	(1392)	(242)
	c	26.76	17.86	39.50	24.24	88.57	68.18	83.27	40.82	76.15	52.27
		(71)	(28)	(119)	(33)	(490)	(154)	(251)	(49)	(931)	(264)
total		23.98	12.80	40.79	23.81	88.91	75.65	81.79	56.38	77.09	52.16
		(221)	(289)	(429)	(336)	(2046)	(846)	(670)	(243)	(3366)	(1714)
		17.65		33.33		85.03		75.03		68.68	
		(510)		(765)		(2892)		(913)		(5080)	

Notes: Numbers of cases are in parentheses. For single-shot games and for the first repeated game the outcome of the last round in the previous part is considered as a previous outcome.

With regard to the other behavioral mechanisms, the presence of *criticalness* is generally confirmed by data from the medium and high clustering conditions. In combination with *forward-looking bandwagon effects*, criticalness can explain high contribution rates in case victory or clash is anticipated and can explain low contribution rates otherwise. On the other hand, we do not see any indication of *backward-looking bandwagon effects*, as contribution rates after a previous victory and defection are relatively low (42.15%). This is significantly lower than the average rate after defection (52.16%; t=1626.6, p<0.001). Furthermore, there is only limited support for the *intergroup reciprocity hypothesis*.

To cope with the *changing predictions of reinforcement learning*, we created two dummy variables. For one of them, we assigned a value of one in case the adjusted reinforcement learning predictions are in the positive direction. The other variable got a value of one, if the adjusted predictions are in the negative direction. For cases, in which we do not have clear predictions, because payoffs for the subject are at the aspiration level of zero, both variables have zero values. These two variables will be included in the multilevel analysis in the subsequent sections as predictors of reinforcement learning. Since in Part II internalized social control effects are difficult to numeralize, for this part we did not make adjustments in the predictions compared to those reported in Table 4.3.2.2.

Mean contribution rates in the cases of positive predictions of reinforcement learning are higher (69.39%; N=3662) than for negative predictions (47.06%; N=6880). The difference is highly significant (t=22.997, p<0.001). As predicted, contribution rates in cases with unclear predictions lay in between (52.13%; N=2494).

In this section, in relation to our main micro hypotheses, we presented statistics about mean contribution rates. First of all, we gained confirmation that we need to incorporate different behavioral mechanisms in a unified model in order to have a satisfactory explanation for individual decisions in the repeated IPG games. Some of the decision rules have obviously higher impact on individual choices than others. We have seen that forward-looking *criticalness* is a major factor, irrespective of previous decisions and of experimental parts. We also found support for *reinforcement learning after a defection choice*. However, it seems that reinforcement does not work in a symmetric way: we did not find confirmation for this hypothesis after a previous contribution. We have only experienced such a tendency in later parts of the low clustering condition. Similar to reinforcement after a defection choice in clustered structures, it seems that subjects used *reinforcement learning more in later stages of the experiment*. Besides, we have only slight support for the intergroup reciprocity and partial support for the local reciprocity hypotheses. For the surprisingly high contribution rates after an anticipation of victory, we formulated an *ex post* explanation that is based on a forward-looking *bandwagon* mechanism. It seems that people like to make contributions to the team success more than enjoying victory as free riders. On the other hand, based on the mean contribution rates, backward-looking bandwagon considerations do not influence subjects in the experiment. For a closer test of the main micro hypotheses, we will consider subjective perceptions of the behavioral mechanisms in the following section, based on results from a post-experiment questionnaire. We will turn to multilevel analysis of decision data in Section 5.6.

5.5 Questionnaire data

In this section, we report results of the post-experiment questionnaire that subjects had to fill in on their computer. Subjects received their payments after they completed the questionnaire. Because of some technical difficulties, we have data from 201 subjects (out of 203).

Table 5.5.1. Reported motivations of subjects during the experiment.

Categories of intentions (I wanted …)	*Major intention* N	%	*Ranks of categories* mean	median	*Factor loadings* 1	2	3
to get as much money as possible	76	37.8	1.99	2	-0.14	-0.04	-0.79
to avoid loosing	47	23.4	1.84	2	0.06	0.78	-0.24
to do the best for my team	47	23.4	2.21	2	0.08	0.71	0.35
to look trustworthy for others	14	7.0	2.91	3	0.47	0.50	0.17
to do the same as my neighbor did	9	4.5	3.05	3	0.91	0.07	-0.06
to help my neighbor(s)	5	2.5	3.32	3	0.89	0.11	0.14
to do harm for the other team	0	0	4.48	5	-0.02	0.03	0.68
to initiate and break the deadlock	1	0.5	-	-	-	-	-
to protest against others' behavior	1	0.5	-	-	-	-	-
an outcome to fulfill my forecast	1	0.5	-	-	-	-	-

Notes: Categories with italics are self-reported to the open "other" category. 1=very important, 5=not important at all. Factor loadings are obtained from an explorative factor (principal component) analysis. Rotated varimax solution with Kaiser normalization. N=201.

Some of the questions concerned the motivations of subjects for their decisions in the experiment. Self-reported evaluation of different intentions can be biased, but it gives an indication of which factors were consciously taken into account at the decisions. Table 5.5.1 displays frequencies of answers for a question that asked about the main intention during the experiment. Seven categories were created in advance, three subjects reported a different main motivation to an open "other" category. These answers are also included in Table 5.5.1. Additionally, subjects were asked to mark the importance of the seven created categories separately on a five-point scale (from 1=very important to 5=not important at all). Means and medians for these questions are also shown in Table 5.5.1.

Most subjects felt monetary incentives the most important in the experiment. On average, loss aversion was ranked as of the highest importance (mean rank 1.84). These two motivations together with intentions to help the team were mentioned as the three most important driving forces in the experiment. Surprisingly, neighborhood effects were ranked quite low by the subjects, although the lowest evaluation was received by the competitive category of doing harm for the other team (mean rank 4.48). A null hypothesis that the mentioned motivations are evaluated in the same way can be rejected (Kendall W=0.45; $\chi^2(6)$=542.47, p<0.001). With regard to the importance of neighborhood effects, they are evaluated significantly lower than the three leading motivations. The hypothesis that the best ranked imitation of neighbors is at least equally important as doing the best for the team can be rejected (Wilcoxon signed rank test Z=6.646, p<0.001).

It seems that the intentions of subjects cannot be put on a unidimensional scale. Such a scale has a reliability of α=0.56. This relatively low reliability is mainly caused by subjective perceptions of monetary rewards that have negative correlations with the ranks of other intentions. By leaving the item "I wanted to get as much money as possible" out of the scale, reliability would jump up to α=0.65. The number of subjects is not sufficient to conduct a powerful multidimensional scaling or factor analysis. However, for illustrative reasons, we report factor loadings in Table 5.5.1 that are results of an explorative factor analysis. Three underlying factors have been found in the analysis. The first factor explains 33.01% of the variance and can be interpreted as *local attachment* with positive loadings for neighborhood specific concerns. It is more difficult to find a meaningful interpretation for the second factor that increases explained variance by 17.23%. Possibly it is related to *responsibility aversion* in a broader sense. This includes risk aversion and group specific concerns (importance of helping the team, trustworthiness) with positive loadings. Adding a third factor increases explained variance by 15.03%. This factor can be conceptualized as a *utilitarianism* scale with negative loadings for monetary concerns. Despite the low number of subjects and items, it seems that subjective intentions can be nicely fitted in a three-dimensional space, in which neighborhood concerns, group concerns, and monetary concerns constitute the different dimensions.

Subjects were also asked to evaluate procedural changes during the experiment. The evaluation of minimal contact and new monetary incentives was relevant for the discussion of single-shot games and therefore we analyzed these questions (see Section 3.6.2). In comparison to other procedural changes, the introduction of minimal contact and new monetary incentives were evaluated as less important than information about the outcome of the previous round (mean=1.791, median=2) or information about the decisions of the neighbors (mean=1.920, median=2). The difference is highly significant (comparing the latter with minimal contact, the Wilcoxon signed rank test provides Z=8.299, p<0.001). It is interesting to see that while subjects expressed no particular relevance for imitating or helping their neighbors, they highly appreciated information about their decisions.

Another group of questions was related to conscious application of certain simple strategies that are conditional on the past. These were simple yes or no questions related to whether subjects considered the given strategy during the experiment. Reinforcement learning was consciously taken into account by many subjects, but mainly only for repeating successful choices and less frequently for shifting from unsuccessful choices (see Table 5.5.2). Intergroup reciprocity was reported to be important after a defeat and clash. Many subjects were willing to retaliate nasty actions of the other team, but quite few of them appreciated nice actions of the other side. Local reciprocal strategies were considered by half of the subjects, but substantially more of them felt pressure from the direction of their fellow neighbors.[4] On the other hand, almost nobody reported pressure from the direction of a neighbor of the opposite team. Besides, many subjects claimed to consider a backward-looking bandwagon effect.

Table 5.5.2. Reported conditional strategies of subjects in the experiment.

behavioral mechanism	*conditional strategy asked*	*%*
reinforcement learning	win-stay	64.2
	loose-change	48.3
changes in reinforcement	win new incentives-stay	70.7
	no win of new incentives-change	37.3
	defeat-contribute	73.6
intergroup reciprocity	clash-contribute	74.6
	clash-defect (inverse)	26.9
	peace-contribute (inverse)	65.2
local reciprocity	imitation (repetition) of neighbor decision	54.7
internalized social control	traitor pressure	6.5
	internalized selective incentives	72.1
bandwagon effect (backward)	victory-contribute	66.0
	victory-defect framed in long term benefits	23.9
other	victory-defect framed in free riding benefits	32.8
	random	18.4

Note: Proportion of subjects who considered the given strategy, N=201.

Questionnaire data can provide insights whether subjects developed a certain *identification* with other participants during the experiment. To get some indication of this, we asked subjects about how would they divide 100 guilders of their hypothetical gain between them and certain other subjects. Answers show a strong identification with fellow team members. Fellows who were not neighbors would get significantly higher proportions of the money than would members of the other team (13.30 guilders

[4] Subjects, who had no fellow neighbors were asked about a “hypothetical” fellow neighbor.

versus 5.69 guilders; t=8.420; p<0.001). On average, non-neighbor fellows were rewarded even more than neighbors (12.25 NLG for left neighbors, 12.38 NLG for right neighbors). However, it really made a difference to which team the neighbor belonged. Strong bonds evolved towards fellow neighbors (16.49 guilders for left fellow neighbors, 17.97 guilders for right fellow neighbors) and weak bonds towards neighbors from the other team (9.01 NLG and 8.19 NLG respectively). The difference is significant for both left (t=3.761, p<0.001) and right neighbors (t=4.737, p<0.001). An attachment to a neighbor from the other team was still stronger than to other members of the opposite team.

In this section, we summarized results from the post-experiment questionnaire. We tried to show how subjects perceived and evaluated their decisions in the repeated IPG game. Most subjects reported that their major intention was to receive as much money as possible. They consciously took into account reinforcement of successful choices, although data about actual behavior shows the contrary (see previous section). Questionnaire data also supports that subjects consciously applied hawkish intergroup retaliations and intended to join the effort of others after victory.

5.6 Results of the multilevel analyses

In this section we will test our main micro hypotheses by using multilevel logistic regression analysis. The exact specification of the models can be found in the appendix to Chapter 4. We tested two types of models. In the first type we fixed the size of prediction parameters over subjects. In the second type we allowed a random variation of the slopes of main explanatory variables. We assumed that these random variations follow a normal distribution around their mean. The analyses have been done by using the MLwiN 1.02 software (Rasbash et al., 1998).

We included all cases in the analysis, when a subject made a real decision and excluded all decisions that are implemented by the computer. For handling missing values for independent variables, we used imputation of the mean for variables on interval scale and we imputed the median for categorical variables. Exceptions were dummies concerning the outcome of the previous round: they were all given a zero value (there was no peace, no defeat, no victory and no clash in the previous round). For predictions about reinforcement learning two dummy variables were used. One of them got a value of one in case of positive predictions and the other one for negative predictions. For the first single-shot games and for cases with no clear predictions both variables have zero values.

It should be noted that all likelihood statistics are rough approximations. Since the number of random components in the model is limited, in the second type of models we restricted all random covariance parameters to zero. Hence, only random variances of the slopes are estimated. For testing these random effects we used deviance tests instead of *t*-tests (cf. van Duijn, van Busschbach, and Snijders, 1999: 192-193).

Incorporating all independent and control variables in the analyses results in an extensive model. In such extensive models, testing several null hypotheses and conducting a huge amount of *t*-tests simultaneously can only be done with reservations (Cohen, 1990). The incremental validity of new variables is very low and the predicting power of the model does not increase substantially, although improvement statistics provide significant values. These limitations are only of secondary importance, because driven by theoretical arguments, we *should* include these variables. We need controls to determine what are the real explanatory factors behind individual decisions. This is the only way to judge our main micro hypotheses correctly.[5]

5.6.1 The effect of decision heuristics

In this section, we report results of a multilevel logistic regression analysis that included all of our main independent variables. Table 5.6.1.1 includes estimates of two models, one that includes only fixed effects of main variables and one that allows for a random variation of slopes. Log likelihood statistics of the first model are compared to a baseline model that includes only a constant with interindividual and intersession variation.[6] None of the models in Table 5.6.1.1 show variation of contribution propensities between the sessions. It means that we did not find any indication of session specific scenarios or session specific variation of contribution propensities. The variation between sessions can be entirely attributed to effects of our main explanatory variables. In this way, there are no substantial differences between results from a two-level (decisions-subjects) and from a three-level (decisions-subjects-sessions) model. Still, grounded on the original theoretical argument that explanatory variables are not likely to cover all intersession variation, we will report results from the three-level model.

As in the case of single-shot games, most hypotheses about social control effects are supported by the data. *Social selective incentives and behavioral confirmation both in an internalized and in a monetary form have a significant effect on individual decisions*. Fellow neighbor pressure is a strong determinant of decision. *Internalized forms of social control are not as strong as direct social control, but their effects are still significant except the effect of internalized traitor rewards*. As a macro consequence of social control effects, conflict is more likely in segregated structures.

[5] Strictly speaking, in our analyses we do not test hypotheses as “the given variable has a positive effect on contribution propensities”. Instead, as conventional, we test null hypotheses as “the given variable has no effect on contribution propensities”.

[6] The baseline model provides a parameter estimate of 0.177 (0.124) for the α_0 baseline contribution rate, 0.432^{+++} (0.051) for interindividual variance and 0.270^{+++} (0.099) for intersession variance. The -2 Log Likelihood of the baseline model is 18560.7.

Table 5.6.1.1. Results of multilevel logistic regression on contribution propensities

independent variable	hypothesis about the direction of effect	model with fixed slopes	model assuming random slopes
FIXED EFFECTS			
α_0 baseline contribution propensity	?	.745*** (.082)	.749*** (.083)
STRUCTURAL EMBEDDEDNESS			
s_0 internalized selective incentives	+	.194*** (.046)	.188*** (.057)
s_1 direct selective incentives	+	.573*** (.063)	.531*** (.082)
b_0 internalized behavioral confirmation	+	.245*** (.039)	.215*** (.048)
b_1 direct behavioral confirmation	+	.611*** (.061)	.632*** (.083)
t_0 internalized traitor rewards	-	-.006 (.040)	.016 (.042)
TEMPORAL EMBEDDEDNESS			
Criticalness			
p anticipated peace	-	-2.722*** (.097)	-2.594*** (.144)
d anticipated defeat	-	-2.038*** (.090)	-2.036*** (.111)
c anticipated clash	+	-.716*** (.079)	-.742*** (.089)
v anticipated victory (not victory in r-1)	+	.268** (.091)	.270** (.099)
Intergroup reciprocity			
p peace in r-1	-	-.151 (.092)	-.115 (.095)
d defeat in r-1	+	.156 (.088)	.149 (.092)
c clash in r-1	+	-.019 (.096)	-.041 (.099)
Reinforcement learning			
predicted increase of contribution propensity	+	-.110 (.068)	-.135 (.073)
predicted decrease of contribution propensity	-	-.212*** (.059)	-.181** (.062)
Local reciprocity			
t_1 towards opposite neighbors	+	.003 (.030)	-.013 (.030)
b_2 towards fellow neighbors	+	.119** (.045)	.114* (.047)
b_3 towards fellow neighbors, if b_1 is introduced	+	.043 (.064)	.018 (.068)
RANDOM EFFECTS			
interindividual variance		.625^{+++} (.071)	.601^{+++} (.080)
intersession variance		.000 (.000)	.000 (.000)
variance of s_0 internalized selective incentives			.075^{+++} (.036)
variance of s_1 direct selective incentives			.117^{++} (.067)
variance of b_0 internalized behavioral confirmation			.070^{+++} (.029)
variance of b_1 direct behavioral confirmation			.335^{+++} (.091)
variance of t_0 internalized traitor rewards			.000 (.000)
variance of p anticipated peace			1.829^{+++} (.301)
variance of d anticipated defeat			.588^{+++} (.146)
variance of c anticipated clash			.214^{++} (.086)
variance of v anticipated victory (not v in r-1)			.191^{+++} (.071)
variance of p peace in r-1			.000 (.000)
variance of d defeat in r-1			.001^{++} (.055)
variance of c clash in r-1			.002^{+} (.077)
variance of positive reinforcement			.078 (.061)
variance of negative reinforcement			.043 (.045)
variance of local reciprocity t_1 towards opposite neighbors			.000 (.000)
variance of local reciprocity b_2 towards fellow neighbors			.007^{+} (.019)
variance of local reciprocity b_3 towards fellow neighbors			.000 (.000)

cont. next page

-2 Log Likelihood model	10523	10363
Improvement χ^2 (df) for model in right column		160*** (17)
vs. previous model	8037.7*** (17)	

Notes: N=14244 decisions for 203 subjects in 21 experimental sessions. Numbers in parentheses are standard errors. Iterative Generalized Least Squares estimates. * significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{+}$ significant at the 5% level, $^{++}$significant at the 1% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes)

With regard to hypotheses about temporal embeddedness, analysis shows that effects of all variables that are related to expectations of subjects about the forthcoming round are highly significant. As we predicted, contribution rates decrease when peace or defeat is anticipated. Contribution rates significantly increase when victory is expected, but not experienced. *This supports our criticalness hypothesis*. However, an anticipation of a clash decreases contribution rates (estimates are -0.716 and -0.742 in the two models), which is contrary to our predictions. This finding is related to high contribution rates after the anticipation of victory (77.6%). The reference category for the criticalness dummy variables is the case in which victory is experienced and anticipated. In this case, contribution rates are unexpectedly high; estimates for other variables therefore are all biased in the negative direction. These high rates after the anticipation of victory cannot be entirely explained by the principle of criticalness (cf. Section 5.4.1). Therefore we offered a supplementary hypothesis. We claimed that subjects like to take a constructive part from the victory of their side and therefore they join the bandwagon and leave behind free rider benefits. Criticalness combined with this forward-looking bandwagon effect explains huge negative effects of anticipated peace and defeat and justifies the surprisingly negative effect of clash anticipations. *Consequently, if we control for bandwagon tendencies, we can confirm the criticalness hypothesis.* However, a direct test of forward-looking bandwagon effects is not available, because if we include anticipated victory as a dummy among the predictors, we would run into a perfect multicollinearity of independent factors.

There is *less support for the intergroup reciprocity hypothesis*. A previous result of peace and defeat drives in the predicted direction, but controlling for other main factors, these effects are not significant. Moreover, the sign of the parameter estimate of the effect of previous clash is in the opposite direction than predicted (-0.019 and -0.041). These results are also related to the perceptions of subjects about victory. When the team won in the previous round, subjects contributed more than what we could derive from the intergroup reciprocity principle. Since this was the reference category, the effects of dummy variables concerning other outcomes than victory in the previous round all shifted in the negative direction. We do not get significantly positive parameters for the effect of previous clash and defeat, because we do not control for a possible *backward-looking bandwagon effect*. It is more surprising that the effect of previous peace is not significant, which contradicts the intergroup reciprocity hypothesis even if we control for possible backward-looking bandwagon tendencies.

There is *firm support for reinforcement in the negative direction, but there is no indication for generating contribution by reinforcement learning*. As we have seen by looking at conditional contribution rates, reinforcement learning works asymmetrically in the experiment. We found that the reinforcement principle was more likely to be activated *after* a previous defection (see Section 5.4.1). The analysis in Table 5.6.1.1 shows that reinforcement works if it drives *towards* defection. The repetition of a beneficial defection choice is much more likely than the repetition of a successful contribution decision. Estimates for the dummy that indicates when contributions are predicted based on reinforcement, have surprisingly negative signs (-0.110 and -0.135). This indicates that contribution rates in the reference category are *ceteris paribus* higher than in cases, when contribution could be expected based on the reinforcement principle. This result does not give support for the reinforcement learning hypothesis, even if the reference point is not at zero payment. *We have to reject the hypothesis that reinforcement enters the considerations of subjects as a universal principle*.

We have to reach the same conclusion with regard to local reciprocity. *The hypothesis that subjects reciprocate behavior of their neighbors from the other group is not supported by the data*. On the other hand, *we find confirmation of reciprocation or imitation of the decisions of fellow neighbors*. Honoring local reciprocal behavior by new monetary incentives in Parts III and IV did not have an additional effect.

Since most independent variables are dummies, comparison of effect size is not meaningless. Expectations of subjects make the largest impact on actual decisions in the manner that is likely to be a *combined effect of criticalness and forward-looking bandwagon principles*. Monetary social control has also a very strong influence on individual contribution choices. The relative strength of these and other effects can provide a basis for an *ex post* estimation of the importance of different principles. In a follow-up study, these results can be used to determine assumptions about weights of these behavioral mechanisms in individual decision making that can consequently lead to more accurate predictions of intergroup processes.

The second model in Table 5.6.1.1 allows for a random variation of the slopes of the main variables. Meanwhile the number of parameters to be estimated is almost the double in comparison to the first model, there are no big changes in the parameter values or in their significance. It is also not much of a surprise that we found significant random variations for the slope of those variables that have highly significant fixed effects. There is a disparity between subjects in the extent to which they rely on criticalness and bandwagon principles and in the magnitude they are influenced by different forms of social control. On the other hand, we cannot claim that any of the insignificant effects were caused by the diversity of the strength and direction of these variables on different subjects.

To summarize, our multilevel analyses that contained the main explanatory variables provided support for most of our micro hypotheses about the effects of structural embeddedness. Social control of fellow neighbors both in an internalized and in a monetary form had a significant effect on individual decisions. On the other hand, our micro hypotheses about the effects of temporal embeddedness received only partial support. The criticalness hypothesis had to be adjusted for a bandwagon tendency, there was no indication of intergroup reciprocity, and reinforcement seemed to work only, if it drives towards defection. In addition, the local reciprocity hypothesis also received conditional support, as subjects reciprocated actions of their fellow neighbors only.

5.6.2 The effect of personal characteristics and other control variables

In this section, we first extend the multilevel analysis of Section 5.6.1 by including certain subject-level characteristics. Individual decisions in the repeated IPG game cannot be assumed as independent from certain attributes of the subjects. Second, there are effects of temporal embeddedness that are not embraced by our main independent variables. These effects might have a significant impact on contribution propensities. Moreover, in case of close associations, their omission might cause bias in the parameter estimates of the main independent variables. For this reason, we include these control variables in the subsequent models.

Table 5.6.2.1 summarizes the effects of control variables and displays the results of the multilevel analyses. As results show, the inclusion of personal characteristics in the model did not cause any substantial change in the parameters of our main independent variables. Significant effects remained significant and insignificant effects remained insignificant. There are no big differences even in the size of parameter estimates.

The *unimportance of personal characteristics* for individual decisions is persuasively demonstrated by the analysis. None of the subject-level variables has a significant effect on contribution propensities. This includes the insignificant *gender* effect. The differences in mean contribution rates (55.23% versus 52.19%) in favor of women can be explained by effects of structural and temporal embeddedness. Similarly, the difference in contribution rates of subjects who already graduated (59.57%) and students (53.46%) was ruled out by our main independent variables.

As in the single-shot games, *study direction* did not have a significant effect. Students from the Faculty of Economics (that does not include business and spatial sciences) had even the highest contribution rate (58.63%) in comparison with students from all other faculties. This also means that students of economics behaved differently in the single-shot and in the repeated games. This is probably an indication that they recognized the difference between the two, and that they were thinking more likely in equilibrium terms than others in the experiment. Still, controlling for other effects, training in economics had no significant effect. On the other extreme, students of spatial sciences and law had the lowest contribution rates (42.88% and 46.07%, respectively). These are also insignificant predictors in the multilevel regression model. Another sign that questions the importance of previous *experience* is that participating in a similar experiment before did not have a significant influence on contribution propensities. The difference in mean contribution rates (52.43% if a subject did take part in a similar experiment before and 55.02%, if he did not) can be attributed to other effects.

Risk preferences also turned to be insignificant predictors. The only personal characteristics we measured and found significant in explaining individual decisions in the single-shot games were *social orientations*. However, they had no significant effects on decisions in the repeated IPG games. The reason for this striking difference between single-shot and repeated games could lie in the relationship between temporal embeddedness and social orientations. In case expectations of prosocials and proselfs differ, the net effect of social orientations can disappear. To test this *ex post* hypothesis, we will return to the analysis of the relationship between social orientations and expectations in Section 5.7. Another possible reason is that prosocial (and egalitarian) orientation is no longer directed so closely towards in-group members. Feedback information may allow prosocials to feel sympathy also towards out-group members.

Table 5.6.2.1. Results of multilevel logistic regression on contribution propensities with personal characteristics

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes	model with random slopes
FIXED EFFECTS			
α_0 baseline contribution propensity	?	.886* (.361)	.737* (.375)
STRUCTURAL EMBEDDEDNESS			
s_0 internalized selective incentives	+	.186*** (.046)	.184** (.056)
s_1 direct selective incentives	+	.596*** (.064)	.544*** (.084)
b_0 internalized behavioral confirmation	+	.252*** (.039)	.222*** (.048)
b_1 direct behavioral confirmation	+	.612*** (.061)	.628*** (.086)
t_0 internalized traitor rewards	-	-.001 (.040)	.024 (.043)
TEMPORAL EMBEDDEDNESS			
Criticalness			
p anticipated peace	-	-2.744*** (.098)	-2.612*** (.143)
d anticipated defeat	-	-2.059*** (.091)	-2.058*** (.112)
c anticipated clash	+	-.721*** (.080)	-.749*** (.091)
v anticipated victory (not v in r-1)	+	.271** (.092)	.276** (.100)
Intergroup reciprocity			
p peace in r-1	-	-.159 (.093)	-.122 (.096)
d defeat in r-1	+	.156 (.089)	.150 (.092)
c clash in r-1	+	-.018 (.097)	-.039 (.100)
Reinforcement learning			
predicted increase of contribution propensity	+	-.114 (.068)	-.142 (.074)
predicted decrease of contribution propensity	-	-.215*** (.059)	-.186** (.062)
Local reciprocity			
t_1 towards opposite neighbors	+	.001 (.030)	-.016 (.030)
b_2 towards fellow neighbors	+	.122** (.046)	.119* (.048)
b_3 towards fellow neighbors, if b_1 is introduced	+	.043 (.065)	.012 (.068)

cont. next page

PERSONAL CHARACTERISTICS		
gender (1=male)	-.028 (.129)	.018 (.135)
student at the university (1=yes)	.086 (.335)	.233 (.349)
studies at the law faculty	-.336 (.333)	-.472 (.351)
studies natural sciences	-.025 (.313)	-.230 (.328)
studies economic, business, or spatial sciences	-.255 (.305)	-.428 (.319)
studies social sciences	-.122 (.282)	-.215 (.295)
student of literary studies or arts	-.127 (.290)	-.354 (.302)
did a similar experiment before	-.124 (.124)	-.134 (.129)
strong risk aversion towards gains	-.080 (.122)	-.046 (.127)
strong loss aversion	-.022 (.120)	-.054 (.126)
consistency on social orientation questions	.005 (.163)	.012 (.170)
prosocial orientation	.297 (.165)	.320 (.172)
egalitarian orientation	.261 (.159)	.268 (.166)
number of acquainted subjects in the experiment	-.022 (.080)	-.011 (.083)
quiz questions answered correctly %	-.002 (.003)	.000 (.004)
RANDOM EFFECTS		
interindividual variance	.582^{+++} (.071)	.534^{+++} (.074)
intersession variance	.001 (.023)	.000 (.000)
variance of s_0 internalized selective incentives		.071^{+++} (.035)
variance of s_1 direct selective incentives		.124 (.069)
variance of b_0 internalized behavioral confirmation		.075^{+++} (.031)
variance of b_1 direct behavioral confirmation		.367^{+++} (.096)
variance of t_0 internalized traitor rewards		.000 (.000)
variance of *p* anticipated peace		1.742 (.293)
variance of *d* anticipated defeat		.600^{+++} (.148)
variance of *c* anticipated clash		.252^{+++} (.091)
variance of *v* anticipated victory (not *v* in *r-1*)		.193^{+++} (.071)
variance of *p* peace in *r-1*		.000 (.000)
variance of *d* defeat in *r-1*		.000 (.000)
variance of *c* clash in *r-1*		.000 (.000)
variance of positive reinforcement		.101^{++} (.063)
variance of negative reinforcement		.027 (.044)
variance of local reciprocity t_1 towards opposite neighbors		.000 (.000)
variance of local reciprocity b_2 towards fellow neighbors		.009^{++} (.019)
variance of local reciprocity b_3 towards fellow neighbors		.000 (.000)
-2 Log Likelihood model	10372	10269.2
Improvement χ^2 (df) for model in right column		102.8*** (17)
vs. previous model	151*** (15)	93.8*** (15)

Notes: N=14244 decisions for 203 subjects in 21 experimental sessions. Numbers in parentheses are standard errors. Iterative Generalized Least Squares estimates. * significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{+}$ significant at the 5% level, $^{++}$significant at the 1% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes).

Besides subject-level variables, we have to control for certain *procedural effects* and for influences of *temporal embeddedness* other than our main independent variables. Before extending the multilevel model with these predictors, let us briefly report about a procedural variable that will not be included in the multilevel analysis. This is the *color label* of the team (red or green). Even though some scientists working on visual perceptions would disagree, we presumed that the color of the flag on the computer has nothing to do with actual choice. After color labels were introduced in Part II, members of the red team contributed in 55.47% of all decisions and members of the green team contributed in 56.40% of all cases. The null hypothesis that contribution rates are independent from color labels cannot be rejected (t=0.966, two-tailed p=0.334). The difference was higher in the control condition (46.55% versus 49.00%), which indicates that subjects, who were assigned into the green team had higher baseline contribution rates by pure chance, although the difference is not significant (t=1.452, two-tailed p=0.146). Color labels did not make a significant difference for individual decisions indeed.

Table 5.6.2.2 reports the results from two multilevel analyses. Procedural effects together with other control variables that are related to temporal effects are incorporated in these models. Although the inclusion of these variables resulted in significant improvement statistics, there was no substantial enrichment in the explanatory power of the model. Still, unlike subject characteristics, some temporal effects had a significant influence on contribution propensities.

The introduction of these variables resulted in two major changes with regard to testing our main independent variables. *First*, defeat in the previous round increases contributions significantly. It means that subjects retaliate the severe loss caused by the other group. On the other hand, other elements of intergroup reciprocity are not significant. For instance, a previous clash *ceteris paribus* does not generate more contributions at all (estimates are -0.018 and -0.032). Since the retaliation of defeat was also insignificant in the previous models, we have to consider why it has become significant in this extended analysis. The possible main reason is controlling for different trends in the experiment. Subjects become more familiar with the payoff structure during the experiment, which results in a natural tendency towards defection in case keeping the bonus is beneficial. On the other hand, it also might take time until players realize that newly introduced incentives make contribution more attractive. These linear tendencies might demolish some effects of intergroup reciprocity in case we do not control for them in the analysis. For instance, contribution rates of subjects in the defeated team does not increase nor decrease significantly, because retaliating intentions are in conflict with motivations to defect as a result of equilibrium concerns that are learned during the game. If we leave one of these variables out of the analysis, the parameter estimate for the other one will be biased.

Second, internalized traitor rewards have a significant effect in the model that assumes variation in the slopes of main independent variables. However, this effect is in the opposite direction from that predicted (0.104 and 0.135). Estimates for internalized traitor rewards behaved strangely also in the single-shot game. The pattern is similar here: in our parsimonious models that contain only the main explanatory variables, the estimate is in the right direction or close to zero. However, when we control for time effects, the estimate becomes positive, meaning that the presence of a neighbor from the opposite group *increases* contribution propensities. A possible explanation for this can be based on the silent identification effect (Bohnet and Frey, 1999). Visual contact between the subjects decreases social distance between them and leads to empathy and to some sort of silent identification. As a consequence of silent identification, after the inclusion of a between part trend variable, our analysis overestimates the effect of internalized selective incentives, but also shifts the estimate of internalized traitor rewards in the positive direction (see more in Section 3.6.5). Another indication for the presence of silent identification with neighbors of the opposite group is provided by questionnaire data. Subjects in the hypothetical dictator games were more generous towards members of the opposite group, who were neighbors, than towards other out-group members (see Section 5.5).

Results show that *most control variables related to temporal effects had no significant influence* on individual contribution rates. Insignificant effects include session-level variables as delay time at the start of the experiment and length of play. Length of play has no net effect also because of the inclusion of trend variables in the analysis. We found a clear downward trend in the first two experimental parts. In these parts, the IPG game was played in the original payoff structure, without additional monetary incentives. In this setting, it seems that subjects learned over time that keeping the bonus is more beneficial than contributing. We found another significant trend in the random slope model for the medium clustering condition in Part III when monetary rewards were introduced for behavioral confirmation. This is somewhat surprising, because other trend variables (for instance, the trend within the high clustering conditions in Part III) were not significant.

Besides, the analyses show that all peaks we experienced in contribution rates over time (see Figures 5.3.3.1 and 5.3.3.2) are the results of other temporal effects. These fluctuations, including the local peaks of contribution around rounds 5, 10, and 15, *can be derived as consequences of the main behavioral principles*. Hence, we do not need to rely on additional arguments about hypothetical endgame effects and the speculations we raised in Section 5.3.2 are not justified. On the other hand, it is somewhat puzzling that the significant endgame effect of the single-shot games has disappeared. This might be another indication that the repeated games caused significant distortion in behavior in the single-shot games, but it is difficult to determine the mechanisms of how. There is also no indication of the fact that subjects increase their contribution rates in the repeated games in order to gain reputation from others or to establish a good image.

Unlike in the single-shot games, we did not find a significant decay of contribution between different parts of the game. It seems that if independent learning occurred at all, it occurred as a continuous process without structural breaks. Therefore the differences in contribution rates between experimental parts (see Section 5.3.3) did not result from an independent between parts trend. It is reasonable to believe that the significant between parts trend in the single-shot games is also a consequence of within part trends through the repeated games.

Table 5.6.2.2. Results of multilevel logistic regression on contribution propensities with personal characteristics and control variables of temporal embeddedness

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes	model assuming random slopes
FIXED EFFECTS			
α_0 baseline contribution propensity	?	1.138** (.410)	1.030* (.429)
STRUCTURAL EMBEDDEDNESS			
s_0 internalized selective incentives	+	.217*** (.053)	.217*** (.063)
s_1 direct selective incentives	+	.589*** (.085)	.544*** (.105)
b_0 internalized behavioral confirmation	+	.255*** (.039)	.223*** (.048)
b_1 direct behavioral confirmation	+	.602*** (.063)	.624*** (.089)
t_0 internalized traitor rewards	-	.104 (.056)	.135* (.059)
TEMPORAL EMBEDDEDNESS			
Criticalness			
p anticipated peace	-	-2.720*** (.098)	-2.592*** (.144)
d anticipated defeat	-	-2.074*** (.092)	-2.068*** (.112)
c anticipated clash	+	-.751*** (.080)	-.770*** (.091)
v anticipated victory (not *v* in *r-1*)	+	.251** (.092)	.267** (.101)
Intergroup reciprocity			
p peace in *r-1*	-	-.058 (.096)	-.050 (.099)
d defeat in *r-1*	+	.213* (.093)	.194* (.096)
c clash in *r-1*	+	-.018 (.099)	-.032 (.102)
Reinforcement learning			
predicted increase of contribution propensity	+	-.050 (.072)	-.086 (.077)
predicted decrease of contribution propensity	-	-.167** (.061)	-.142* (.063)
Local reciprocity			
t_1 towards opposite neighbors	+	-.016 (.030)	-.027 (.031)
b_2 towards fellow neighbors	+	.139** (.046)	.128** (.048)
b_3 towards fellow neighbors, if b_1 is introduced	+	-.004 (.048)	-.021 (.071)
PERSONAL CHARACTERISTICS			
gender (1=male)		-.018 (.130)	.024 (.136)
student at the university (1=yes)		.113 (.340)	.233 (.352)
studies at the law faculty		-.352 (.337)	-.460 (.352)
studies natural sciences		-.046 (.317)	-.242 (.329)
studies economic, business, or spatial sciences		-.269 (.309)	-.422 (.321)
studies social sciences		-.137 (.285)	-.217 (.296)
student of literary studies or arts		-.153 (.293)	-.358 (.304)
did a similar experiment before		-.115 (.125)	-.121 (.129)
strong risk aversion towards gains		-.064 (.123)	-.032 (.128)
strong loss aversion		-.022 (.122)	-.053 (.126)
consistency on social orientation questions		-.011 (.165)	.005 (.171)
prosocial orientation		.309 (.166)	.325 (.173)
egalitarian orientation		.263 (.160)	.269 (.166)
number of acquainted subjects in the experiment		-.019 (.081)	-.009 (.083)
quiz questions answered correctly %		-.002 (.003)	-.001 (.004)

cont. next page

CONTROL VARIABLES RELATED TO TEMPORAL EFFECTS		
delay (minutes) at the start of the experiment	.006 (.007)	.006 (.007)
length of play (number of rounds)	.005 (.008)	.007 (.009)
within part trend, if no new incentives	-.023* (.010)	-.024* (.011)
within Part III trend, *s*, medium clustering	-.006 (.015)	-.010 (.016)
within Part III trend, *s*, high clustering	-.018 (.019)	-.018 (.020)
within Part III trend, *b*, medium clustering	-.023 (.014)	-.034* (.015)
within Part III trend, *b*, high clustering	-.020 (.015)	-.021 (.016)
within Part IV trend, medium clustering	.004 (.015)	-.009 (.016)
within Part IV trend, high clustering	.025 (.017)	.023 (.018)
endgame effect in round 5	-.066 (.100)	-.048 (.100)
endgame effect in round 10	.020 (.094)	.019 (.094)
endgame effect in round 15	.092 (.106)	.092 (.105)
between part trend	-.208 (.149)	-.232 (.168)
reputation effect	.022 (.081)	.031 (.081)
long term criticalness	-.118* (.060)	-.074 (.061)
RANDOM EFFECTS		
interindividual variance	.591^{+++} (.072)	.538^{+++} (.075)
intersession variance	.005 (.024)	.000 (.000)
variance of s_0 internalized selective incentives		.075^{+++} (.036)
variance of s_1 direct selective incentives		.117 (.068)
variance of b_0 internalized behavioral confirmation		.070^{+++} (.030)
variance of b_1 direct behavioral confirmation		.398^{+++} (.100)
variance of t_0 internalized traitor rewards		.000 (.000)
variance of *p* anticipated peace		1.747 (.294)
variance of *d* anticipated defeat		.594^{+++} (.148)
variance of *c* anticipated clash		.239^{+++} (.090)
variance of *v* anticipated victory (not *v* in *r-1*)		.195^{+++} (.071)
variance of *p* peace in *r-1*		.000 (.000)
variance of *d* defeat in *r-1*		.003 (.055)
variance of *c* clash in *r-1*		.000 (.000)
variance of positive reinforcement		.102^{+} (.064)
variance of negative reinforcement		.029^{++} (.044)
variance of local reciprocity t_1 towards opposite neighbors		.000 (.000)
variance of local reciprocity b_2 towards fellow neighbors		.006^{++} (.019)
variance of local reciprocity b_3 towards fellow neighbors		.000 (.000)
-2 Log Likelihood model	10326.7	10226.4
Improvement χ^2 (df) for model in right column		100.3*** (17)
vs. previous model	45.3*** (15)	42.8*** (15)

Notes: N=14244 decisions for 203 subjects in 21 experimental sessions. Numbers in parentheses are standard errors. Iterative Generalized Least Squares estimates. * significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{+}$ significant at the 5% level, $^{++}$significant at the 1% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes).

Among the control variables, we also included the long-term effect of *criticalness*. In case the outcome of the game shows stability in a series of games (except when it is clash) subjects might recognize that their decision is not critical, therefore they will be less likely to contribute afterwards. The first model in Table 5.6.2.2 finds this effect significant, which shows *support for the long-term effect of criticalness*. This effect is in addition to the short term effect of the criticalness principle. However, in the model that allows a random variation in the slopes of the main explanatory variables, this effect is no longer significant. The most likely cause of the difference in the results of the two models is the significant variation in the effect of short-term criticalness between subjects. Some subjects are highly guided by this forward-looking principle, while others are influenced only to a limited extent. This variance rules out the net effect of long term criticalness.

In this section, we extended our multilevel analysis by control variables about personal characteristics and time effects. We did not find any indication that such personal characteristics as gender, study direction, risk preferences, and social orientations had significant effects on individual decisions. Among the controls for time effects, we found a significant independent downward trend in the absence of new monetary incentives and a long-term effect of criticalness. The inclusion of control variables caused only small changes in the parameter values of our main explanatory factors.

5.6.3 Interaction effects

In this section, we extend our analysis by incorporating certain cross-level interaction effects into the model explaining contribution propensities. These effects contain partly interactions of structural embeddedness and personal characteristics that were also included in the analysis of single-shot games. Besides, in the analysis of repeated IPG games we include interactions of temporal embeddedness, especially of reciprocity and personal characteristics.

Table 5.6.3.1 displays results of the extended analyses. As improvement statistics show, these variables enriched the explanatory power of the model. With regard to testing our main micro hypotheses, we cannot report much change as compared with the previous analyses that also included personal characteristics and control variables of temporal embeddedness among the predictors (cf. Table 5.6.2.2).

The one and only substantive change in the significance of main explanatory variables is of *local reciprocity towards fellow neighbors*. The significant effect of this variable disappeared after controlling for its interaction with personal characteristics and for its interaction with the gender of the neighbor. The significance of some of these effects demonstrates that *local reciprocity does not work universally* even towards fellow neighbors. Gender did not make a difference for the application of local reciprocal rules towards fellows, but social orientations did. Everything else is being as given *prosocials applied the local reciprocity rule towards fellows significantly more often than proselfs*. Subjects with prosocial orientation were strongly influenced by the previous decision of fellow neighbors, not like proselfs, who were more likely to disregard such information.

The application of local reciprocal strategies was not only conditional on personal characteristics of the subject, but also on the *target of reciprocity*. Unfortunately we do not have information about subjective perceptions of neighbors, but such perceptions were certainly major determinants of conditioning behavior on the action of neighbors. Subject might have used these perceptions to detect trustworthiness of neighbors and anticipate their behavior in the forthcoming rounds. These perceptions and sympathy might be based on some simple telltale signs (Frank, 1988; Macy and Skvoretz, 1998). What could be these telltale signs between strangers, however, is not clear. One sign that is easily distinguishable for everyone and can be partly a determinant of trust is gender. Since previous experiments found contradictory results about the likelihood of contribution depending on the gender of opponent (cf., Ortmann and Tichy, 1999; Solnick and Schweitzer, 1999), we also did not formulate hypotheses about subjects' perceptions about neighbors of different gender (see Section 3.4.2). On the other hand, we recognize that our analysis should be controlled for the target of local reciprocity at least in this respect. Results of the random slope model in Table 5.6.3.1 justified the inclusion of these interaction effects. *Decisions of fellow neighbors were reciprocated less likely, if fellow neighbors were women* (estimates are 0.331 and 0.332). Consequently, there are no gender differences with regard to those who choose to reciprocate or imitate, but there are gender differences of whom to reciprocate or imitate.

We also have interesting results for interactions with other forms of *local reciprocity*. We were interested whether the effect of fellow local reciprocity was strengthened after the introduction of monetary behavioral confirmation. We did not find confirmation of this effect in the previous models or here. However, interactions show that *local reciprocity was strengthened for some subjects and towards certain fellow neighbors*. As the model with fixed slopes shows, *men were less likely influenced by the introduction of monetary behavioral confirmation to reciprocate actions of fellow neighbors* (-0.266). This result contradicts to the related findings in the two-person PD of Rapoport and Chammah (1965) that men are more inclined to play TFT, especially when there are more economic incentives to do so. Our result is puzzling also in the light that neither the main effect of gender nor the main effect of strengthened local reciprocity was significant.

With regard to the interaction effects of social orientations and the application of local reciprocal strategies after the introduction of monetary confirmation rewards, results show that proselfs were more influenced by these additional incentives (estimates are -0.676 and -0.549). This supports an explanation that these new rewards are clearly framed as pure monetary incentives and not as guidelines of behavioral confirmation. When application of local reciprocity was not rewarded by monetary payoffs, prosocials reciprocated decisions of fellow neighbors more likely (estimates are 0.263 and 0.208). The presence of these two counterpolar interaction effects partly explains why there are no significant main effects of social orientations.

Besides, the strengthening of local reciprocity towards fellow neighbors was also conditional on the target of reciprocity. Similar to baseline fellow reciprocity, the gender of fellow neighbors made a significant difference. However, unlike that case, the application of local reciprocal strategies was strengthened in case fellow neighbors were women. It means that in the absence of additional monetary incentives subjects imitated actions of male neighbors more likely, but when there were monetary reasons for imitation, this discrimination has changed its direction.

There were no significant differences between men and women with regard to reciprocating or imitating decisions of neighbors from the opposite group. A personal characteristic that made a significant difference for the application of local reciprocal strategies towards opposite neighbors was social orientations. We found that prosocials follow local reciprocal rules more likely than proselfs (estimates are 0.182 and 0.156). This result indicates that prosocials reciprocated actions of their neighbors from both sides, unlike proselfs, who were not influenced by the previous decisions of their neighbors. All this means that prosocials are not likely to be initiators of neither conflict nor peace, but they are the ones, who most likely mobilized by their environment to join and support ongoing tendencies. As a consequence, a sufficient number of prosocial subjects can establish inflating scenarios of clashes or peace.

The target of reciprocity also matters for the application of local reciprocal strategies. But unlike in the case of fellow neighbors, there are no net differences, in relation to the number of female neighbors. For reciprocating or imitating neighbors from the opposite group, it is more important, whether the neighbor has the same sex or not. We found that reciprocation was less likely in case there were more neighbors from the other sex (estimates are -0.093 and -0.113). Subjects reciprocated actions of members of the opposite group when they were similar to them, at least with regard to their gender.

Although there are many interesting interaction effects of gender and local reciprocity, *we did not find significant interactions between gender and intergroup reciprocity*. Gender was also unimportant in relation to how quickly subjects learned the structure of the game, for which we used an independent within part trend as an indicator.

Besides interactions of the application of reciprocal strategies and subject-level characteristics as gender and social orientations, we included cross-level interactions of social control and personal characteristics. These effects were included also in the analysis of single-shot games. In the single-shot games, we found that prosocial individuals were more influenced by internalized traitor rewards than proselfs. Sacrifices of prosocials made them responsible for higher likelihood of conflicts between the groups. On the other hand, this interaction effect was not significant in the repeated IPG games. Other insignificant effects among the interactions were not significant in the single-shot games, either.

We found a significant interaction between the influence of internalized traitor rewards and the number of neighbors from the other group with another sex in the single-shot games. In Section 3.6.6, we provided the following arguments for this result. Only neighbors of the opposite gender provide a significant social control in the form of traitor rewards. For most subjects the color of the flag attached to the computers is not sufficient to activate these internalized incentives, but an additional difference in an apparent characteristic, such as gender, might help to make the substantive distinction to consider traitor pressure at the decisions. Results in Table 5.6.3.1 show that this interaction variable had a significant effect also in the repeated games (estimates are -0.183 and -0.157).

Table 5.6.3.1. Results of multilevel logistic regression on contribution propensities with personal characteristics, control variables of temporal embeddedness, and cross-level interactions

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of main effects	multilevel model assuming random slopes of main effects
FIXED EFFECTS			
α_0 baseline contribution propensity	?	1.217** (.413)	1.133** (.434)
STRUCTURAL EMBEDDEDNESS			
s_0 internalized selective incentives	+	.198*** (.055)	.187** (.064)
s_1 direct selective incentives	+	.559*** (.085)	.525*** (.106)
b_0 internalized behavioral confirmation	+	.294*** (.071)	.280** (.093)
b_1 direct behavioral confirmation	+	.604*** (.064)	.635*** (.085)
t_0 internalized traitor rewards	-	.172 (.088)	.205* (.095)
TEMPORAL EMBEDDEDNESS			
Criticalness			
p anticipated peace	-	-2.733*** (.099)	-2.599*** (.144)
d anticipated defeat	-	-2.101*** (.093)	-2.088*** (.113)
c anticipated clash	+	-.763*** (.081)	-.781*** (.091)
v anticipated victory (not v in $r\text{-}1$)	+	.252** (.093)	.263** (.102)
Intergroup reciprocity			
p peace in $r\text{-}1$	-	-.012 (.111)	-.018 (.115)
d defeat in $r\text{-}1$	+	.224* (.107)	.216 (.112)
c clash in $r\text{-}1$	+	-.103 (.117)	-.115 (.123)
Reinforcement learning			
predicted increase of contribution propensity	+	-.062 (.072)	-.094 (.078)
predicted decrease of contribution propensity	-	-.168** (.061)	-.137* (.064)
Local reciprocity			
t_1 towards opposite neighbors	+	-.105 (.062)	-.082 (.064)
b_2 towards fellow neighbors	+	.125 (.102)	.199 (.105)
b_3 towards fellow neighbors, if b is introduced	+	-.023 (.145)	-.151 (.153)

cont. next page

PERSONAL CHARACTERISTICS		
gender (1=male)	-.071 (.160)	-.047 (.163)
student at the university (1=yes)	.128 (.341)	.253 (.355)
studies at the law faculty	-.395 (.337)	-.486 (.355)
studies natural sciences	-.070 (.317)	-.264 (.332)
studies economic, business, or spatial sciences	-.300 (.309)	-.454 (.323)
studies social sciences	-.179 (.285)	-.251 (.298)
student of literary studies or arts	-.168 (.293)	-.370 (.306)
did a similar experiment before	-.131 (.125)	-.134 (.131)
strong risk aversion towards gains	-.066 (.124)	-.035 (.129)
strong loss aversion	-.017 (.122)	-.054 (.127)
consistency on social orientation questions	-.037 (.166)	-.013 (.173)
prosocial orientation	.265 (.178)	.292 (.183)
egalitarian orientation	.237 (.174)	.194 (.178)
number of acquainted subjects in the experiment	-.011 (.081)	.007 (.084)
quiz questions answered correctly %	-.001 (.003)	.000 (.004)
CONTROL VARIABLES RELATED TO TEMPORAL EFFECTS		
delay (minutes) at the start of the experiment	.005 (.007)	.004 (.007)
length of play (number of rounds)	.007 (.009)	.009 (.010)
within part trend, if no new incentives	-.030** (.011)	-.033** (.012)
within Part III trend, *s*, medium clustering	-.014 (.016)	-.020 (.017)
within Part III trend, *s*, high clustering	-.025 (.020)	-.024 (.021)
within Part III trend, *b*, medium clustering	-.033* (.015)	-.045** (.016)
within Part III trend, *b*, high clustering	-.029 (.016)	-.030 (.017)
within Part IV trend, medium clustering	-.005 (.015)	-.019 (.016)
within Part IV trend, high clustering	.017 (.017)	.013 (.018)
endgame effect in round 5	-.067 (.101)	-.048 (.101)
endgame effect in round 10	.010 (.095)	.008 (.095)
endgame effect in round 15	.096 (.106)	.095 (.106)
between part trend	-.248 (.151)	-.278 (.171)
reputation effect	.024 (.082)	.036 (.082)
long term criticalness	-.136* (.060)	-.091 (.062)
CROSS-LEVEL INTERACTIONS		
Interactions with gender		
gender * within part trend	.013 (.008)	.015 (.009)
gender * previous peace (intergroup reciprocity)	-.092 (.129)	-.070 (.141)
gender * previous defeat (intergroup reciprocity)	.003 (.125)	-.043 (.137)
gender * previous clash (intergroup reciprocity)	.214 (.145)	.204 (.156)
gender * local reciprocity towards opposite neighbors	.099 (.075)	.108 (.076)
gender * local reciprocity towards fellow neighbors	.116 (.096)	.038 (.099)
gender * local reciprocity towards fellows nbs, if *b* is introduced	-.266* (.134)	-.183 (.140)
Interactions with social control		
t_0* number of opposite neighbors of the other sex	-.183** (.057)	-.157* (.061)
t_0* number of male opposite neighbors	.105 (.061)	.049 (.065)
b_0* number of fellow neighbors of the same sex	.090 (.049)	.036 (.068)
b_0* number of female fellow neighbors	-.099* (.049)	-.069 (.070)
t_0* prosocial orientation	.122 (.088)	.098 (.093)
b_0* prosocial orientation	-.047 (.076)	-.027 (.106)
t_0* egalitarian orientation	.029 (.098)	.122 (.105)
b_0* egalitarian orientation	-.017 (.080)	-.051 (.114)

		cont. from previous page
Interactions with local reciprocity		
loc. rec. tow. opposite nbs * number of opp. nbs of the other sex	-.093 (.053)	-.113* (.054)
loc. rec. tow. opposite nbs * number of male opposite neighbors	.105 (.058)	.111 (.059)
loc. rec. tow. fellow nbs * number of fellow nbs of the same sex	.016 (.070)	-.014 (.072)
loc. rec. tow. fellow nbs * number of female fellow neighbors	-.111 (.071)	-.145* (.074)
loc. rec. tow. fel. nbs, with *b* * n of fellow nbs of the same sex	.037 (.104)	.088 (.110)
loc. rec. tow. fellow nbs, with *b* * number of female fellow nbs	.331** (.108)	.332** (.113)
local rec. towards opposite neighbors * prosocial orientation	.182** (.067)	.156* (.068)
local rec. towards fellow neighbors * prosocial orientation	.263* (.117)	.208 (.121)
local rec. towards fellow nbs, if *b* is introduced * prosocial or.	-.676*** (.158)	-.549** (.167)
local rec. towards opposite neighbors * egalitarian orientation	.005 (.074)	-.034 (.075)
local rec. towards fellow neighbors * egalitarian orientation	-.088 (.126)	-.023 (.131)
local rec. towards fellow nbs, if *b* is introduced * egalitarian or.	.312 (.172)	.261 (.181)
RANDOM EFFECTS		
interindividual variance	.596^{+++} (.069)	.547^{+++} (.076)
intersession variance	.000 (.000)	.000 (.000)
variance of s_0 internalized selective incentives		.069^{+++} (.035)
variance of s_1 direct selective incentives		.131 (.070)
variance of b_0 internalized behavioral confirmation		.067^{+++} (.029)
variance of b_1 direct behavioral confirmation		.309 (.088)
variance of t_0 internalized traitor rewards		.000 (.000)
variance of *p* anticipated peace		1.763 (.296)
variance of *d* anticipated defeat		.621^{+++} (.152)
variance of *c* anticipated clash		.241^{+++} (.091)
variance of *v* anticipated victory (not *v* in *r-1*)		.213^{+++} (.074)
variance of *p* peace in *r-1*		.000 (.000)
variance of *d* defeat in *r-1*		.000 (.000)
variance of *c* clash in *r-1*		.000 (.000)
variance of positive reinforcement		.107^{+} (.064)
variance of negative reinforcement		.029 (.045)
variance of local reciprocity t_1 towards opposite neighbors		.000 (.000)
variance of local reciprocity b_2 towards fellow neighbors		.001^{++} (.018)
variance of local reciprocity b_3 towards fellow neighbors		.000 (.000)
-2 Log Likelihood model	10085.2	10011.8
Improvement χ^2 (df) for model in right column		73.4*** (17)
vs. previous model	241.5*** (27)	203.6*** (27)

Notes: N=14244 decisions for 203 subjects in 21 experimental sessions. Numbers in parentheses are standard errors. Iterative Generalized Least Squares estimates. * significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects it is more appropriate to use deviance tests: $^{+}$ significant at the 5% level, $^{++}$significant at the 1% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes).

In this section, we extended our multilevel analysis by certain cross-level interaction effects, including interactions of structural embeddedness and personal characteristics and interactions of temporal embeddedness, especially of reciprocity and personal characteristics. These variables enriched the explanatory power of the model and also revealed some peculiarities about the predicted main behavioral mechanisms.

Particularly, results showed that local reciprocity works conditional on personal characteristics, especially on social orientations, and also conditional on the gender of the target of reciprocity.

5.7 How beliefs coincide with reality

In this section, we will have a closer look at how subjects form their expectations about the subsequent outcome in the repeated IPG game. In this way, we aim to find deeper roots of forward-looking behavior, of which presence was well supported by our analysis. On the basis of our results, we concluded that *criticalness* with a correction of forward-looking *bandwagon* tendencies is an important predictor of individual behavior in the experiments. This conclusion was drawn because dummies indicating expectations of subjects had significant effects on contributions.

A natural step towards a deeper investigation is to try to determine the basis of these expectations and the connections how do they depend on the past. Besides, it might help to find their origins, if we consider whether or not these expectations match with some sense of objective reality. If they do not, then we should come up with possible reasons for the systematic distortions.

Another major reason for a closer look at expectations is a concern about the validity of our explanation. We used expectations as predictors of decisions, assuming that subjects evaluate what is likely to happen in the subsequent round and they base their decisions on this evaluation. However, as subjects had to provide their forecasts parallel to their decisions, they could adjust their decisions in order to fulfill their predictions about the subsequent outcome. There was one subject, who explicitly reported that he adjusted his decisions in order to fulfill his expectations (see Section 5.5). A closer look at expectations might help to verify that this is not a serious concern for most subjects and most decisions.

To achieve these goals, in this section, we will report certain statistics about the expectations of subjects. To find their roots we report how these expectations were conditional on the past and on social orientations. We will check how often were expectations about the subsequent outcome correct and what could be the causes for systematic distortions. Finally, we will provide an indirect test for the concern about the relationship between expectations and decisions.

First, we try to ascertain determinants of the expectations of subjects. In Table 5.4.1.3 we reported average contribution rates conditional on the previous outcome, on previous decision, and on the expectation of the subject. Contribution rates were the focus of this table, but in parentheses the number of cell-relevant cases were also included. These numbers are important for the purpose of this section, which is to check for a possible effect of the previous round on the expectations of subjects. For the sake of clarity, we report these figures again in Table 5.7.1 in a different structure. Now the emphasis is on conditional probabilities of forming different expectations when the previous outcome and decision are given.

Table 5.7.1. Expectations of subjects in the *entire* experiment by previous decisions and previous outcomes

		prev. dec.	subject's expectation peace (*p*)	defeat (*d*)	victory (*v*)	clash (*c*)	total
previous outcome	*p*	C	32.84% (311)	7.39% (70)	49.74% (471)	10.03% (95)	100% (947)
		D	50.94% (1435)	8.38% (236)	32.23% (908)	8.45% (238)	100% (2817)
	d	C	10.52% (167)	27.22% (432)	46.00% (730)	16.26% (258)	100% (1587)
		D	12.86% (261)	23.80% (483)	45.76% (969)	15.57% (316)	100% (2029)
	v	C	7.29% (208)	7.18% (205)	70.05% (2000)	15.48% (442)	100% (2855)
		D	15.02% (114)	12.38% (94)	56.79% (431)	15.81% (120)	100% (759)
	c	C	9.09% (131)	13.88% (200)	54.34% (783)	22.69% (327)	100% (1441)
		D	11.74% (60)	12.13% (62)	59.69% (305)	16.44% (84)	100% (511)
total		C	11.96% (817)	13.30% (907)	58.33% (3984)	16.43% (1122)	100% (6830)
		D	30.56% (1870)	14.31% (875)	42.72% (2613)	12.39% (758)	100% (6116)
total			20.76% (2687)	13.76% (1782)	50.96% (6597)	14.52% (1880)	100% (12946)

Note: Numbers of cases are in parentheses.

As Table 5.7.1 shows, subjects were very optimistic about the outcome of the game. Victory was expected (50.96%) far more often than other outcomes and far more often than it has happened. In the absolute majority of decisions, victory was expected after a previous victory and clash. More astonishingly, in the relative majority of the cases, subjects expected the victory of their team even after defeat. Victory was expected quite often also after peace. However, in most cases, if they defected in the previous round, subjects expected a recurrence of peace. Recurrence of other outcomes was expected also relatively often (defeat was expected most likely after defeat, victory was expected most likely after victory, and clash was expected most likely after a clash). It is no surprise that clash is expected least likely after peace and defeat is expected least likely after peace and victory.

Subjects were very optimistic regardless they contributed or defected in the previous round. They were more likely to expect victory if they contributed before, but they were also quite optimistic after defection. There were no large differences with respect to the expectation of defeat and clash conditional on previous action. Subjects were more likely to expect peace to occur if they had defected in the previous round.

To sum up, Table 5.7.1 demonstrates that the previous outcome and decision had an effect on the forecasts of subjects, but that *the major determinant of expectations was an unfaltering optimism*. This optimism could be a result of hope in victory or wishful thinking that is excited partly by the competitive structure of the experimental situation and possibly also by general personality traits. This might also indicate an in-group bias, due to which subjects have a more positive view of their group fellows than of members of the out-group (cf. Turner, 1982; Bettencourt et al., 1992; Brewer, 1996b).

A personal characteristic that might influence expectations is social orientation. We found no significant main effects of social orientations on individual decision in the repeated games, although these effects were significant in the single-shot games. This difference could possibly be the result of the intervening effect of expectations that were not encountered in the single-shot games. If prosocial subjects are predominantly optimistic and strongly influenced by forward-looking *bandwagon* tendencies, then the net effect of social orientations might disappear after controlling for expectations. This way of reasoning is parallel to the arguments we provided for the main effects of social orientations in the single-shot games. We observed that prosocials made more sacrifices for their teams and caused more trouble in intergroup relations, probably because they evaluated social identity that can be attained from the intergroup competition more than did the proselfs. This could ground also their stronger belief in victory and higher optimism.

However, as Table 5.7.2 shows, there are no large differences between prosocials and proselfs, neither in their optimism about victory nor in their expectations about other outcomes. The hypothesis that expectation of victory is independent of prosocial values cannot be rejected ($\chi^2(1)=0.392$, two-tailed $p=0.531$). This has the consequence that the argument above about the intervening effect of expectations is not supported by the data.

Table 5.7.2 Expectations of subjects by social orientations

subject's expectation	peace (p)	defeat (d)	victory (v)	clash (c)	total
prosocial subjects	20.3%	12.6%	51.0%	16.1%	100%
	(1091)	(680)	(2745)	(867)	(5383)
others (proselfs or	20.9%	14.1%	50.8%	14.2%	100%
not consistent)	(1808)	(1223)	(4407)	(1233)	(8671)
total	20.6%	13.5%	50.9%	14.9%	100%
	(2899)	(1903)	(7152)	(2100)	(14054)

Note: Numbers of cases are in parentheses. All single-shot and repeated games are included.

Second, we discuss how accurate the expectations of subjects were and try to find factors that correlate with inaccurate forecasts. As we have seen, inaccuracy usually meant farfetched optimism about the outcome of the game. This optimism lead several subjects to make sacrifices for their groups, as they intended to join the bandwagon of victory that actually never came. Here we aim to explore reasons for these misbeliefs.

One possibility is that subjects could not foresee a shift in the outcome, because it was the result of changes in the decision of others. For instance, a previous victory suddenly disappeared because teammates became comfortable with the result and tried to free ride on the effort of others, meanwhile defeat mobilized contributions in the other team. On the other hand, as we have seen in Table 5.3.2.1, victory occurred most likely after a previous victory. Therefore, expectations of victory are most likely to be accurate after a previous victory.

Table 5.7.3. Results of multilevel logistic regression on *expectations of victory* [7]

independent variable	hypothesis about the direction of effect	model with decision-level variables only	multilevel model with subject-level variables
FIXED EFFECTS			
Decision-level variables			
(constant) baseline inaccuracy rate	?	.621*** (.067)	.749** (.233)
v victory in the previous round (*r-1*)	-	-.531*** (.055)	-.535*** (.055)
no feedback (*r*<7)	+	-.072 (.055)	-.074 (.055)
experience (*r*)	-	.003** (.001)	.003** (.001)
Personal characteristics and other subject-level variables			
gender (1=male)			-.023 (.092)
studies at the law faculty			-.007 (.195)
studies natural sciences			-.105 (.171)
studies economic, business, or spatial science			-.134 (.163)
studies social sciences			.096 (.142)
student of literary studies or arts			-.004 (.147)
strong risk aversion towards gains			-.076 (.088)
strong loss aversion			-.073 (.087)
prosocial orientation			.052 (.104)
egalitarian orientation			-.003 (.111)
quiz questions answered correctly %			-.001 (.002)
RANDOM EFFECT			
interindividual variance		$.231^{+++}$ (.036)	$.224^{+++}$ (.035)
-2 Log Likelihood model		9677.83	9667.81
Improvement χ^2 (df in parentheses) for model in right column			10.02* (11)
vs. baseline model		122.06*** (3)	132.08*** (14)

Notes: N=7199 expectations of victory for 203 subjects in 21 experimental sessions. Numbers in parentheses are standard errors. Iterative Generalized Least Squares estimates. * significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).

[7] The baseline model provides a parameter estimate of 0.514*** (0.047) for the baseline inaccuracy rate and for 0.312^{+++} (0.044) for interindividual variance. The -2 Log Likelihood of the baseline model is 9799.89.

Another possible reason for farfetched optimism is inexperience. Subjects did not know what to expect about the outcome in the single-shot games and in the first repeated game, because no feedback has been provided yet about previous results. We predict that subjects have become experienced gradually, hence the proportion of inaccurate forecasts should decrease over time.

Furthermore, certain personal characteristics could also play a role, including gender, study direction, risk preferences, social orientations, and the level of understanding of the experimental game. We do not have strong arguments to pledge ourselves for certain hypotheses about these variables. Another possible predictor is the understanding of the experimental situation that we measured by the proportion of quiz questions answered correctly.

Table 5.7.3 summarizes our predictions and reports the results of a multilevel logistic regression model that takes into account these variables. All cases (N=7199) are included in the analysis, in which a subject expected victory, including single-shot and repeated games. 61.37% (N=4418) of these expectations were inaccurate, that is the actual outcome was not victory. The dichotomous variable of accuracy (inaccurate=1) is the dependent variable of the model. We constructed a two-level design, in which single expectations are at the first level and subjects are at the second level. This was necessary, since personal characteristics affect all expectations of the subject in the same way and consequently the expectations cannot be handled as independent observations. On the other hand, there was no reason to include a third level of analysis.

As results in Table 5.7.3 show, the personal characteristics we included as predictors did not play an important role in which subjects had farfetched optimism about the outcome of the game. None of these variables had significant effects and their inclusion did not improve the explanatory power of the model. As expected, expectations of victory were more accurate in case the previous round also ended with victory. On the other hand, predictions about the lack of information and inexperience are not supported by the data. Subjects did not guess better when they knew the previous outcome. Moreover, their forecasts became less accurate over time.

After all, this analysis did not reveal the real reasons for farfetched optimism. We could conclude that the lack of information, inexperience, and the above mentioned personal characteristics are not responsible for misbeliefs.

Third, we will briefly deal with the concern about the relationship between expectations and decisions. Subjects had to provide their forecasts at the same time when they decided to keep their bonus or to contribute it to their team. Therefore, they could in principle adjust their decisions to fulfill their expectations. The idea that subjects try to maximize the probability that their expectations are right has correspondence with the reduction of cognitive dissonance (Festinger, 1962). In our experiments, fulfilling expectations would be especially strange, if subjects expected a clash or a defeat of their team. Throughout this chapter, when we discussed the effect of *criticalness* and *forward-looking bandwagon mechanisms* on individual decisions, we assumed that this concern could be neglected. Here we provide an indirect test of this assumption.

Table 5.7.4 Differences between fulfilling expectations and the adjusted criticalness hypothesis

subject's expectation	previous outcome and decision	*fulfilling expectations*	*criticalness and bandwagon*
peace (*p*)	any	-	-
defeat (*d*)	any	-	-
victory (*v*)	any	+	+
clash (*c*)	peace (*p*) or defeat (*d*)	+	+
	victory (*v*)	-	+
	clash and contribution (*c*C)	+	+
	clash and defection (*c*D)	-	+

In case subjects try to match their decisions with their expectations, it has the following consequences. In case they expected peace or defeat, independent of the previous outcome and of their previous decision, they should have kept their bonuses. In the situation where they anticipated victory they should have contributed in any case. In case their forecast was a clash there was no clear directive what they should do. When their team had lost in the previous round or when there was peace contribution would have been the appropriate reaction. Assuming that others will not often change their decisions after victory, defection could be a more efficient way to fulfill the expectation of a clash. After a clash, subjects have a solid rationale to believe in the status quo when they do not change their decisions. These implications are summarized in Table 5.7.4. The table also includes predictions that are based on the criticalness principle with adjustments of a forward-looking bandwagon effect. This helps to clarify what are the differences between the two explanations and at which points could the motive of fulfilling expectations be a serious concern for drawing conclusions about forward-looking behavior.

As Table 5.7.4 shows, there is almost a perfect overlap between the alternative hypothesis of *fulfilling expectations* and the predictions derived from the *criticalness hypothesis adjusted for forward-looking bandwagon tendencies*. Therefore, on the basis of conditional contribution rates it is difficult to carry out a test that decides between these two hypotheses. There are only differences in case the subject expects a clash to occur in the subsequent round. To demonstrate which predictions are closer to reality in this case, in Table 5.7.5 we recall conditional contribution rates, if the subject expected a clash in the coming round. These statistics have been already reported in Table 5.4.1.3. Here we also include the rival predictions of fulfilling expectations and the adjusted criticalness principle.

Table 5.7.5 The effect of fulfilling expectations and adjusted criticalness on contribution rates (%), if the subject expected a clash in the subsequent round

		fulfilling expectations		*criticalness and bandwagon*		contribution rates	
previous decision		C	D	C	D	C	D
previous outcome	*peace (p)*	+	+	+	+	56.84% (95)	43.28% (238)
	defeat (d)	+	+	+	+	78.68% (258)	59.49% (316)
	victory (v)	-	-	+	+	69.46% (442)	45.83% (120)
	clash (c)	+	-	+	+	78.29% (327)	41.67% (84)
total						73.08% (1122)	50.26% (758)

Notes: N=1880. Numbers of cases are in parentheses. The average contribution rate, if clash expected is 63.88%.

There are three cells in Table 5.7.5, for which the predictions of fulfilling expectations and adjusted criticalness are different. In the *v*C cell, the contribution rate (69.46%) is lower than after contribution in general, but the difference is not significant (t=1.652, two-tailed p=0.099). On the other hand, this contribution rate is significantly higher than the average contribution rate when clash is anticipated (t=2.543, two-tailed p=0.011). This supports the *adjusted criticalness* hypothesis rather than the tendency of fulfilling expectations. In the *v*D cell, the contribution rate (45.83%) is significantly lower than the average contribution rate when clash is expected (t=3.951, p<0.001), but the difference in comparison with the contribution rate after defection (50.26%) is not significant (t=0.969, two-tailed p=0.334). In the *c*D cell, the contribution rate is even lower (41.67%). The difference is significant compared to the overall mean in case clash is expected (t=4.105, p<0.001), but it is not significant in comparison with the average contribution rate after defection (t=1.588, two-tailed p=0.116). On the basis of these data, the alternative hypothesis of *fulfilling expectations* seems more plausible. Consequently, although it is unlikely, *we cannot exclude the possibility that subjects tried to adjust their decisions in order to fulfill their predictions*.

In this section we took a closer look at the forward-looking behavior of subjects in the experiment. *First*, we analyzed what could be the determinants of subjects' view about the future. We found that the previous outcome and decision had an effect on the forecasts of subjects, but the major determinant of expectations was an unfaltering optimism. Besides, we showed that expectations were independent of social orientations. *Second*, we looked at some possible origins of farfetched optimism we experienced. We found a significant effect of the previous outcome, but we had to conclude that the lack of information, inexperience, and personal characteristics are not significant determinants of farfetched optimism. *Third*, we discussed the possibility that subjects adjusted their decisions in order to fulfill their expectations. On the basis of an indirect test, we could not reject this alternative hypothesis.

5.8 Summary of results

In this chapter, we presented the results of the analysis of repeated IPG game experiments. First we explored the aggregated outcomes of the games and tried to trace the predicted scenarios in the experiments. Then we turned to the analysis of individual behavior and tested our main micro hypotheses. We also drew attention to the effects of intervening and control variables and to interactions. In this concluding section, we provide a summary of our main results in the order as we structured our main research questions in Table 4.2.1. The discussion of societal implications and limitations of this analysis are left for Chapter 6.

We classified our interests as *micro* and *macro* questions, depending on whether they related to individual behavior or to aggregated outcomes in the experiment. We emphasized that the emergence of intergroup conflict and peace cannot be explained as a direct consequence of group-level factors, such as segregation. In the spirit of methodological individualism, the explanation has to reveal the causes that determine individual actions leading to harmful consequences in intergroup relations. Therefore, we preserved the bottom-up character of the explanation. After testing the presence of macro phenomena in the experiments we concentrated on the roots of individual behavior.

We classified the main causes into three categories: effects of *structural embeddedness*, effects of *temporal embeddedness*, and their *interactions*. For the sake of testing our hypotheses about structural embeddedness we manipulated structural configurations in the experiment. We developed an experimental design, in which seating patterns of subjects were varied between experimental sessions. Structural manipulations were used for both the single-shot and the repeated games. Chapter 3 discussed the analysis of the single-shot games, in which effects of temporal embeddedness could be excluded. Effects of temporal embeddedness and the interactions of structure and time could only be analyzed in the repeated games. By a within session manipulation we made sure that we have also data about situations, in which effects of structural embeddedness can be excluded. In the control condition of every session (Part I), subjects were separated, hence they could not have been influenced by their neighbors. Further parts of the experiment ensured that we could analyze also the simultaneous effect and interactions of structural and temporal embeddedness.

Now let us summarize the answers we could give to the research questions we addressed in this chapter. Table 5.8.1 displays the main hypotheses we formulated in Chapter 4 for these questions and the conclusions we drew from the results in Chapter 5. Table 5.8.1 follows the structure of Table 4.2.1, in which we summarized our main research questions and the structure of Table 4.3.3.1, in which we formulated our hypotheses.

Table 5.8.1 Summary of main hypotheses (*H*) and results (*R*)

	Structural embeddedness	*Temporal embeddedness*	*Interactions of structure and time*
Macro hypotheses *dependent variable:* outcome(s) of the game	*H*: Conflict is more likely in segregated structures. *R*: Partly confirmed. Conflict was least likely in the low clustering condition. Both medium and high clustering has high likelihood of conflict.	*H*: Stable peace. *R*: Confirmed. *H*: Durable conflict. *R*: Largely supported. *H*: Spiral of peace. *R*: Largely supported. *H*: Spiral of conflict. *R*: Not supported.	*H*: Stable peace and the spiral of peace is more likely, if segregation is low and durable conflict and the spiral of conflict is more likely in segregated structures. *R*: Partly supported.
Micro hypotheses *dependent variable:* individual decision(s)	*H*: Subjects are influenced by internalized and direct forms of social control. *R*: Largely confirmed. Subjects are influenced by internalized and direct forms of social control from fellow neighbors and conditionally by traitor rewards.	*H*: Subjects follow (a combination of) simple behavioral rules. *R*: Partly confirmed. Criticalness adjusted for forward-looking bandwagon effects plays an important role. Reinforcement learning works only towards defection. There is little support for intergroup reciprocity.	*H*: Subjects reciprocate previous behavior of their neighbors. *R*: Conditionally supported. *H*: There is no direct effect of structure on which behavioral rules are followed. *R*: Partly supported.

With regard to effects of *structural embeddedness*, we were interested in the effects of manipulating structural configurations in the experiment on the outcome of the repeated IPG game and on individual decisions. Precisely, we were looking for an answer to whether segregation has an effect on the likelihood of intergroup conflict. We constructed three conditions (low, medium, and high clustering) in order to test the hypothesis that conflict is more likely in segregated structures. As we compared frequencies of intergroup conflict between structural conditions, we gained a partial confirmation of this hypothesis. Conflict was least frequent in the low clustering condition, but it did not occur more likely in the high clustering condition than in medium clustering (cf. Table 5.3.1.1). Further analysis has revealed that this discrepancy was partly a consequence of a ceiling effect and partly of a selection distortion. Subjects, who by chance were assigned to the medium clustering condition, had high baseline contribution rates. The initial difference is accumulated as subjects reciprocated contribution decisions (cf. Table 5.3.1.2). Furthermore, we also tested whether segregation increases the likelihood of conflict more when monetary selective incentives are introduced in comparison to when monetary confirmation rewards are present. Results confirm the hypothesis that the segregation effect is stronger under normative pressure than under confirmation pressure.

The mechanism behind the main effect of segregation we believe is the social control of neighbors. We predicted that different forms of social control influence decisions in a direct and in an internalized form. Social control in an internalized form gets activated in the presence of interdependent others, even when they are complete strangers, they have no channels of communication, and they have no ways to enforce each others' decisions in any way. We also tested the effects of direct forms of social control that we introduced as additional monetary rewards in Parts III and IV in the experiment. This allowed us to provide a comparison between the size of effects of monetary rewards and internalized forms of social control. We assumed that group membership of neighbors would activate different forms of social control, causing the segregation effect at the aggregated level.

Between fellow neighbors, we predicted the influence of internalized selective incentives and behavioral confirmation. Internalized selective incentives exert pressure towards contribution if there are fellow neighbors around. Behavioral confirmation is realized in an internalized form if subjects adjusted their decisions in order to match the expected decisions of their fellow neighbors. Selective incentives and behavioral confirmation were introduced also as monetary rewards that were distributed conditional on neighbor decisions in Parts III and IV of the experiment. Our analysis confirmed all hypotheses about the influence of these incentives. Internalized forms of fellow social control did not have as strong effects as monetary side-payments, but they were still highly significant predictors of individual decisions.

Between neighbors from opposite groups, we predicted the existence of an internalized pressure that drives towards defection. The analysis did not confirm the hypothesis about the universal presence of these incentives that we called traitor rewards. Subjects felt pressure to defect and betray the interests of their group only if they were surrounded by members of the other group who belonged to the opposite sex. It seems that gender worked as a telltale sign that helped to make the substantive distinction to consider traitor pressure at the decisions. Neighbors of the same sex did not activate this pressure. From another perspective, subjects were more competitive towards the out-group when their neighbors from the opposite group had the same gender.

With regard to effects of *temporal embeddedness*, we were interested in how the outcomes of intergroup competition change over time and we tried to trace typical scenarios in the experiment. Particularly, we focused on the dynamics of intergroup conflict and contribution rates. We predicted that once peace is established, when no changes are introduced, it would be stable over time. We got a solid confirmation of this hypothesis, as peace followed peace relatively often not only once, but in a sequence of peaceful outcomes. Similarly, we predicted that once it occurs, conflict would be durable. The analysis of results revealed that conflict was indeed repeated quite often, but this did not mean that one particular outcome was stabilized (see Section 5.3.2).

As a further consequence of this, the *spiral of conflict* hypothesis that we formulated about the process of self-reinforcing dynamics of harmful clashes was not supported by the data. Contribution rates very often stabilized at a certain high level, especially in Part IV in the high clustering condition, but there were always some defectors, who prevented the emergence of all-in clashes. On the other hand, we found some evidence for the *spiral of peace* hypothesis. Contribution rates decreased gradually in experimental sessions if no new incentives were introduced. However, there was a difference between the sessions, in which stages of the experiment this decrease has occurred. Furthermore, this spiral has also not reached its extremity, as some hardliners kept trying to contribute (see Section 5.3.3).

The dynamics of intergroup conflict and of contribution rates are the results of changes in individual behavior over time. The primary focus of this chapter was to test predictions about these changes. We aimed therefore to determine the strategies individuals follow in the repeated IPG game. We believed that these strategies are rather simple and softwired, as they are combinations of different heuristics and are not based on elaborate calculations. Furthermore, we debated both the rational forward-looking and the purely backward-looking views on human action. We discussed that individual decisions make use of both forward-looking calculations and past experience. Hence, we formulated hypotheses about the influence of simple mechanisms on individual choices that are partly oriented towards the future and partly towards the past.

With respect to the shadow of the future, we predicted that subjects count on the principle of *criticalness*. Criticalness prescribes contribution in case subjects perceive their decision in the forthcoming round as a decisive one that makes a difference for the aggregated outcome. We asked subjects directly at each round about their expectations about the outcome. The influence of these expectations on actual decisions was the basis of our test of the criticalness hypothesis. Results confirmed that expectations are strong predictors of individual choices in the repeated IPG game. As predicted, the expectation of peace and defeat decreased the willingness to contribute. However, the anticipation of clash was also associated with a decrease in contribution propensities, which contradicts the criticalness hypothesis. This negative result was related to the unconditional high contribution rates when victory was expected (see Table 5.4.1.1). It seems that subjects liked to make active contributions to the victory of their team, although they could have free ridden on the effort of others. Hence we concluded that the criticalness hypothesis should be adjusted for a forward-looking bandwagon tendency.

With respect to the shadow of the past, we formulated hypotheses about the influence of two behavioral principles: reinforcement learning and intergroup reciprocity. The *reinforcement learning* hypothesis predicted that subjects would be more likely to stick to their actions if they gained money in the previous round and that they would be more likely to change their decisions if they lost. When formulating this hypothesis, we did not make any additional assumptions on the speed and accuracy of learning and we assumed that subjects consider the zero payment as their aspiration level throughout the entire experiment. Results provided only conditional support for the reinforcement learning hypothesis. We found that reinforcement works if it drives towards defection and that it is more likely to be activated after a previous defection (for a similar result, see Flache, 1996).

The *intergroup reciprocity* hypothesis predicted that subjects reciprocate the observed collective behavior of the other group. They decrease their contribution propensity if the other group behaved peacefully in the previous round, and they are more likely to contribute if they observed a competitive behavior. On the basis of our analysis, this hypothesis did not receive confirmation. Although means of conditional contribution rates provided some evidence for intergroup reciprocity (see Section 5.4), the multilevel regression analysis proved that this effect was caused by other variables and only a previous defeat elicited significant retaliations.

A comparison of the relative size of effects provide more support for forward-looking considerations than for backward-looking strategies, which is due to the strong influence of subjective expectations. On the other hand, a partial confirmation of the hypotheses about backward-looking mechanisms justifies our efforts to integrate different behavioral mechanisms in a unified model of human action.

With regard to *interaction effects of structure and time*, at the macro level we were interested whether we can trace different scenarios in different structural conditions of the experiment. We predicted that stable peace and the *spiral of peace* would be most likely in the low clustering condition. Results supported this hypothesis, although the *spiral of peace* occurred in different stages of the experiment (see Section 5.3.3). On the other hand, we predicted that durable conflict and the *spiral of conflict* would be more likely in segregated structures. In accordance with the previous finding, these scenarios did not occur at all in the low clustering condition. There was also support for the stabilization of conflict in the high clustering condition. However, contribution rates did not increase gradually within experimental parts, but between the parts. Consequently, we did not find any support for the *spiral of conflict* hypothesis, not even in the most segregated structures. Besides, there was no indication of any recognizable scenarios in the medium clustering condition (see Section 5.3.3).

We were interested in finding the micro causes of different scenarios. We aimed to answer whether or not subjects apply different strategies in different structural conditions and if they are influenced by other interaction effects of structure and time, particularly by previous decisions of their neighbors. We showed that for the explanation of interactions at the aggregated level a direct effect of the structural condition on which behavioral rules are applied does not necessarily have to exist.

Instead, differences in the scenarios can be the consequence of main effects of structural and temporal embeddedness on individual decisions. Subjects cannot apply certain conditional strategies if they do not have the opportunity to do so. For instance, they cannot retaliate conflict if there was no conflict. We predicted that further reasons behind the macro scenarios are the influence of previous decisions of neighbors and independent learning of the structure of the game.

We formulated a hypothesis about *local reciprocity* that prescribes contribution as a reaction to contribution of neighbors and drives towards defection, if neighbors kept their bonus in the previous round. Results provided conditional support for this hypothesis depending on personal characteristics of the subject and on the target of reciprocity. We found a significant main effect of local fellow reciprocity only (see Sections 5.6.1 and 5.6.2). Looking deeper at the interaction effects, results showed a complex picture about the conditions under which local reciprocity influences individual decisions. We found significant interaction effects between social orientations and all forms of local reciprocity. Prosocials reciprocated actions of both fellow and opposite neighbors more likely than proselfs. On the other hand, the introduction of monetary behavioral confirmation elicited less additional reciprocity from prosocials than from proselfs. These monetary incentives also had an interaction with gender: men were less likely influenced by them to reciprocate their fellow neighbors than women. Furthermore, responses to decisions of neighbors were not only conditional on the source, but also on the target of local reciprocity. Subjects reciprocated (or imitated) their fellow neighbors more likely if they were males. On the other hand, reciprocating actions of neighbors from the other group was conditional on identicalness of gender. Local reciprocity was applied less likely if there were more neighbors from the other sex. To sum up, local reciprocity did not work universally, but only under certain circumstances.

The hypothesis that subjects do not use different strategies under different structural conditions is partly supported by the data. Criticalness and forward-looking bandwagon tendencies affected decisions under every structural condition and in each experimental part. On the other hand, reinforcement learning had an effect only in later stages of the experiment, but irrespective of the structural condition. It seems that subjects have “learned the principle of learning” during the game (see Section 5.4.3).

We also discussed that the interaction effect of structure and time at the aggregated level can be partly the consequence of independent learning trends. Subjects might learn the structure of the game during the experiment and they might adjust their decisions towards a rational choice over time, irrespective of previous outcomes and future expectations. For testing this preposition, we included trend variables as controls in the analysis (see Sections 5.6.2 and 5.6.3). We found a significant downward trend that is independent from intergroup reciprocity and the other main behavioral mechanisms if no monetary incentives were introduced in the experiment. This is also partly the reason for the occurrence of the *spiral of peace* at the aggregated level.

"… both knowledge about and acquaintance with out-groups lessen hostility toward them."

Gordon W. Allport: The Nature of Prejudice (1954: 265)

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 Summary of research questions and results

6.1.1 Research problem

Harmful conflict between groups is among the most costly enigma of mankind. The high complexity of intergroup conflict in our everyday life, however, makes the solution of this enigma a challenging task. Instead of dealing with all the complexities, by making simplifying assumptions and by relying on a sound theoretical basis, this study contributed to the investigation of certain factors that influence the likelihood of intergroup conflict. First of all, we emphasized that *intergroup conflicts arise as groups compete for the possession of scarce material or immaterial resources*. In pursuit of the underlying processes that drive competitions toward lethal consequences, we focused on effects of *structural and temporal embeddedness*. Particularly, we intended to determine some conditions under which *segregation* is associated with conflict and the causal mechanisms that are responsible for this effect. As most intergroup relations cannot be torn from their historical context, we also investigated how conflict changes over time and what are the endogenous determinants of positive and negative changes.

Such an investigation should not remain at the intergroup level, as groups are not unitary entities. They consist of consciously acting individuals, who make their decisions based on their own free will. For instance, citizens volunteer for military service in a war situation, soldiers volunteer to be sent to the front lines, front liners volunteer to be asked for special assignments (Coleman, 1990). Terrorists who commit suicide bombings not only risk their life, but they sacrifice it in the belief that it helps their groups to obtain a certain goal, such as independence. Individual participation in intergroup conflict is rarely so demanding, but it always involves a cost that members have to bear to help their group. Contributions of this sort are seemingly rather more senseless than valuable. The fundamental question this study has raised is *why are people willing to make such sacrifices, even though they are not required to do so and their behavior could lead to lethal consequences*. Forces and motives that facilitate individual participation are necessary to discover for the explanation of intergroup conflict.

In this study, we focused on factors that can be classified according to their origin in *intergroup competition*, in the *social network of group members*, or in the *time horizon of intergroup and interpersonal relations*. Intergroup competition and comparison motivates participation by *public good* rewards, such as territory, pride, or social identity that are distributed among the members of the winning side. Social relationships with relevant others constrain individual decisions by transmitting different forms of *social control* that can either enforce or reduce participation. The historical context is important as people behave according to certain *decision heuristics*. These heuristics are partly built on their *experience*, for instance, they learn from the past or revenge previous outrages. Similarly, *expectations* about future encounters also influence individual actions strongly.

6.1.2 Structural embeddedness and intergroup conflict

The major theoretical goal of this study was to *construct a model of intergroup competition* that can help to understand the emergence of harmful conflicts that are the consequence of structural and temporal embeddedness. As a basis of model building, we departed from the *Intergroup Public Goods (IPG) game* model of competitive intergroup relations (Rapoport and Bornstein, 1987). The major potential of the IPG game model is that it coherently represents the dual interdependence structure of intergroup competitions; the negative interdependency between the groups on one hand and the free riding problem within the groups on the other hand. However, the IPG model in its original form could not cope with *effects of structural embeddedness*. Structural embeddedness affects individual decisions through different *social control mechanisms*. Social *selective incentives*, such as respect, are distributed as rewards for fellows who made sacrifices for the group and as punishments for those fellows who failed to take part in group efforts. *Behavioral confirmation* is received for an action that is identical to the behavior of fellow group members. *Traitor rewards* are sent to relevant members of the opposite group who refused to help their own group. As a substantial contribution of this study, we incorporated these forms of social control mechanisms into the IPG model of competitive intergroup relations.

As social control is transmitted through network relations, the *network structure of individual ties has an impact on the outcome of intergroup competition*. One of our main research questions targeted to determine the *structural conditions under which the likelihood of intergroup conflict is higher* and the conditions under which peaceful coexistence might be expected.

In the structurally embedded IPG game, under certain structural conditions strong social control might lead to an unusual social dilemma situation. In this social dilemma, everyone is better off by *contributing* for the group, but overall contribution would result in an outcome that for everyone is worse than peaceful coexistence. In this interdependence structure, unintended conflict is the aggregated consequence of intentional individual actions. The model predicts that such a social dilemma would occur where dense contacts within the groups are distribution channels for social selective incentives, rather than spreading routes of behavioral confirmation. An additional requirement for the emergence of suboptimal conflict situations is that interpersonal ties between members of the competing groups should be scarce or traitor rewards should be of little importance.

Our interest was not purely in determining the conditions under which overall conflict evolves as a social dilemma, but in general in showing the structural conditions that increase the likelihood of intergroup conflict. In particular, a main research question concerned why and under what conditions *segregation* promotes intergroup conflict. For the investigation of this question we used simulations and experiments.

Simulations were applied in order to explore the general theoretical relationship between segregation and intergroup conflict and to derive hypotheses about the effect of structural embeddedness for the experimental investigation. In the simulations, the structurally embedded IPG game was played by agents arranged in simplified network structures. As a general result, we found that the effect segregation on the likelihood of intergroup conflict can be typically characterized by an S-shape function. This means that *segregation is likely to promote intergroup conflict*. However, depending on other parameter values, *in certain ranges of segregation, an additional change does not result in an increase in the likelihood of conflict*. These cases were labeled as *floor* and *ceiling* effects. Furthermore, we found *a stronger segregation effect where local selective incentives were relatively important when compared to behavioral confirmation from neighbors*.

Simulation results were reformulated as hypotheses for the *experiments*. Besides the theoretical developments, testing these hypotheses in experiments was another major scientific contribution of this study. For the implementation of structural conditions in laboratory, a *new experimental design* needed to be invented. In the experiments, subjects played the structurally embedded IPG game without communication in two teams consisting of five members. In order to test the presence of a segregation effect and the underlying social control mechanisms, seating patterns were varied and visibility conditions were manipulated. Four structural conditions were applied, one in which subjects were separated, one in which the two groups were arranged in a highly segregated pattern, one with medium, and one with low levels of segregation. The segregation effect on the likelihood of intergroup conflict was tested by comparing outcomes of the game under different seating patterns.

We predicted that social control would affect individual decisions also in an *internalized* form. For detecting internalized forms of social control, in one part of the experiment subjects established eye contact with their direct neighbors. In later parts of the experiment, also *direct forms of social control* were introduced as monetary rewards in order to test the relative impact of internalized and direct forms of social control on individual decisions.

Experimental results *partly confirmed the segregation hypothesis*. As predicted, intergroup conflict was least likely when members of the two teams were seated in a mixed pattern. However, conflict was not less likely in the medium segregation condition than in high segregation. This could partly be explained as a ceiling effect and partly as a consequence of high baseline contribution rates in the medium segregation setting.

In general, monetary social control strongly influenced individual decisions. We found strong evidence of internalized behavioral confirmation as subjects adjusted their decision towards the expected decision of their fellow neighbors even though there was only eye contact between them. There was also some support for the effect of internalized selective incentives. These effects were not as strong as the influence of monetary social control, but internalized social control from fellow neighbors was still a major predictor of individual contribution propensities. Internalized traitor incentives were found only under certain conditions. Subjects felt a pressure to betray the interests of their group only when they were surrounded by members of the other group, who belonged to the opposite sex.

To summarize, the new experimental design provided important insights for understanding structural effects and the influence of social control in intergroup situations. Particularly, results demonstrated how important *eye contact* between the subjects in the laboratory is in activating internalized mechanisms of social control.

What follows is a brief summary of our main *research questions* (**Q**), our derived hypotheses (**H**), and **experimental results** (**R**) about the effect of segregation on intergroup conflict.

Q *Why and under what conditions does segregation promote intergroup conflicts? In an experimental setting, is the likelihood of intergroup conflict higher when group members are arranged in a segregated pattern? Is there such an effect when individuals only have eye contact with each other?*

H The network structure affects the likelihood of conflict because it is the channel of social control mechanisms. In a segregated structure, the likelihood of intergroup conflict is usually higher because dense connections between group fellows allow for the spread of selective incentives that support mobilization. Moreover, scarce relations between members of the competing groups mitigate the distribution of suppressing motives. Because social control is often internalized, segregation increases the likelihood of intergroup conflict also where individuals only have eye contact with each other.

R **By comparing results from different structural configurations of the experiment, we found that intergroup conflict was least likely when members of the two teams were seated in a mixed pattern. However, conflict was not less likely in the medium segregation condition than in high segregation. This was also the case, when only minimal contact was introduced between neighboring subjects.**

Answering these questions required an explanation of individual behavior that is causing these macro consequences.

Q *Why and under which conditions do individuals participate in collective action that hurts the interest of another group and might result in mutually harmful consequences? What are the underlying mechanisms of network effects at the interpersonal level?*

How do different forms of social control, namely selective incentives, behavioral confirmation, and traitor rewards contribute to intergroup conflict and what is the impact of their relative size? Do these forms of social control affect individual decisions in a direct and in an internalized form in an experimental environment?

H Individuals might be mobilized to participate if direct and internalized forms of social control compensate for the cost of contribution. Social control implies a distribution of positive and negative incentives that constrain individual decisions, conditional on expectations and behavior of the individual and relevant others. Depending on the nature of social control and on the composition of the ego-network of the individual, social control can facilitate or suppress participation. As the web of interpersonal relations defines the social network, macro structural effects can be derived as a consequence of social control mechanisms.

Social selective incentives reward group fellows for participation. Traitor rewards are conveyed in interpersonal relations between members of the competing groups and inhibit participation. Consequently, these forms of social control are responsible for the segregation effect. Behavioral confirmation, however, has a double edge; it can contribute to the establishment of widespread activism as well as to the dissemination of peaceful behavior. The larger the size of normative pressure (selective incentives) compared to confirmation pressure, the stronger the effect of segregation on intergroup conflict.

R Subjects in the experiment were strongly influenced by direct forms of social control. We found strong evidence of internalized behavioral confirmation and some evidence of internalized selective incentives. Internalized traitor incentives were found only under certain conditions. These social control effects were sufficient to cause a difference in the likelihood of conflict between structural conditions. The segregation effect was stronger when normative pressure was introduced in comparison to when confirmation rewards were present.

Besides these main research questions, in the simulations we also investigated whether the effects and macro consequences of social control mechanisms are dependent on assumptions about *individual rationality* and access to information or not. We examined the effect of social control mechanisms and structural configurations on intergroup conflict under four behavioral models. These models differed in the assumptions regarding the level of calculating rationality of individual actors and regarding the amount of information individuals have access to. We were particularly interested in the effect of segregation on the likelihood of conflict under the different model specifications. Simulation results showed that under certain structural conditions, rational individuals with higher amount of information were more likely to be trapped in harmful conflict than less rational actors. Furthermore, rigid assumptions about individual rationality slightly strengthened the effect of segregation on intergroup conflict.

We demonstrated that not only segregation, but also *other properties of the social network* are associated with the likelihood of intergroup conflict. For instance, minority hostages can suppress mobilization, bridging ties can play a brokerage in the spread of contribution, and loosely connected subgroups may either elicit or inhibit intergroup conflict depending on the behavioral assumptions.

6.1.3 Temporal embeddedness and intergroup conflict

Besides the structural embeddedness of behavior, *temporal embeddedness* plays also a crucial role in intergroup relations. The historical record of previous encounters and prospects of future relations have firm effects on present attitudes and actions in the intergroup context. Empirical examples of inflating clashes and durable conflicts between groups challenged us to explain the emergence of such scenarios. Fortunately, intergroup relations are more frequently peaceful than violent, which drove us towards the exploration of the conditions under which peaceful coexistence prevails.

Our explanation of the dynamics of intergroup relations concentrated on mechanisms at the individual level that aggregate to macro scenarios. Individual decisions are strongly affected by both experiences from the past and expectations about the future. With regard to the temporal embeddedness of action, in this study, we adopted a view of *bounded rationality*. We claimed that individuals do not recall all past events when making decisions and neither do they make extensive calculations about long-term consequences. Instead, they are guided by *simple behavioral heuristics*. With regard to the question, what are these guiding rules, based on recommendations of previous literature we formulated hypotheses about the existence of three mechanisms, namely criticalness, reinforcement learning, and reciprocity. *Criticalness* dictates contribution when the individual expects a single decision to change the outcome of the competition. *Reinforcement learning* prescribes sticking with a decision that gave rise to a satisfactory outcome. It calls for a change when the outcome was unsatisfactory. *Reciprocal mechanisms* evoke peaceful responses to observed peaceful behavior of others and evoke retaliations to previous contributions to conflict. Individuals might reciprocate the collective action of the other group, but also actions of relevant individuals. In this study, we have taken into consideration all these behavioral mechanisms. The main argument to present this approach was that behavioral strategies might differ between and also within actors.

An original aspect of this research was that it tested the influence of different behavioral heuristics on individual decisions simultaneously in the experimental laboratory. We could test the effect of temporal embeddedness as in parts of the experiment subjects played *repeated* IPG games. Unlike in the *single-shot games* where subjects did not receive any feedback on previous outcomes, in the repeated games they received information about the outcome of the previous game and eventually about the decision of their neighbors. As structural conditions were introduced also in the repeated games, *interaction effects of structure and time* could be also tested.

Let us now provide a brief summary about the main *research questions* (**Q**) about the effect of temporal embeddedness and about the interactions of structure and time. We also summarize the hypotheses (**H**) that were formulated for these questions and the conclusions that were drawn from the experimental results (**R**).

Q *How do intergroup relations change over time? Are there typical scenarios, such as an endless regression of conflict or a spiral of peace? Do these scenarios differ according to structural conditions?*

H We predicted that some typical scenarios would occur, such as stable peace, durable conflict, a spiral of peace, and a spiral of conflict. Our forecast was that stable peace and a spiral of peace would be more likely in a mixed seating configuration and durable conflict and spiral of conflict would be more likely when segregation is high.

R **Results show that the scenarios of spiral of peace and stable peace occurred frequently when members of the two teams were seated in a mixed pattern. In sessions with high segregation, durable conflict emerged, but with changing fortune for the two groups and not gradually, as the spiral of conflict hypothesis had predicted. Contribution rates increased due to the introduction of social control between experimental parts.**

Q *What are the simple heuristics that guide individual choice and as a consequence, are responsible for the emergence of macro scenarios? Are criticalness, reinforcement and reciprocity important determinants of individual action in repeated intergroup relations?*

H Our hypothesis was that individual choice in the experiments would be guided by simple behavioral rules that are partly based on experience from previous encounters and partly based on expectations about future interactions. We predicted that the combination of criticalness, reinforcement learning, intergroup reciprocity, and local reciprocity mechanisms would influence actual individual decisions.

R **Results justify the use of a combination of different behavioral principles in the explanation of individual contribution propensities. The hypothesis about criticalness is confirmed, but with an adjustment for forward-looking bandwagon tendencies. We found surprisingly high willingness for contribution when subjects expected a victory for their group in the forthcoming game. Results show however that reinforcement learning does not**

work symmetrically, as we did not find evidence of reinforcement towards a contribution choice. Moreover, there is little support for the presence of intergroup reciprocity. Only a previous defeat elicited significant retaliations, but this might also be caused by simple hawkish attitudes. Local reciprocity was found conditional to personal characteristics of the subject and conditional to the gender and group affiliation of the neighbor.

6.2 Implications and societal relevance

In this section, we discuss some implications of this study for societal applications. Our model was built on the presumption that the origin of intergroup conflict is the *competition* between the groups for certain limited resources. If there is no competition, intergroup relations have a completely different nature. In most cases, the lack of negative interdependency is a guarantee for peaceful coexistence, even if there are dark memories of the past or there is a high level of segregation. In these situations, there could be other difficulties, such as coordination problems. For instance, using the same standards would be beneficial for interacting groups, but naturally enough, none of the groups are keen on changing its own established system. This research has not directly dealt with such situations, but it might have some valid implications also for these cases. The underlying mechanisms of social control and decision heuristics work in a similar way in these contexts, causing effects of structural and temporal embeddedness. For example, people experience strong confirmation pressure from relevant others to use identical standards, such as speaking the same language as they speak. Consequently, members of a minority, who are exposed to members of the majority group, are easily forced to adopt the standards of the other group. Furthermore, segregation of the social network directly leads to the same standards within the segments, such as to the evolution of dialects in remote parts of a language area.

A closer correspondence can be made between the results of this study and situations in which groups actually compete for a certain goal. However, *not all competitions are social dilemmas*, as mutual collective action does not always have suboptimal consequences. Intergroup rivalry might result in a draw that is not harmful for either side or the groups may reach a *compromise* and divide the public good. Furthermore, competitions might have a positive value for the larger community such as in the case of team sports or competition between R&D teams. In these cases, there is still a free riding problem within the groups, but mobilization has positive externalities for the other group. The community has an interest in enforcing participation. Therefore policy suggestions that follow from our analysis are the opposite to the case of harmful competition. In order to facilitate contribution, dense relations within the group and strong selective incentives are needed. Besides, ties between members of the rival groups should be minimized or should be kept at a neutral level.

These implications also hold for in-group *collective action problems*. The key mechanism to the solution of these social dilemmas is the distribution of strong selective incentives and their internalization. Our model predicts, similarly to Coleman (1990), that this works best in a dense network with transitive ties. On the other hand, strong confirmation pressure in a dense network might lead to widespread contribution as well as to overall defection. In this way, behavioral confirmation is double edged for cooperation, just as in the case of approval exchange (Flache, 1996).

There are possibly many more situations, to which the results of this study can be implied. However, the major goal of this research was to understand and explain the emergence of *mutually harmful intergroup conflict* situations and to show under which conditions these can be avoided. In this respect, our major contribution was to reveal mechanisms that explain how and why conflicts emerge as a result of structural and temporal embeddedness of individual actions.

In relation to *structural embeddedness*, our major conclusion was that *segregation* is likely to promote intergroup conflict. Starting from a certain structural configuration, an increase in the number of relations within the group or a decrease in the number of relations toward members of the other group will definitely not facilitate conflict resolution. For instance, this result supports policy arguments to encourage interethnic relations and decrease residential segregation in order to help the resolution of ethnic conflicts. However, as both theoretical and experimental results suggest, such a policy will not always be effective. Conflict can just as likely occur in middle ranges of clustering as it can in the completely segregated setup, due to weak traitor incentives and the strong influence of (internalized) fellow pressure. Furthermore, already a few zealots might be sufficient to initiate conflict even in a relatively mixed configuration. In these cases, especially if the costs of desegregation are high, it is better to seek institutional or external solutions for the management of damaging intergroup relations.

Desegregation policies also have to be aware of *other influential network properties*. As the examples of minority hostages and bridging ties show, it is often more efficient to place certain persons to the right position than to implement a costly wide-range desegregation policy. A presence of a small minority in an otherwise homogenous environment might be sufficient to excite sympathy for the rival group and to suppress activism. Bridging ties between isolated subgroups might play a crucial role in the dissemination of radical, but also of peaceful attitudes. We illustrated that one gatekeeper is usually not enough to play such a brokerage. Multiple bridging ties do this better, when supported by strong bridgeheads, that is by influential group members, who are in contact with the intermediary persons. To keep hawkish tendencies at a low level, the most efficient way is to close down the radical groups' bridging contacts to the outside world, and thus to isolate radicalism. With regard to the question of how can such structural strategies be implemented in practice, one can benefit from recommendations of applied social network analysis (e.g., Kratzer, 2001; Leenders, Kratzer, and van Engelen, 2002).

There will always be heroes who are prepared to die to advance the position of their group. In the experiments of this study, a significant portion of subjects sacrificed the received bonus, even when their decision remained anonymous and the effort was completely fruitless. We did not find any personality trait that would strongly correlate with such behavior, except prosocial orientation, to a certain extent. The interesting result that zealots often have prosocial attitudes implies that it is the individualistic “invisible hand” that would save intergroup relations from lethal rivalry.

Individualization is a remedy for conflict also in another sense. Apart from fanatics, most individuals are sensitive to social control effects. Experimental results show that social control from fellow group members even in an internalized form impacts upon individual action. As fellows are more likely to demand contribution, downgrading fellow contacts would further conflict resolution. In *isolation*, individuals build primarily on their egoistic incentives and do not make costly sacrifices for the group.

Minority hostages are also isolated from their fellows. Besides free riding incentives, an additional motivation for them to defect is traitor pressure from their neighbors. Minority groups do not have many choices, apart from assimilating into their environment. If they are too few, they cannot even hope to evoke tolerant attitudes from their neighbors.

As experimental results show, traitor incentives are only internalized and affect decisions when group affiliation is combined with visible characteristics, such as gender. In practice, social relations are much stronger than in the laboratory. We could therefore expect the transmission of immaterial traitor rewards between members of the different groups. This prescribes another route out of troubled intergroup relations, which is *globalization*. As relations across groups are propagated and network paths between individuals are shortened, exposure to other cultures increases and individuals are influenced by stronger traitor incentives that work against the emergence of intergroup conflict of a traditional kind.

In relation to *temporal embeddedness*, we concluded that intergroup competition that is repeated in a similar fashion often has similar outcomes because individual decisions are also relatively stable over time. On one hand, peaceful relations are likely to be preserved. On the other hand, conflict is also frequently a follow-up of previous conflict, although the winners of the competition might change. The pessimistic prediction that vengeance drives conflict toward ultimate escalation is not supported by our experimental data. Individuals were tempted again and again by the free rider benefits of peaceful behavior.

In general, our results suggest that effects of temporal embeddedness are not as strong as effects of the payoff structure and social control. Among temporal effects, the *shadow of the future* proved to be more influential than the shadow of the past as individuals were rather forward than backward-looking. We demonstrated that the proper model of action should also consider influences of previous encounters. Experiments showed that subjects were reinforced to adopt peaceful behavior, but they were not likely to learn activism in the Pavlovian sense. In general, being aware of the implications of lessons learned in the past for one's future were developed over time and reinforcement was detected more often in later stages of the experiment. The *asymmetry of learning effects in favor of peaceful attitudes* gives hope for endogenous solutions of intergroup conflict. In principle, a deadlock of a clash can be broken by a critical mass of defecting initiatives: others will follow as they learn over time the benefits of peaceful coexistence. This process requires a lot of patience if people indeed need time to adopt reinforcement principles. In practice, this could be facilitated by making the negative consequences of the past transparent and by helping coordination between initiators of peace.

Another positive result of the experiments for conflict resolution is that subjects *did not retaliate* the collective action of the rival group except where they were provoked by other incentives. This suggests that vengeance can be avoided in the intergroup context. On the other hand, we found conditional support for the existence of local reciprocity. Under certain conditions subjects retaliated the previous contribution decisions of other subjects seated adjacent to them. This result implies that unlike in the intergroup context, revenge can be an important mechanism in interpersonal relations. Furthermore, a contagion of interpersonal retaliations might gradually lead to an outburst of intergroup conflict.

Still, such macro dynamics were not detected in the experiments. There was no indication that conflict would have been solved as individuals adopted reinforcement over time. Furthermore there was also no sign that local retaliations were leading to a spiral of conflict. These scenarios did not occur in the repeated games because of the presence of intervening effects, of which the most influential was the role of *expectations about the future*.

Subjects did evaluate possible outcomes of subsequent games, which is very much in line with the rational choice perspective on individual action. On the other hand, these evaluations were far from being sophisticated and perfect, which shows that individual rationality also in this aspect has strong limitations. Subjects helped their group only if they felt that a single contribution could change the outcome. This happened frequently as they largely overestimated the criticalness of their choice (for a similar result, see Kerr, 1989). Furthermore, if subjects expected a victory of their team, they did their best to take their shares not only from the rewards but also from the group efforts. However, they largely overestimated the chances of their group to win and the efforts made by their fellows (for a similar result, see Brewer, 1996a; 1999). All this resulted in a *blind optimism* about the future and their individual role. After all, this optimism explains hawkish tendencies in the experiment and much of why intergroup conflicts were observed so frequently.

This implies that a *cognitive approach* of conflict resolution might have an enormous power. Methods that can lower individual expectations and beliefs about criticalness would contribute substantially to peace. In practice, this means that the media and for that matter anyone, who has control over the flow of information, carries a large responsibility with regard to the management of intergroup relations. Major channels of information could be used effectively to educate individuals about realistic thinking and the irrationality of radicalism. Manipulating individual concepts about the future in this positive sense would help the establishment of peace in the world. However, this is extremely difficult as the groups have a collective interest that is in contradiction with this end. They have an incentive to advertise criticalness principles, such as stating that "we need *you* to win". Besides, bandwagon effects explain why it is always in the interest of the most popular political party to publish survey results before the elections.

6.3 Limitations

In this section we critically discuss the limitations of this study and the problems we face in generalizing the results. Our conclusions have remained at a highly abstract level, as it was not our intention to provide precise policy instructions. We would need to be much more cautious with regard to relating the results to concrete cases of intergroup competitions. Empirical situations have special characteristics that should be taken into account in their analysis. Most limitations of this study concern such *complexities* that actually shape intergroup relations in practice. We intended to stay at a relatively abstract level in order to focus on fundamental processes of structural and temporal embeddedness. In this section, we summarize some points to which we need to add complexity to understand concrete cases of intergroup conflict better.

Some of the restrictions and assumptions we needed to make can quite easily be relaxed. Both in the simulations and in the experiments, we considered a specific *group size*. As we discussed, in small groups individual decisions are more likely decisive, therefore it is easier to establish collective action, even in the absence of social selective incentives (cf. Olson, 1965). In large groups, relative differences in group sizes are more important because of the disproportional effect of segregation. We also discussed the importance of minimal contributing sets (MCS). The higher the *thresholds* to establish collective action, the higher the chances for intergroup peace are. The *payoff structure* of the IPG game can easily be adjusted to model different empirical situations. For instance, the assumption of zero reward for peace can be relaxed.

Furthermore, similar assumptions on structural and temporal embeddedness can be applied to completely different interdependence structures, such as to *other team games*. For instance, there might be objections to the too *narrow concept of clash* in the IPG game. In case of civil war, urban gang fights, and other violent competitions, when the number of participants is not exactly equal in the groups, conflict outcomes are still lethal for both sides. A game that takes this into consideration is illustrated in

Figure 6.3.1. A significant change compared to the original IPG game is that overall clash is no longer a Nash-equilibrium. However, by comparing the structurally embedded version of this game to the structurally embedded IPG game, we would not detect significant differences. Both the conditions for contribution being a dominant strategy and the conditions for overall contribution being a suboptimal equilibrium are exactly the same as those presented in this study. Hence, our major conclusions might be also generalized for these situations.

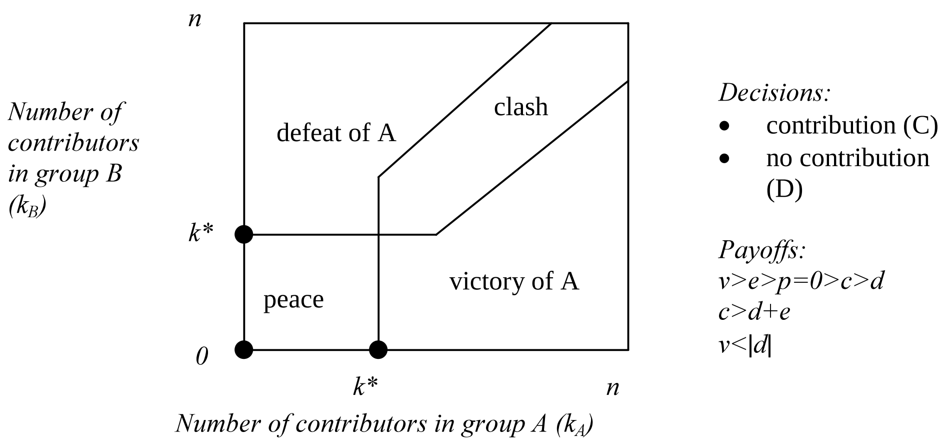


Figure 6.3.1 Graphical representation of a game with an extended concept of clash
Note: Nash equilibria are indicated by bullets.

It was necessary to decrease abstraction and consider some intervening variables already in the experiments in comparison to the simulation analysis. In the simulations, we assumed a linear *utility function* that was independent from the utility of others. Furthermore, we assumed that the set of players is *homogeneous* with regard to the applied behavioral rules. In the experiments, we found that utility functions are sensitive to attitudes towards risk and to rewards of others. Besides, experiments showed that subjects also differ in their strategic behavior. Subsequent simulations that are follow-up of this study could incorporate this and could introduce parameters about the distribution of social orientations, risk attitudes, and decision heuristics in the analysis. Advanced models might also incorporate further individual differences. An example is a distinction between forward-looking leaders and backward-looking followers or aggressive and peaceful types. Such differences might be caused by intrinsic values of fighting or differences in individual incentive structures. For instance, some people may have a stronger preference for the public good of victory, and in this privileged situation they would therefore be more inclined to make sacrifices for their group (cf. Olson, 1965).

Another major discrepancy between the simulations and the experiments is that we did not consider *repeated interactions* in the simulations. The theoretical analysis of iterated intergroup competitions would be a natural development of subsequent research that might build on the experiences of our experiments.

There are also major limitations to the generalization of experimental results to empirical situations. Although intergroup competitions often have the same target over and over again, such as regaining control over a certain territory, *interdependencies are seldom repeated exactly in the same structure*. Additionally, history cannot repeat itself because structural contacts between individuals change. Neighbors move to another house, old friendships dissolve and new friendships develop, or member countries of rival alliance blocks change their dyadic relations. These changes might be partly the consequence of previous intergroup conflicts (for empirical evidence see Doherty, 1990; Liska and Bellair, 1995; Liska, Logan, and Bellair, 1998). Extreme examples of feedback effects of intergroup conflict on the community structure are the incidences of massive refugee flows. In most cases, feedback effects promote a segregation process in the community structure. A segregated network can remain a stable configuration, in which intergroup conflict is repeated. As segregation and conflict stabilize it is difficult to intervene and help the community shifting to peaceful equilibrium. Major research implications proceeding from these empirical processes would see the formation of a dynamic model that incorporates feedback effects of conflict on the residential structure.

Furthermore, in the long run, *group boundaries may also change*. Assimilation, for instance, can be considered as an optimal long-term strategy to avoid the emergence of harmful conflicts between groups.

A major concern for the validity of this study is the difficulty of the *measurement* of payoff parameters in empirical situations. Even the utility of material rewards of intergroup competition, such as territory, is difficult to determine. There are even more problems with conceptualizing the value of immaterial rewards, such as social identity, and also with providing meaningful parameter values for rewards of social control.

A possible objection to the behavioral assumptions of this study is that *social control might take quite different forms* in interpersonal relations than that which we have explicated. There are several regularities at the micro level that can be related to undesired macro consequences (cf. Coleman, 1990). For instance, not every acquaintance is a good one. There can be interpersonal feuds not only between fellows, but also between members of the competing groups. Intergroup conflict very often originates or results in negative interdependencies between individuals. Furthermore, some ties are stronger and transmit more efficient social control than others. There are also qualitative differences between different social contacts. Certain individual characteristics, such as gender, might imply that ties are very different in terms of content (e.g., Burt, 1998). As we hear about contemporary Romeos and Juliets, we know well how the meaningless fight between the modern counterparts of Capulets and Montagues lead to personal tragedies (cf. movies “Before the Rain” or “Torn Apart”). Moreover, social control can also be completely independent from the intergroup context. Policies that aim to strengthen these forms of social control are ineffective for the improvement of intergroup relations.

Another possible objection to the behavioral assumptions of this study is that *social control is not always gratuitous to produce*. The establishment of effective sanctioning systems, such as norms, can also be characterized as a social dilemma, which includes a second order free rider problem (Heckathorn, 1989; 1993). Furthermore, sanctioning those who do not take part in the enforcement of sanctions includes higher order free rider problems (Kuran, 1995; Wintrobe, 1995: 59). An example of how such dilemmas can be solved by *hierarchies* is provided in Stalinist regimes, where top leaders executed the highest level officials who failed to punish those party members who did not enforce enough “contributive” behavior in the population.

The role of in-group hierarchies is another factor that was disregarded in this study. Formal hierarchical control is an effective way of solving social dilemmas, but it does not improve intergroup relations. On the other hand, internal hierarchies clarify responsibilities within the groups for certain actions. If the set of powerful actors is clearly defined, *negotiations* between the two sides are also easier. Agreements between leaders can bring settlements over highly segmented division lines. Furthermore, a *higher body* that controls, at least to a certain extent, actions of the groups, such as the teacher in the classroom or the United Nations in international conflicts, can impose on rival parties that they be committed to peaceful coexistence. This higher body can induce positive changes also indirectly by manipulating the structural conditions of the intergroup competition. For instance, national governments can implement and even enforce desegregation policies or can change the competitive character of intergroup relations, for example, by dividing the public good.

In general, *institutions* can play a crucial role in conflict resolution, even if they are not external supervising powers but established internally by opposite sides. Intense competition can sometimes be defused by institutional interventions of seemingly little importance (for instance, by a fixed negotiation schedule, a referee or a mediator). Similar help might come from the introduction of *new division lines*, from the intervention of *third parties* or from the *media*. These interventions might change the incentive structure, might manipulate the available information or create cognitive interdependencies. However, as Rubin (1980: 385) claims, this help is only conditional: “It appears that certain tried and true techniques of third-party intervention, such as the introduction of communication between the parties, the recommendation that the disputants consider multiple issues as a package, and the use of such issue identification procedures as role reversal, facilitate concession making and agreement only when conflict is relatively low in amount and intensity. When conflict is intense, these very techniques may prove ineffectual and may even exacerbate the conflict.”

There are probably several more limitations and objections to this study that concern the complexity of empirical situations. Despite these reservations, our analysis was nevertheless able to demonstrate the effect of intergroup competition, structural and temporal embeddedness on the likelihood of group conflict. This may help conflict resolution in empirical situations that are along these lines.

6.4 Directions for future research

We hope that future studies of intergroup conflict and peace can benefit from the results of this research. In this section we provide some suggestions for subsequent investigation. *Analytical developments* might concern different types of equilibria in repeated games or the derivation of aggregated consequences of bounded rationality of players. *Ex-post simulations* in close relation with our experimental findings might be designed to model macro consequences of certain combinations of behavioral mechanisms given certain distributions of personal characteristics in a predetermined population. Further simulation developments might include computer tournaments, genetic algorithms, cellular automata, neural networks, and other *agent-based simulation* techniques for the analysis of repeated interactions and structural dynamics. *Future experiments* might be concentrated on the role of communication, on the process of identification with the group, on structural changes, or on the emergence of social control. *Field experiments, case studies,* and *survey research* might provide ways to test theoretical hypotheses empirically and might reveal further underlying mechanisms for the explanation of intergroup conflict.

This constitutes a wide range of possibilities. Section 6.4 is devoted to provide more concrete suggestions for some of the specific directions that were originally planned to be part of this study.

6.4.1 Analytical developments

As this research has demonstrated, social control can provide a structural solution to the collective action problem within groups. On the other hand, this solution might create a social dilemma in the intergroup context, especially in a segregated setting. Individual contribution can be the dominant choice for everyone. This choice would result in an aggregated outcome that is suboptimal compared to peace. Is there a way out of this social trap? How can the harmful outcome of overall conflict be avoided? Similar to Axelrod (1984), we can argue that *infinite repetitions* provide a resolution to conflict. However, the strategy that triggers collective action of the other group is *not* in equilibrium against itself in the structurally embedded IPG game. Hence, intergroup reciprocity has no theoretical support in this specific context.

A reciprocal strategy that offers a way out of the social dilemma is TIT FOR ONE TAT (TF1T). It defects in the first round and contributes only, if there was at least one contributor in the other group in the previous round. A homogenous population of the TF1T strategy is an equilibrium, if the shadow of the future is bright enough (the probability of continuation is high). The proof is similar to the derivation of the possibility of TFT|TFT equilibrium in the two-person Prisoner's Dilemma (PD) (Axelrod, 1984)[1].

6.4.2 Individuals with short-term memory: analysis of two-state Moore machines

Subsequent research that assumes intentional individual action should examine *what are the successful individual strategies in long-term intergroup competitions*. For the purpose of this theoretical inquiry, hardwired strategies should be targeted that have the same conditional rules throughout the repetitions.

To impose a restriction on the number of possible strategies and to make the analysis more realistic, research interest could be limited to the analysis of strategies with short time memory. An example would be to consider strategies with only one round of memory. These behavioral rules are called *two-state Moore machines* (Linster, 1992; Bicchieri, 1997; Klos, 1997). A Moore machine is a deterministic automaton that provides conditional responses for all possible combinations of previous outcomes and decisions. Different Moore machines can represent permanent

---

[1] For the sake of simplicity, let us introduce the notations S_i for $f_i s$, T_i for $g_i t$, and B_i for $f_i b$. It is sufficient to show that the homogenous TF1T population cannot be invaded by ALL C nor by an alternation of D and C. For the first statement,

$$S_i + \frac{\delta(T_i + B_i + e + d) + \delta^2 (S_i + B_i + v)}{1-\delta^2} \leq \frac{T_i + B_i + e}{1-\delta} \qquad (6.4.1.1)$$

should hold for all *i*, where δ is the probability of continuation. This simplifies to

$$\delta d + \delta^2 (B_i + v) \leq T_i + B_i + e - S_i . \qquad (6.4.1.2)$$

A continuation probability $\delta>0$ exists, if $d+v \leq T_i+e-S_i$ is met for everyone.

For the second statement (the homogenous TF1T population cannot be invaded by an alternation of D and C),

$$S_i + \delta(S_i + d) + \frac{\delta^2 (S_i + B_i + c)}{1-\delta} \leq \frac{T_i + B_i + e}{1-\delta} \qquad (6.4.1.3)$$

should hold for all *i*. This simplifies to

$$\delta d + \delta^2 (B_i + S_i + c) \leq T_i + B_i + e - S_i . \qquad (6.4.1.4)$$

A continuation probability $\delta>0$ exists, if $d+c \leq T_i+e-2S_i$ is met for everyone.

trouble-makers, firm free riders, revenging individuals, and many other types. Examples of Moore machines are shown in Table 6.4.2.1. Cells show individual decisions in round *r* as functions of the outcome and decision in round *r-1*. These strategies use only restricted information from the previous round. They are conditional on the previous collective outcome and not on the exact number of contributors. There are two possible past decisions and if we assume information about the establishment of collective action, six possible outcome states (see Table 6.4.2.1). Every strategy is defined by its binary responses to these twelve different states. Therefore, strategies can be expressed as 12-bit strings or can be uniquely represented by the decimal number translations of these strings (these are shown as strategy codes in Table 6.4.2.1). Additionally, for each strategy, the very first choice also has to be defined. This could be a quite crucial determinant of fitness, as it is the case with a property of being *nice* in the two-person PD (Axelrod, 1984).

Table 6.4.2.1 Examples of two-state Moore machines

outcome in round r-1		*victory*				*peace*		*clash*		*defeat*			
collective action in round r-1		*yes*		*no*		*no*		*yes*		*yes*		*no*	
own decision in round r-1		C	D	C	D	C	D	C	D	C	D	C	D
Strategy	*code*												
ALL D	0	D	D	D	D	D	D	D	D	D	D	D	D
Change	1365	D	C	D	C	D	C	D	C	D	C	D	C
Likely Critical	2096	C	D	D	D	D	D	C	C	D	D	D	D
Win-Stay Loose-Change	2709	C	D	C	D	C	D	D	C	D	C	D	C
Intergroup TFT	3135	C	C	D	D	D	D	C	C	C	C	C	C
Intergroup GRIM	3775	C	C	C	D	C	D	C	C	C	C	C	C
TFT + Bandwagon	3903	C	C	C	C	D	D	C	C	C	C	C	C
ALL C	4095	C	C	C	C	C	C	C	C	C	C	C	C

Notes: Cells contain decisions in round *r*. C=contribution (binary value 1), D=defection (value 0).

Among the strategies that are displayed in Table 6.4.2.1, “Likely Critical” resembles the criticalness behavioral principle of the present study. However, the “Likely Critical” Moore machine is strictly backward-looking, unlike subjects, who base their decisions on expectations about the next outcome. “Intergroup TFT” resembles intergroup reciprocity and the strategy “Win-Stay Loose-Change” is the counterpart of reinforcement learning. The main difference is that in the experiments we did not consider deterministic strategies, and we formulated our hypotheses in probabilistic terms, since other factors also influence decisions.

Strategies that were shown in Table 6.4.2.1, do not use information about previous decisions of relevant others. However, one of the major conclusions of this study is that human action is strongly determined by social control from related group members. For instance, previous decisions of neighbors might influence the behavior of individuals in the form of local reciprocity. In order to analyze what are the successful individual strategies *in different neighborhoods* in long-term intergroup competitions, we need to consider much more complex strategies that are also conditional on the behavior of relevant others.

The conceptualization of what is successful, however, is quite problematic. Strategies can have overwhelming success in certain neighborhoods and ignominious failure in others. Even in strategically equivalent network positions or at a fixed location different strategies may do best, depending on the distribution of strategies among others. Therefore, a study that aims to search for successful strategies has to make simplifying assumptions, for instance by considering simple network structures, such as a cellular world (cf. von Neumann, 1966; Nowak and May, 1992; Nowak and Vallacher, 1998; Hegselmann and Flache, 1998; Flache and Hegselmann, 1999a).

6.4.3 Computer tournaments and evolutionary selection of successful strategies

The study that examines Moore-machines and searches for successful individual strategies would need to specify the *population* of strategies that is analyzed, the *selection criteria*, and also the *algorithm* that is used to determine success.

To avoid the criticism that strategies are chosen in an ad hoc fashion for the analysis, strategies should be systematically selected or sampled (Binmore, 1998). For instance, similar to the two-person PD computer tournaments of Linster (1992), a careful design can encounter *all possible* Moore machines with one round of memory. On the other hand, for the illustration of certain tendencies, a *qualitative selection* also suffices, as in the computer tournament of Axelrod (1984) or in the case of strategy selection by Kollock (1993).

One possibility to determine the success of strategies is based on an evolutionary perspective (Maynard Smith and Price, 1973; Dawkins, 1976; Maynard Smith, 1982). *Ecological success* covers all phases of a strategy-life from birth to overwhelming dominance and it consists of three criteria: initial viability, robustness, and stability (Axelrod, 1984). *Initial viability* means that a strategy is successful as a mutant in a uniform population. The fabled TFT, for instance, individually is not viable in an ALL D population in the two-person iterated PD. *Robustness* means that a strategy does well in mixed environments. In the iterated PD computer tournaments of Axelrod (1984) TFT and slight modifications of TFT were the most robust strategies. If all possible two-state Moore machines participate, GRIM[2] is the most robust strategy (Linster, 1992). Finally, a strategy is *evolutionary stable*, if no other strategy can invade its uniform environment. In the iterated two-person PD, ALL D is always stable and TFT is evolutionary stable, if the discount parameter is sufficiently high (Friedman, 1971; Axelrod, 1984).

[2] GRIM is also known as FRIEDMAN (Axelrod, 1984). It contributes in the first round and never contributes again, once the other player has defected.

Computer tournaments can be used to investigate the robustness of different strategies (cf. Axelrod, 1984). Evolutionary algorithms can test also other criteria of ecological success employing the principle of "survival of the fittest". *Learning* and spread of successful heuristics can either happen *globally* in the entire population as the best performing strategies reproduce (Axelrod, 1984; 1997a) or can take place *locally* as individuals adopt better codes of behavior from neighbors (Ellison, 1993; Chwe, 2000). If strategies are also contingent on previous actions of neighbors, the amount of possible codes is extremely large, even if only two-state Moore machines are considered. *Genetic algorithms* are able to cope with such a variety as they introduce mutation and crossover to the reproduction process (Holland 1975; Holland and Miller, 1991; Macy, 1996; Klos, 1997). Hence there is no need to encounter all possibilities; simulations take care of the selection of successful behavioral traits.

6.4.4 Fully connected networks

In this study, we assumed that all interpersonal relations are equally important. However, the impact of good friends and acquaintances on individual behavior might be radically different. An advanced model design might take into account that the size of social control effects is dependent on the strength of the tie between given individuals. A further complication is that individual relations might be asymmetrical. The subjective strength of a tie could be different for the connected persons. Subjective values, ranging from zero (no contact, indifference) to maximum influence (dictatorship), determine the weights of social control effects.

The consideration of relational strength adds substantial complexity to the model and calls for efficient computational methods. *Agent-based simulation* techniques could be used for this purpose. Agent-based simulations assume that individuals have intentions and make choices that affect other agents (Macy, 2002). With their help, emergent macro phenomena can be "grown" from simple rules of micro behavior (Epstein and Axtell, 1996). For the study of structural embeddedness of intergroup dynamics, neural network models might provide a powerful example of agent-based techniques (Bainbridge, 1995; Macy, 1996; Nowak and Vallacher, 1998). *Attractor neural networks* that create a compelling analogy of mechanisms in the brain and in society would particularly fit the purpose of the study of social network dynamics (Minsky, 1985; Kitts, Macy, and Flache, 1999). In attractor neural networks, nodes are agents and ties between nodes have weights corresponding to the strength of the relationship. Agents adjust these weights over time due to their network strategy. Meanwhile, they receive from and impose social control on others proportional to the strength of the tie. These relational strategies combined with the decisions of whether to participate in intergroup conflict could be the basic decision elements in an agent-based simulation model of a subsequent research program.

6.4.5 Structural dynamics

In repeated interactions, intergroup conflict or peace has a feedback effect on the network structure. People might reconsider, who are their real friends, depending on their previous actions. As a consequence of intergroup conflict and of dyadic social control, some ties might become stronger and other ties might dissolve almost entirely.

In order to increase future satisfaction, individuals have an interest in strengthening their ties with those who rewarded them in previous rounds with either selective incentives or with behavioral confirmation. All this is at the expense of less profitable relations. This kind of *structural learning* (cf. Kitts, Macy, and Flache 1999) is an optimal network strategy in the short run. However, it does not provide a way out of ongoing intergroup conflict. After clashes, structural learning provides a directive to decrease the strength of ties with members of the other group and to increase the strength of ties within the group. This leads to higher segregation and durable intergroup conflict, which Newcomb (1947) called *autistic hostility*. It could happen that the suboptimality of overall participation disappears and group members enjoy continuous conflict, since they are compensated by selective incentives and behavioral confirmation. But this is definitely not a desirable resolution for the social dilemma. Individual structural strategies that show the way out of harmful clashes should be of a different kind.

6.5 Walking out of social traps

It is certainly challenging to find ways out of the spiral of ongoing conflict and segregation without an inclusion of third parties or institutions. In this section, we discuss some intuitively appealing *endogenous solutions* of *reciprocity*, *tolerance*, *raising new issues*, *assimilation*, *exit*, and *positive discrimination* of members of the opposite group. Subsequent research in the directions we considered in Section 6.4 might provide theoretical support for these solutions. Moreover, we would like to demonstrate that suggested solutions can be explicitly linked to empirical situations and provide ways out of the social trap of harmful intergroup relations.

6.5.1 The emergence and collapse of live-and-let-live systems

In this section, we consider an endogenous solution of intergroup conflict by *reciprocity*. At the same time, we also illustrate how this research and its proceedings might help to explain intergroup conflicts better than previous studies.

In his highly influential work, Axelrod (1984) argues that the relevance and success of TFT in the repeated PD can help to explain the emergence of the live-and-let-live system in the trench warfare of World War I. There is no doubt that mutual restraint became possible in the antagonistic fight because of the static nature and infiniteness of the opposition. However, the two-person PD model of trench warfare is too simple to represent the interdependencies on the battlefield. First of all, opponents cannot be considered as unitary players. Soldiers as individual actors start socializing with soldiers from the other side and they are the ones, who apply the live-and-let-die principle otherwise. It is completely their decision to shoot accurately or not, although there is extremely high pressure from their commanders to comply with group efforts. On the other hand, besides the effort of pulling the trigger, they face a much higher cost of breaking the strong moral rule of “not to kill”. Because of these concerns, trench warfare is better described as a team game, in which individual actions aggregate to the collective performance of the armies. As the competition at the national level can indeed be described as a two-person PD, for the two levels of interdependencies the IPG game provides an accurate representation.

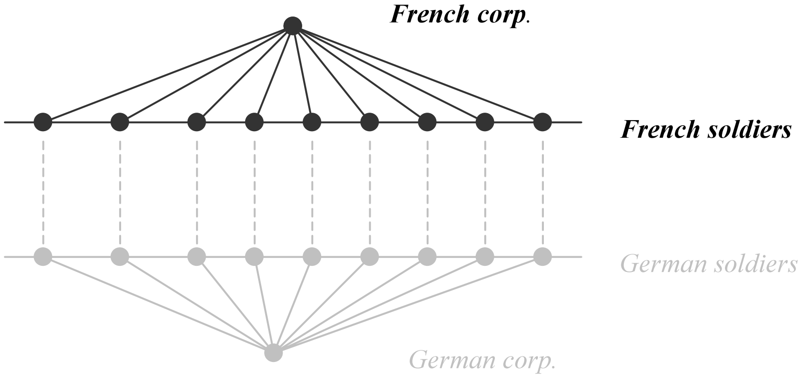


Figure 6.5.1.1 Application of the structurally embedded IPG game to trench warfare in World War I

Social control effects entering the model as soldiers have an influence on the actions of comrades and on warriors on the opposite side. These include behavioral confirmation as well as social selective incentives between fellows and traitor rewards towards people on the other side. Furthermore, in the form of orders, rewards, and punishments, officers distribute strong selective incentives between their troops. Social control effects are transmitted along the spatial arrangement of soldiers on the front line that is more or less stable over time (see Figure 6.5.1.1 for a rough illustration).

To get a closer impression of the social dilemma of trench warfare, consider a numerical example. Let us fix the private shares of the public rewards from the intergroup competition to $p=0$; $v=2$; $d=-5$; and $c=-4$. Let us assume a benefit for not shooting that contains the value of a bullet, a spared effort, and moral satisfaction of being $e=1$. Additionally, side payments of social control are in effect. The supervising officer enforces compliance strongly by $s=4$. Behavioral confirmation from each neighboring comrade is set to $b=0.5$. Furthermore, assume that traitor rewards that concern sympathy to humans on the other side are negligible ($t=0.01$). In the structurally embedded IPG game of this payoff structure, shooting is a dominant strategy of all soldiers. Independently of what others do, individuals are always better off by fighting. Rationality dictates the application of the live-and-let-die principle for everyone. The aggregated consequence of overall participation in the bloodshed is dominant strategy equilibrium. However, in this outcome everyone is rewarded less (1) than in overall peace (2.01), which shows the social dilemma character of trench warfare.

Due to the static nature of opposition, the interdependency is repeated in the same structure. As Section 6.4.1 shows, if the likelihood of continuation is large enough ($\delta \geq 0.657$), the strategy TF1T ensures a possibility of an endogenous solution for the conflict. The emergence of the live-and-let-live system can be explained by the overall adoption of this rule. On the other hand, the easy collapse of this system is the consequence of the extreme vulnerability of this equilibrium to a single trembling hand (cf. Selten, 1975).

6.5.2 Glory for tolerance

Another internal solution has its roots in the *contact hypothesis* of Allport (1954). Facilitating contact formation and interaction between members of the rival groups might contribute to the dissemination of traitor incentives and consequently to the resolution of intergroup conflict. However, it is more difficult to find a rationale for actors to behave tolerantly and form such contacts voluntarily, especially during ongoing conflict. Unlike structural learning (cf. Section 6.4.5), *tolerance* is a farsighted network strategy that is based on short-term sacrifices in order to gain long-term benefits of intergroup peace. Subsequent research might examine under which conditions, if at all, strategic formation of intergroup ties would bring individual benefits in the long run. Another question concerns whether or not tolerant network strategies can invade a segregated population that is full of prejudice, and whether or not tolerant societies can be in evolutionary stable equilibrium.

This investigation might take place by using attractor neural networks (cf. Section 6.4.4). In this perspective, the strategy “tolerance” gives new weights to the strength of ties in favor of intergroup contacts. The success of the strategy might depend on the likelihood of continuation and on how quickly individuals can change their ego-networks. A key to success might be in the moderateness of the strategy. An exaggerated application of tolerance would simply lead to treachery and not to conflict resolution. Important fellow ties should not be sacrificed, as they are the sources of propagating peaceful behavior. Furthermore, the network location of tolerant actors might also be of crucial importance. For instance, individuals with high centrality in their group can provide sufficient confirmation pressure towards several fellows, but they are also at risk of being turned back easily.

6.5.3 Raising new issues: crosscutting cleavages and consociational democracies

The idea that crosscutting social circles promote good intergroup relations dates back to Simmel (1955[1908]). This theorem was elaborated further in several respects (LeVine and Campbell, 1972; Deschamps and Doise, 1978; Blau and Schwartz, 1984; Brewer and Miller, 1984; Flap, 1988). In empirical research, the hypothesis was tested, for instance, by using data on intermarriage (Blau and Schwartz, 1984). In political science, crosscutting dimensions have been recognized to overcome deeply rooted cleavages in Western democracies (cf. Lipset and Rokkan, 1967). However, there is also contradicting evidence, which shows that crosscutting social circles do not necessarily influence behavior, for instance, voting (Nieuwbeerta and Flap, 2000).

A subsequent study could analyze the effects of introducing a new issue that causes a similar negative interdependency as the old issue of intergroup competition, but along another division line. In the case of crosscutting cleavages, the dimensions would be clearly distinct. A further specification could concern the salience of actors for old and new issues, as is the case in collective decision making models (e.g., Bueno de Mesquita and Stokman, 1994). Conflict on the old issue might end, if the new issue is salient enough for actors in key structural positions to build intergroup ties. On the other hand, conflict might evolve on the new issue, if the old division becomes unimportant for most individuals and the population becomes segregated along the new division line. Even worse, raising a new issue might lead to conflict on both dimensions, even when cleavages are crosscutting. For instance, hooligans at one time may fight as football supporters and other times they may be side by side as left or right wing extremists.

Peaceful resolutions might emerge on both issues, for instance, if key actors in the rival groups develop a high salience for the new issue, on which they agree. In this case, the majority of the population does not necessarily have to be interested in the new issue nor should they decrease their negative sentiments towards members of the other group. It is sufficient if they remain interested in and influenced by the key actors in their group. The popularity of influential actors ensures that conflict is suppressed on the old dimension. This is probably why peaceful coexistence can be maintained in consociational democracies (cf. Lijphart, 1969; 1977). For instance, in Belgium, as long as the Flemish and the Walloon intellectual, political and economic elite remains united on major policy issues, there is a chance that ethnic relations can be rendered harmonious.

6.5.4 Assimilation and exit

As Simmel (1955[1908]) claimed, a natural end to conflict is the disappearance of its object. This refers to the case where a limited resource for which groups compete, cannot be attained anymore. Another quick way of conflict resolution is when one of the contestants disappears. This occurs where one group *dies out* completely, it *assimilates* to the other group, or *quits* the interdependence. Unfortunately, there are several empirical examples, in which participants of conflict aim at the complete *annihilation* of the other side. Ethnic cleansing and genocide are the most horrible ways of getting rid of the rivalry. *Assimilation* means giving up the original group membership. It might be a beneficial strategy, especially if traitor pressure puts a heavy burden on the individual who is surrounded by members of the other group. The possibility of changing sides under such conditions might be incorporated into an agent-based simulation design. Sometimes it is possible to avoid taking part in the competition or to *exit* the intergroup contest (cf. Hirschman, 1970). The exit option has already been incorporated into agent-based simulations for the two-person PD. Results show that enlarging the set of alternatives by the possibility of not playing might help the evolution of cooperation and morality (Schüssler, 1989; Vanberg and Congleton, 1992; Macy and Skvoretz, 1998). Individual exit does not obviously have such desirable consequences in the structurally embedded IPG game, as gatekeepers with extensive contacts to members of both groups would be especially keen on refusing to play.

6.5.5 Discriminative sanctioning

Another strategy that advocates conflict resolution and might be of central interest in subsequent research is *positive discrimination*. As we discussed in Section 6.3, the production of social control often involves costs. Consequently, individuals have limitations on how much social control they impose on each other (cf. Kuran, 1995). Costs require them to decide whom to sanction negatively or positively (cf. Flache 1996; Fehr and Schmidt 1999). This means relaxing the assumption that social control towards in-group and out-group members is distributed automatically and infinitely. In this perspective, a multitude of possible strategies can be constructed that are based on different distribution policies of sanctioning resources. Positive discrimination, as one of these, implies more likely imposing positive sanctions on out-group members and negative sanctions on group fellows. Similar to tolerance (cf. Section 6.5.2), it is difficult to find an individual rationale for the application of such a strategy. A significant proportion of individuals may follow it if a backup group specializes in higher order sanctioning and provides selective incentives for positive discriminators or punishments for negative discriminators. A further support could come from the inclusion of both forward-looking and backward-looking agents in the population. Far sighted individuals might teach reinforcement learners to adopt peaceful behavior by using discriminative sanctioning. Exact conditions for the emergence of equilibria, in which individuals in sufficient number apply positive discrimination to protect intergroup peace, can be highlighted by agent-based simulation.

6.5.6 Epilogue

In this chapter, we summarized and concluded our study on competitive intergroup relations. An overview of the main results was provided and some strengths and weaknesses of the analysis were discussed. We also supported further investigations with some suggestions. We concentrated mainly on recommendations for simulation research, but we strongly believe that analytical model developments, laboratory experiments, survey research, field and case studies might also reveal further underlying mechanisms and enrich the explanation of intergroup conflict substantially.

After all, this study should be considered as a modest contribution to the search for a solution for the costly enigma of intergroup conflict. Hopefully, this search will provide the opportunity to challenge the heavy pessimism delivered from the time of the Bosnian war, in yet another graffiti message in the heart of Budapest stating “Szarajövõ” (the future is crap).

References

Allison, P. (1992): The Cultural Evolution of Beneficent Norms. *Social Forces*, 71: 279-301.

Allport, G. W. (1954): *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Anderson, B. (1983): *Imagined Communities.* New York, Verso.

Andreoni, J. (1988): Why Free Ride? Strategies and Learning in Public Goods Experiments. *Journal of Public Economics*, 37: 291-304.

Andreoni, J. and Miller, J. H. (1993): Rational Cooperation in the Finitely Repeated Prisoner's Dilemma: Experimental Evidence. *Economic Journal*, 103: 570-585.

Asch, S. (1956): *Studies of Independence and Conformity*. Washington, DC: American Psychological Association.

Au, W. T.; Chen, X. P., and Komorita, S. S. (1998): A Probabilistic Model of Criticality in a Sequential Public Good Dilemma. *Organizational Behavior and Human Decision Processes*, 75(3): 274-293.

Axelrod, R. (1984): *The Evolution of Cooperation*. New York, Basic Books.

Axelrod, R. (1986): An Evolutionary Approach to Norms. *American Political Science Review,* 80: 1095-1111.

Axelrod, R. (1987): The Evolution of Strategies in the Iterated Prisoner's Dilemma. In: Davis, L. (ed.): *Genetic Algorithms and Simulated Annealing*. London, Pitman.

Axelrod, R. (1997a): *The Complexity of Cooperation. Agent-Based Models of Competition and Collaboration*. Princeton NJ, Princeton University Press.

Axelrod, R. (1997b): The Dissemination of Culture. A Model with Local Convergence and Global Polarization. *Journal of Conflict Resolution,* 41: 203-226.

Bainbridge, W. S. (1995): Neural Network Models of Religious Belief. *Sociological Perspectives,* 38(4): 483-495.

Bateson, P. (2000): The Biological Evolution of Cooperation and Trust. In: Gambetta, D. (ed.): *Trust: Making and Breaking Cooperative Relations.* Department of Sociology, University of Oxford. http://www.sociology.ox.ac.uk/papers/bateson14-30.pdf

Berninghaus, S. K. and Schwalbe, U. (1996): Evolution, Interaction, and Nash Equilibria. *Journal of Economic Behavior and Organization,* 29: 57-85.

Bethwaite, J. and Tompkinson, P. (1996): The Ultimatum Game and Non-Selfish Utility Functions. *Journal of Economic Psychology*, 17(2): 259-271.

Bettencourt, B. A.; Brewer, M. B.; Croak, M. R., and Miller, N. (1992): Cooperation and the Reduction of Intergroup Bias: The Role of Reward Structure and Social Orientation. *Journal of Experimental Social Psychology,* 28: 301-319.

Bhavnani, R. and Backer, D. (2000): Localized Ethnic Conflict and Genocide. Accounting for Differences in Rwanda and Burundi. *Journal of Conflict Resolution*, 44(3): 283-306.

Bicchieri, C. (1993): *Rationality and Coordination*. Cambridge, Cambridge University Press.

Bicchieri, C. (1997): Learning to Cooperate. In: Bicchieri, C.; Jeffrey, R., and Skyrms, B. (eds.): *The Dynamics of Norms*. (Studies in Probability, Induction, and Decision Theory.) Cambridge, Cambridge University Press.

Bienenstock, E. J. (2001): *The Evolution of Reputation and Cooperation in Emergent Social Networks*. Manuscript.

Billig, M. G. and Tajfel, H. (1973): Social Categorization and Similarity in Intergroup Behaviour. *European Journal of Social Psychology,* 3: 27-52.

Binmore, K. (1998): The Evolution of Fairness Norms. *Rationality and Society,* 10(3): 275-301.

Blalock, H. M. (1967): *Toward a Theory of Minority-Group Relations.* New York, Wiley.

Blalock, H. M. (1986): A Model for Racial Contact in Schools. In: Prager, J.; Longshore, D., and Seeman, M. (eds.): *School Desegregation Research. New Directions in Situational Analysis*. New York, Plenum Press.

Blau, P. M. and Schwartz, J. E. (1984): *Crosscutting Social Circles. Testing a Macrostructural Theory of Intergroup Relations.* Orlando, Academic Press.

Blumer, H. (1958): Race Prejudice as a Sense of Group Position. *Pacific Sociological Review,* 1(1): 3-7.

Bohnet, I. and Frey, B. S. (1999): The Sound of Silence in Prisoner's Dilemma and Dictator Games. *Journal of Economic Behavior and Organization*, 38: 43-57.

Bonacich, P. and Schneider, S. (1992): Communication Networks and Collective Action. In: Liebrand, W. B. G.; Messick, D. M., and Wilke, H. A. M. (eds.): *Social Dilemmas. Theoretical Issues and Research Findings*. Oxford, Pergamon Press.

Bornstein, G. (1992): The Free-Rider Problem in Intergroup Conflicts over Step-Level and Continuous Public Goods. *Journal of Personality and Social Psychology*, 62: 597-606.

Bornstein, G. and Ben-Yossef, M. (1994): Cooperation in Intergroup and Single-Group Social Dilemmas. *Journal of Experimental Social Psychology*, 30: 52-67.

Bornstein, G. and Rapoport, Amnon (1988): Intergroup Competition for the Provision of Step-Level Public Goods: Effects of Preplay Communication. *European Journal of Social Psychology*, 18: 125-142.

Bornstein, G.; Erev, I., and Rosen, O. (1990): Intergroup Competition as a Structural Solution to Social Dilemmas. *Social Behavior*, 5: 247-260.

Bornstein, G.; Winter, E., and Goren, H. (1996): Experimental Study of Repeated Team-Games. *European Journal of Political Economy*, 12: 629-639.

Bowles, S. and Gintis, H. (1998): The Moral Economy of Communities: Structured Populations and the Evolution of Pro-Social Norms. *Evolution and Human Behavior*, 19: 3-25.

Brewer, M. B. (1981): Ethnocentrism and Its Role in Interpersonal Trust. In: Brewer, M. B. and Collins, B. (eds.): *Scientific Inquiry and the Social Sciences*. San Francisco, Jossey-Bass.

Brewer, M. B. (1996a): When Contact is not Enough: Social Identity and Intergroup Cooperation. *International Journal of Intercultural Relations,* 20: 291-303.

Brewer, M. B. (1996b): In-Group Favoritism: The Subtle Side of Intergroup Discrimination. In: Messick, D. M. and Tenbrunsel, A. E. (eds.): *Codes of Conduct. Behavioral Research Into Business Ethics.* New York, Russell Sage Foundation.

Brewer, M. B. (1999): The Psychology of Prejudice: Ingroup Love or Outgroup Hate? *Journal of Social Issues,* 55(3): 429-444.

Brewer, M. B. and Miller, N. (1984): Beyond the Contact Hypothesis: Theoretical Perspectives on Desegregation. In: Miller, N. and Brewer, M. B. (eds.): *Groups in Contact: The Psychology of Desegregation*. Orlando, Academic Press.

Brewer, M. B. and Miller, N. (1996): *Intergroup Relations.* Buckingham, Open University Press.

Brown, G. (1951): Iterative Solution of Games by Fictitious Play. In: *Activity Analysis of Production and Allocation*. New York, Wiley.

Brown-Kruse, J. and Hummels, D. (1993): Gender Effects in Laboratory Public Goods Contribution: Do Individuals Put Their Money Where Their Mouth Is? *Journal of Economic Behavior and Organization*, 22: 255-267.

Bryk, A.S. and Raudenbush, S. W. (1992): *Hierarchical Linear Models. Applications and Data Analysis Methods*. London, Sage.

Budescu, D. V.; Rapoport, Amnon, and Suleiman, R. (1990): Resource Dilemmas with Environmental Uncertainty and Asymmetric Players. *European Journal of Social Psychology*, 20: 475-487.

Bueno de Mesquita, B. and Stokman, F. N. (eds.) (1994): *European Community Decision Making. Models, Applications, and Comparisons.* New Haven, Yale University Press.

Burt, R. S. (1992): *Structural Holes: The Social Structure of Competition*. Cambridge (Mass.), Harvard University Press.

Burt, R. S. (1998): The Gender of Social Capital. *Rationality and Society*, 10(3): 5-46.

Bush, R.R. and Mosteller, F. (1955): *Stochastic Models for Learning*. New York, Wiley.

Buskens, V. (1999): *Social Networks and Trust*. Amsterdam, Thesis Publishers.

Buskens, V. and Snijders, C. (1997): Individual Heuristics and the Dynamics of Cooperation in Large Groups: Additional Results Using Analytical Methods. *Psychological Review*, 104(4): 792-800.

Buskens, V. and Weesie, J. (2000): An Experiment on the Effects of Embeddedness in Trust Situations: Buying a Used Car. *Rationality and Society,* 12: 227-253.

Butler, D. and Stokes, D. (1974): *Political Change in Britain: The Evolution of Electoral Choice*. New York, St. Martin's.

Cadsby, C. B. and Maynes, E. (1998): Gender and Free Riding in a Threshold Public Goods Game: Experimental Evidence. *Journal of Economic Behavior and Organization,* 34: 603-620.

Camerer, C. F. (1997): Progress in Behavioral Game Theory. *Journal of Economic Perspectives*, 11: 167-188.

Camerer, C. F. and Ho, T-H. (1999): Experience-Weighted Attraction Learning in Games: Estimates from Weak-Link Games. In: Budescu, D., V.; Erev, I., and Zwick, R. (eds.): *Games and Human Behavior. Essays in Honor of Amnon Rapoport*. Mahwah, NJ, Lawrence Erlbaum Publishers.

Caporael, L. R.; Dawes, R. M.; Orbell, J. M., and van de Kragt, A. J. C. (1989): Selfishness Examined: Cooperation in the Absence of Egoistic Incentives; and Response. *Behavioral and Brain Sciences*, 12: 683-739.

Carley, K. (1991): A Theory of Group Stability. *American Sociological Review,* 56: 331-354.

Carter, J. and Irons, M. (1991): Are Economists Different, and If So, Why? *Journal of Economic Perspectives*, 5: 171-177.

Chen, X. P.; Au, W. T., and Komorita, S. S. (1996): Sequential Choice in a Step-Level Public Goods Dilemma: The Effects of Criticality and Uncertainty. *Organizational Behavior and Human Decision Processes*, 69: 179-193.

Chen, Y. and Tang, F. F. (1998): Learning and Incentive Compatible Mechanisms for Public Goods Provision. *Journal of Political Economy*, 106: 633-662.

Chong, D. (1991): *Collective Action and the Civil Rights Movement*. Chicago, Chicago University Press.

Chwe, M. S-Y. (1999): Structure and Strategy in Collective Action. *American Journal of Sociology*, 105: 128-156.

Chwe, M. S-Y. (2000): Communication and Coordination in Social Networks. *Review of Economic Studies,* 67: 1-16.

Cialdini, R. B.; Schaller, M.; Houlihan, D.; Arps, K.; Fultz, J., and Beaman, A. L. (1987): Empathy-based Helping: Is it Selflessly or Selfishly Motivated? *Journal of Personality and Social Psychology*, 52: 749-758.

Coenders, M. (2001): *Nationalistic Attitudes and Ethnic Exclusionism in a Comparatice Perspective. An Empirical Study of Attitudes Toward the Country and Ethnic Immigrants in 22 Countries*. Nijmegen, ICS Dissertation Series.

Cohen, J. (1990): Things That I Have Learned (So Far). *American Psychologist*, 45: 1304-1312.

Coleman, J. S. (1990): *The Foundations of Social Theory*. Cambridge (Mass.), The Belknap Press.

Coser, L. A. (1968) [1956]: *The Functions of Social Conflict*. London, Routledge & Kegan Paul.

Coser, L. A. (1967): *Continuities in the Study of Social Conflict*. New York, Free Press.

Cosmides, L. and Tooby, J. (1992): Cognitive Adaptations for Social Exchange. In: Barkow, J. H.; Cosmides, L., and Tooby, J. (eds.): *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York [etc.], Oxford University Press.

Crawford, V. P. (1997): Theory and Experiment in the Analysis of Strategic Interaction. In: Kreps, D. M. and Wallis, K. F. (eds.): *Advances in Economics and Econometrics: Theory and Applications, 7th World Congress. Vol. 1.* Cambridge, Cambridge University Press.

Dawes, R. M. (1980): Social Dilemmas. *Annual Review of Psychology,* 31: 169-193.

Dawes, R. M. and Thaler, R. H. (1988): Anomalies. Cooperation. *Journal of Economic Perspectives*, 2: 187-197.

Dawkins, R. (1976): *The Selfish Gene*. Oxford, Oxford University Press.

Denitch, B. (1994): *Ethnic Nationalism: The Tragic Death of Yugoslavia*. Minneapolis, University of Minnesota Press.

Deschamps, J. C. and Doise, W. (1978): Crossed Category Membership in Intergroup Relations. In: Tajfel, H. (ed.): *Differentiation between Social Groups.* London, Academic Press.

de Vos, H. and Zeggelink, E. (1997): Reciprocal Altruism in Human Social Evolution: The Viability of Reciprocal Altruism With a Preference for “Old-Helping-Partners”. *Evolution and Human Behavior*, 18: 261-278.

Diez Medrano, J. (1994): The Effects of Ethnic Segregation and Ethnic Competition on Political Mobilization in the Basque Country, 1988. *American Sociological Review*, 59(6): 873-889.

Doherty, P. (1990): Residential Segregation, Social Isolation and Integrated Education in Belfast. *New Community*, 16: 391-401.

Doi, T. (1994): Social Orientation Analysis of the Common and Individual Interest Problem. In: Schulz, U. et al. (eds.): *Social Dilemmas and Cooperation*. Berlin, Springer Verlag.

Downs, A. (1957): *An Economic Theory of Democracy.* New York, Harper and Row.

Drolet, A. L. and Morris, M. W. (2000): Rapport in Conflict Resolution: Accounting for How Face-to-Face Contact Fosters Mutual Cooperation in Mixed-Motive Conflicts. *Journal of Experimental Social Psychology,* 36(1): 26-50.

Duncan, O. D. and Duncan, B. (1955): A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20: 210-217.

Durkheim, E. (1984) [1893]: *Division of Labor in Society.* London, MacMillan.

Ellison, G. (1993): Learning, Local Interaction, and Coordination. *Econometrica*, 61: 1047-1071.

Elster, J. (1990): Selfishness and Altruism. In: Mansbridge, J. J. (ed.): *Beyond Self-Interest*. Chicago, University of Chicago Press.

Epstein, J. M. and Axtell, R. (1996): *Growing Artificial Societies: Social Science from the Bottom Up.* Washington DC, Brookings Institution Press.

Erev, I. and Rapoport, Amnon (1990): Provision of Step-Level Public Goods: The Sequential Contribution Mechanism. *Journal of Conflict Resolution*, 34(3): 401-425.

Erev, I. and Roth, A. E. (1998): Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *American Economic Review*, 88(4): 848-881.

Erev, I. and Roth, A. E. (1999): On the Role of Reinforcement Learning in Experimental Games: The Cognitive Game-Theoretic Approach. In: Budescu, D., V.; Erev, I., and Zwick, R. (eds.): *Games and Human Behavior. Essays in Honor of Amnon Rapoport*. Mahwah, NJ, Lawrence Erlbaum Publishers.

Erev, I.; Bornstein, G., and Galili, R. (1993): Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment. *Journal of Experimental Social Psychology*, 29: 463-478.

Esser, H. (1986): Ethnic Segmentation as the Unintended Result of Intentional Action. In: Diekmann, A. and Mitter, P. (eds.): *Paradoxical Effects of Social Behavior. Essays in Honor of Anatol Rapoport*. Heidelberg, Physica.

Farquhar, P. H. (1984): Utility Assessment Methods. *Management Science*, 30: 1283-1300.

Fearon, J. and Laitin, D. (1996): Explaining Interethnic Cooperation. *American Political Science Review*, 90(4): 715-735.

Fehr, E. and Schmidt, K. (1999): A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics,* 114: 817-868.

Fernandez, R. M. and McAdam, D. (1988): Social Networks and Social Movements: Multiorganizational Fields and Recruitment to Freedom Summer. *Sociological Forum*, 3: 357-382.

Festinger, L. (1962): *A Theory of Cognitive Dissonance*. Stanford, Stanford University Press.

Finkel, S. E. and Opp, K-D. (1991): Party Identification and Participation in Collective Political Action. *Journal of Politics*, 53(2): 339-371.

Finkel, S. E.; Muller, E. N., and Opp, K.-D. (1989): Personal Influence, Collective Rationality, and Mass Political Action. *American Political Science Review,* 83: 885-903.

Fischer, I. and Suleiman, R. (1997): Election Frequency and the Emergence of Cooperation in a Simulated Intergroup Conflict. *Journal of Conflict Resolution,* 41: 483-508.

Flache, A. (1996): *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas*. Amsterdam, Thesis Publishers.

Flache, A. (2002): The Rational Weakness of Strong Ties. Failure of Group Solidarity in a Highly Cohesive Group of Rational Agents. *Journal of Mathematical Sociology*, forthcoming.

Flache, A. and Hegselmann, R. (1999a): Do Irregular Grids Make a Difference? Relaxing the Spatial Regularity Assumption in Cellular Models of Social Dynamics. *Journal of Artificial Societies and Social Simulation*, 4(4), http://www.soc.surrey.ac.uk/JASSS/4/4/6.html

Flache, A. and Hegselmann, R. (1999b): Rationality vs. Learning in the Evolution of Solidarity Networks: A Theoretical Comparison. *Computational and Mathematical Organization Theory*, 5: 97-127.

Flache, A. and Macy, M. W. (1996): The Weakness of Strong Ties: Collective Action Failure in a Highly Cohesive Group. *Journal of Mathematical Sociology,* 21: 3-28.

Flap, H. D. (1988): *Conflict, Loyalty and Violence. The Effects of Social Networks on Behaviour*. Frankfurt, Verlag Peter Lang.

Frank, R. H. (1988): *Passions Within Reason: The Strategic Role of Emotions.* New York, Norton.

Frank, R. H.; Gilovich, T., and Regan, D. (1993): Does Studying Economics Inhibit Cooperation? *Journal of Economic Perspectives*, 7(2): 159-171.

Friedkin, N. E. and Johnsen, E. C. (1999): Social Influence Networks and Opinion Change. *Advances in Group Processes,* 16: 1-29.

Friedman, J. W. (1971): A Non-cooperative Equilibrium for Supergames. *Review of Economic Studies,* 38: 1-12.

Friedman, J. W. (1986): *Game Theory with Applications to Economics.* New York, Oxford University Press.

Frohlich, N. and Oppenheimer, J. A. (1978): *Modern Political Economy*. Englewood Cliffs (NJ), Prentice Hall.

Fudenberg, D. and Kreps, D. M. (1993): Learning Mixed Equilibria. *Games and Economic Behavior*, 5: 320-367.

Fudenberg, D. and Levine, K. (1998): *The Theory of Learning in Games*. Cambridge (Mass.), MIT Press.

Gächter, S. and Fehr, E. (1999): Collective Action as a Social Exchange. *Journal of Economic Behavior and Organization*, 39: 341-369.

Gellner, E. (1983): *Nations and Nationalism*. Ithaca, NY, Cornell University Press.

Gilbert, N. (1995): *Simulation: An Emergent Perspective*. Paper presented at the Conference on New Technologies in the Social Sciences, 27-29 October, Bournemouth, UK. http://www.soc.surrey.ac.uk/research/simsoc/tutorial.html

Gintis, H. (2001): *The Puzzle of Prosociality*. Santa Fe Institute Working Paper Series 01-10-059. http://www.santafe.edu/sfi/publications/Abstracts/01-10-059abs.html

Goldstein, H. (1995). *Multilevel Statistical Models*. London, Edward Arnold Publishers.

Goren, H. and Bornstein, G. (1999): Reciprocation and Learning in the Intergroup Prisoner's Dilemma Game. In: Budescu, D., V.; Erev, I., and Zwick, R. (eds.): *Games and Human Behavior. Essays in Honor of Amnon Rapoport*. Mahwah, NJ, Lawrence Erlbaum Publishers.

Goren, H. and Bornstein, G. (2000): The Effects of Intragroup Communication on Intergroup Cooperation in the Repeated Intergroup Prisoner's Dilemma (IPD) Game. *Journal of Conflict Resolution*, 44(5): 700-719.

Gould, R. V. (1991): Multiple Networks and Mobilization in the Paris Commune, 1871. *American Sociological Review*, 56: 716-729.

Gould, R. V. (1993a): Collective Action and Network Structure. *American Sociological Review,* 58: 182-196.

Gould, R. V. (1993b): Trade Cohesion, Class Unity, and Urban Insurrection: Artisanal Activism in the Paris Commune. *American Journal of Sociology,* 98: 721-754.

Gould, R. V. (1995): *Insurgent Identities: Class, Community, and Protest in Paris from 1848 to the Commune*. Chicago, University of Chicago Press.

Gould, R. V. (1999): Collective Violence and Group Solidarity: Evidence from a Feuding Society. *American Sociological Review*, 64: 356-380.

Grannis, R. (1998): The Importance of Trivial Streets: Residential Streets and Residential Segregation. *American Journal of Sociology*, 103: 1530-1564.

Granovetter, M. (1973): The Strength of Weak Ties. *American Journal of Sociology*, 78: 1360-1380.

Granovetter, M. (1978): Threshold Models of Collective Behavior. *American Journal of Sociology*, 83: 1420-1443.

Granovetter, M. (1985): Economic Action and Social Structure: The Problem of Embeddedness. *American Journal of Sociology*, 91: 481-510.

Granovetter, M. (1986): The Micro-Structure of School Desegregation. In: Prager, J.; Longshore, D., and Seeman, M. (eds.): *School Desegregation Research. New Directions in Situational Analysis*. New York, Plenum Press.

Guttman, J. M. (1986): Matching Behavior and Collective Action: Some Experimental Evidence. *Journal of Economic Behavior and Organization*, 7: 171-198.

Hardin, R. (1982): *Collective Action*. Baltimore, Johns Hopkins University Press.

Hardin, R. (1995): *One for All. The Logic of Group Conflict*. Princeton NJ, Princeton University Press.

Harris, R. (1979): Community Relationships in Northern and Southern Ireland: A Comparison and a Paradox. *Sociological Review*, 27: 41-53.

Hasson, S. (1996): Local Politics and Split Citizenship in Jerusalem. *International Journal of Urban and Regional Research*, 20: 116-133.

Heckathorn, D. D. (1989): Collective Action and the Second-Order Free-Rider Problem. *Rationality and Society*, 1: 78-100.

Heckathorn, D. D. (1990): Collective Sanctions and Compliance Norms: A Formal Theory of Group-Mediated Social Control. *American Sociological Review*, 55: 366-84.

Heckathorn, D. (1993): Collective Action and Group Heterogeneity: Voluntary Provision versus Selective Incentives. *American Sociological Review*, 58: 329-350.

Hegselmann, R. (1996): Social Dilemmas in Lineland and Flatland. In: Liebrand, W. B. G. and Messick, D. M. (eds.): *Frontiers in Social Dilemma Research*. Berlin, Springer.

Hegselmann, R. and Flache, A. (1998): Understanding Complex Social Dynamics: A Plea For Cellular Automata Based Modelling. *Journal of Artificial Societies and Social Simulation,* 1(3), http://www.soc.surrey.ac.uk/JASSS/1/3/1.html

Hirschman, A. O. (1970): *Exit, Voice, Loyalty. Responses to Decline in Firms, Organizations, and States.* Cambridge (Mass.), Harvard University Press.

Ho, T. H. and Weigelt, K. (1996): Task Complexity, Equilibrium Selection, and Learning: An Experimental Study. *Management Science*, 42: 659-679.

Hogg, M. A. and Abrams, D. (1988): *Social Identification.* London, Routledge.

Holland, J. H. (1975): *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor (MI), University of Michigan Press.

Holland, J. H. and Miller, J. H. (1991): Artificial Adaptive Agents in Economic Theory. *American Economic Review*, 81: 365-370.

Hong, C. S. and Konrad, K. A. (1998): Bandwagon Effects and Two-party Majority Voting. *Journal of Risk and Uncertainty,* 16(2): 165-172.

Horwitz, M. and Rabbie, J. M. (1982): Individuality and Membership in the Intergroup System. In: Tajfel, H. (ed.): *Social Identity and Intergroup Relations*. Cambridge, Cambridge University Press.

Insko, C. A.; Schopler, J.; Drigotas, S. M.; Graetz, K. A.; Kennedy, J.; Cox, Ch., and Bornstein, G. (1993): The Role of Communication in Interindividual-Intergroup Discontinuity. *Journal of Conflict Resolution*, 37: 108-138.

Insko, C. A.; Schopler, J.; Graetz, K. A.; Drigotas, S. M.; Currey, D. P.; Smith, S. L.; Brazil, D., and Bornstein, G. (1994): Interindividual-Intergroup Discontinuity in the Prisoner's Dilemma Game. *Journal of Conflict Resolution*, 38: 87-116.

Isaac, R. M., McCue, K. F., and Plott, C. (1985): Public Goods Provision in an Experimental Environment. *Journal of Public Economics*, 26: 51-74.

Jager, W.; Popping, R., and van de Sande, H. (2001): Clustering and Fighting in Two-Party Crowds: Simulating the Approach-avoidance Conflict. *Journal of Artificial Societies and Social Simulation,* 4(3), http://www.soc.surrey.ac.uk/JASSS/4/3/7.html

Jankowski, M. S. (1991): *Islands in the Street. Gangs and American Urban Society*. Berkeley, University of California Press.

Kaplan, R. D. (1993): *Balkan Ghosts*. New York, Vintage.

Kenrick, D. T.; Baumann, D. J., and Cialdini, R. B. (1979): A Step in the Socialization of Altruism and Hedonism: Effects of Negative Mood on Children's Generosity under Public and Private Conditions. *Journal of Personality and Social Psychology*, 37: 747-755.

Kerr, N. L. (1989): Illusions of Efficacy: The Effects of Group Size on Perceived Efficacy in Social Dilemmas. *Journal of Experimental Social Psychology*, 25: 287-313.

Kitts, J. A.; Macy, M. W., and Flache, A. (1999): Structural Learning: Attraction and Conformity in Task-oriented Groups. *Computational and Mathematical Organization Theory*, 5(2): 129-145.

Klos, T. B. (1997): Spatially Coelvolving Automata Play the Repeated Prisoner's Dilemma. In: Conte, R. et al. (eds.): *Simulating Social Phenomena*. Berlin, Springer.

Knight, G. P. and Dubro, A. F. (1984): Cooperative, Competitive, and Individualistic Social Values: An Individualized Regression and Clustering Approach. *Journal of Personality and Social Psychology*, 46: 98-105.

Kollock, P. (1993): "An Eye for an Eye Leaves Everyone Blind": Cooperation and Accounting Systems. *American Sociological Review*, 58: 768-786.

Kollock, P. (1998): Social Dilemmas: The Anatomy of Cooperation. *Annual Review of Sociology*, 24: 183-214.

Komorita, S. S. and Parks, C. D. (1999): Reciprocity and Cooperation in Social Dilemmas: Review and Future Directions. In: Budescu, D., V.; Erev, I., and Zwick, R. (eds.): *Games and Human Behavior. Essays in Honor of Amnon Rapoport*. Mahwah, NJ, Lawrence Erlbaum Publishers.

Komorita, S. S.; Hilty, J. A., and Parks, C. G. (1991): Reciprocity and Cooperation in Social Dilemmas. *Journal of Conflict Resolution*, 35: 494-518.

Komorita, S. S.; Parks, C. G., and Hulbert, L. G. (1992): Reciprocity and the Induction of Cooperation in Social Dilemmas. *Journal of Personality and Social Psychology*, 62: 607-617.

Kosterman, R. and Feshbach, S. (1989): Toward a Measure of Patriotic and Nationalistic Attitudes. *Political Psychology*, 10: 257-274.

Kramer, R. M. and Brewer, M. B. (1984): Effects of Group Identity on Resource Use in a Simulated Common Dilemma. *Journal of Personality and Social Psychology*, 46: 1044-1057.

Kramer, R. M. and Brewer, M. B. (1986): Social Group Identity and the Emergence of Cooperation in Resource Dilemmas. In: Wilke, H.; Messick, M., and Rutte, C. (eds.): *Experimental Social Dilemmas*. Frankfurt, Verlag Peter Lang.

Kratzer, J. (2001): *Communication and Performance: An Empirical Study in Innovation Teams*. Amsterdam, Thesis Publishers.

Kreps, D. M.; Milgrom, P.; Roberts, J., and Wilson, R. (1982): Rational Cooperation in Finitely Repeated Prisoner's Dilemma. *Journal of Economic Theory*, 27: 245-252.

Kuhlman, D. M.; Camac, C., and Cunha, D. A. (1986): Individual Differences in Social Orientation. In: Wilke, H.; Messick, M., and Rutte, C. (eds.): *Experimental Social Dilemmas*. Frankfurt, Verlag Peter Lang.

Kuran, T. (1995): *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge (Mass.), Harvard University Press.

Latané, B. (1981): The Psychology of Social Impact. *American Psychologist*, 36: 343-356.

Ledyard, J. O. (1995): Public Goods: A Survey of Experimental Research. In: Kagel, J. H. and Roth, A. E. (eds.): *The Handbook of Experimental Economics*. Princeton, Princeton University Press.

Leenders, R.; Kratzer, J., and van Engelen, J. M. L. (2002): Managing Product Development Teams Effectively. In: *The PDMA Toolbook*. New York, Wiley.

LeVine, R. A. and Campbell, D. T. (1972): *Ethnocentrism: Theories of Conflict, Ethnic Attitudes and Group Behaviour.* New York, Wiley.

Lieberson, S. (1963): *Ethnic Patterns in American Cities*. New York, Free Press.

Lieberson, S. (1980): *A Piece of the Pie: Blacks and White Immigrants since 1880.* Berkeley, University of California Press.

Lieberson, S. and Carter, D. K. (1982): Temporal Changes and Urban Differences in Residential Segregation: A Reconsideration. *American Journal of Sociology*, 88(2): 296-310.

Liebrand, W. B. G. (1983): A Classification of Social Dilemma Games. *Simulation and Games*, 14: 123-138.

Liebrand, W. B. G. (1984): The Effect of Social Motives, Communication and Group Size on Behavior in an N-Person Multi-Stage Mixed-Motive Game. *European Journal of Social Psychology*, 14: 239-264.

Liebrand, W. B. G. (1986): The Ubiquity of Social Values in Social Dilemmas. In: Wilke, H.; Messick, M., and Rutte, C. (eds.): *Experimental Social Dilemmas*. Frankfurt, Verlag Peter Lang.

Lijphart, A. (1969): Consociational Democracy. *World Politics*, 21(2): 207-225.

Lijphart, A. (1977): *Democracy in Plural Societies. A Comparative Exploration*. New Haven, Yale University Press.

Lindenberg, S. (1986): The Paradox of Privatization in Consumption. In: Diekmann, A. and Mitter, P. (eds.): *Paradoxical Effects of Social Behavior. Essays in Honor of Anatol Rapoport*. Heidelberg, Physica.

Lindenberg, S. (1992): The Method of Decreasing Abstraction. In: Coleman, J. S. and Fararo, T. J. (eds.): *Rational Choice Theory. Advocacy and Critique*. Newbury Park, Sage.

Lindenberg, S. (1993): Framing, Empirical Evidence, and Applications. In: *Jahrbuch für Neue Politische Ökonomie*. Tübingen, Mohr (Siebeck).

Lindenberg, S. (1997): Grounding Groups in Theory: Functional, Cognitive, and Structural Interdependencies. *Advances in Group Processes,* 14: 281-331.

Lindenberg, S., and Frey, B. (1993): Alternatives, Frames, and Relative Prices: A Broader View of Rational Choice Theory. *Acta Sociologica*, 36:191-205.

Linster, B. (1992): Evolutionary Stability in the Repeated Prisoner's Dilemma Played by Two-State Moore Machines. *Southern Economic Journal*, 58: 880-903.

Lipset, S. M. and Rokkan, S. (1967): Cleavage Structure, Party Systems, and Voter Alignments: An Introduction. In: Lipset, S. M. and Rokkan, S. (eds.): *Party Systems and Voter Alignments. Cross-National Perspectives.* New York, Free Press.

Liska, A. E. and Bellair, P. E. (1995): Violent-Crime Rates and Racial Composition: Convergence over Time. *American Journal of Sociology*, 101: 578-610.

Liska, A. E.; Logan, J. R., and Bellair, P. E. (1998): Race and Violent Crime in the Suburbs. *American Sociological Review*, 63: 27-38.

Luce, D. and Raiffa, H. (1957): *Games and Decisions.* New York, Wiley.

Macy, M. W. (1989): Walking Out of Social Traps. A Stochastic Learning Model for the Prisoner's Dilemma. *Rationality and Society*, 2: 197-219.

Macy, M. W. (1990): Learning Theory and the Critical Mass. *American Sociological Review*, 55: 809-826.

Macy, M. W. (1991a): Chains of Cooperation: Threshold Effects in Collective Action. *American Sociological Review*, 56: 730-747.

Macy, M. W. (1991b): Learning to Cooperate: Stochastic and Tacit Collusion in Social Exchange. *American Journal of Sociology*, 97: 808-843.

Macy, M. W. (1993a): Backward Looking Social Control. American Sociological Review, 58: 819-836.

Macy, M. W. (1993b): Social Learning and the Structure of Collective Action. In: Lawler, E. et al. (eds.): *Advances in Group Processes*, 10: 1-35. Greenwich, JAI Press.

Macy, M. W. (1995): PAVLOV and the Evolution of Cooperation: An Experimental Test. *Social Psychology Quarterly*, 58(2): 74-87.

Macy, M. W. (1996): Natural Selection and Social Learning in Prisoner's Dilemma. Coadaption with Genetic Algorithms and Artificial Networks. *Sociological Methods and Research*, 25: 103-137.

Macy, M. W. (1997): Identity, Interest and Emergent Rationality: An Evolutionary Synthesis. *Rationality and Society*, 9(4): 427-448.

Macy, M. W. (2002): Social Simulation. In: Smelser, N. and Baltes, P. (eds.): *International Encyclopaedia of the Social and Behavioral Sciences.* Amsterdam, Elsevier.

Macy, M. W. and Flache, A. (2002): Learning Dynamics in Social Dilemmas. *Proceedings of the National Academy of Sciences*, April.

Macy, M. W. and Skvoretz, J. (1998): The Evolution of Trust and Cooperation between Strangers: A Computational Model. *American Sociological Review,* 63: 638-660.

Mailath, G.; Samuelson, L., and Shaked, A. (1997): Correlated Equilibria and Local Interactions. *Economic Theory,* 9: 551-556.

Marwell, G. and Ames, R. (1981): Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods, IV. *Journal of Public Economics*, 15: 295-310.

Marwell, G. and Oliver, P. E. (1993): *Critical Mass in Collective Action*. Cambridge, Cambridge University Press.

Marwell, G.; Oliver, P. E., and Prahl, R. (1988): Social Networks and Collective Action: A Theory of the Critical Mass. III. *American Journal of Sociology*, 94: 502-534.

Mason, C. F.; Phillips, O. R., and Redington, D. B. (1991): On the Role of Gender in a Noncooperative Game. *Journal of Economic Behavior and Organization*, 15: 215-235.

Massey, G.; Hodson, R., and Sekulic, D. (1999): Ethnic Enclaves and Intolerance: The Case of Yugoslavia. *Social Forces,* 78(2): 669-673.

Maynard Smith, J. and Price, G. R. (1973): The Logic of Animal Conflicts. *Nature*, 246: 15-18.

Maynard Smith, J. (1982): *Evolution and the Theory of Games*. Cambridge, Cambridge University Press.

McAdam, D. (1986): Recruitment to High-Risk Activism: The Case of Freedom Summer. *American Journal of Sociology*, 92: 64-90

McAdam, D. and Paulsen, R. (1993): Specifying the Relationship between Social Ties and Activism. *American Journal of Sociology*, 99: 640-667

McClintock, C. G. and Liebrand, W. B. G. (1988): The Role of Interdependence Structure, Individual Value Orientation, and Another's Strategy in Social Decision Making: A Transformational Analysis. *Journal of Personality and Social Psychology*, 55: 396-409.

Messick, D. M. and Liebrand, W. B. G. (1995): Individual Heuristics and the Dynamics of Cooperation in Large Groups. *Psychological Review*, 102: 131-145.

Minsky, M. (1985): *The Society of Mind.* New York, Simon and Schuster.

Morris, S. (2000): Contagion. *Review of Economic Studies*, 67: 57-78.

Nelson, R. E. (1989): The Strength of Strong Ties: Social Networks and Intergroup Conflict in Organizations. *Academy of Management Journal*, 32: 377-401.

Nieuwbeerta, P. and Flap, H. (2000): Crosscutting Social Circles and Political Choice: Effects of Personal Network Composition on Voting Behavior in the Netherlands. *Social-Networks*, 22(4): 313-335.

Newcomb, T. M. (1947): Autistic Hostility and Social Reality. *Human Relations*, 1: 69-86.

Nowak, A. and Vallacher, R. (1998): Computational Social Psychology: Cellular Automata and Neural Network Models of Interpersonal Dynamics. In: Read, S. and Miller, L. (eds.): *Connectionist Models of Social Reasoning and Social Behavior.* Mahwah, NJ, Lawrence Erlbaum Associates.

Nowak, A.; Szamrej, J., and Latané, B. (1990): From Private Attitude to Public Opinion: A Dynamic Theory of Social Impact. *Psychological Review,* 97: 362-376.

Nowak, M. and May, R. (1992): Evolutionary Games and Spatial Chaos. *Nature,* 359 (29 Oct): 826-829.

Nowak, M. and Sigmund, K. (1992): Tit for Tat in Heterogeneous Populations. *Nature*, 355: 16 January, 250-253.

Nowak, M. and Sigmund, K. (1993): A Strategy of Win-Stay, Lose-Shift that Outperforms Tit-for-tat in the Prisoner's Dilemma Game. *Nature*, 364: 1 July, 56-58.

Nowak, M. and Sigmund, K. (1998): Evolution of Indirect Reciprocity by Image Scoring. *Nature,* 393 (11 June): 573-576.

Nowell, C. and Tinkler, S. (1994): The Influence of Gender on the Provision of a Public Good. *Journal of Economic Behavior and Organization*, 25(1): 25-36.

Oberschall, A. R. (1973): *Social Conflict and Social Movements*. Englewood Cliffs (NJ), Prentice-Hall.

Oberschall, A. R. (1994): Rational Choice in Collective Protests. *Rationality and Society*, 6: 79-100.

Oliver, P.; Marwell, G., and Teixeira, R. (1985): A Theory of the Critical Mass. I. Interdependence, Group Heterogeneity, and the Production of Collective Action. *American Journal of Sociology*, 91: 522-556.

Olson, M. Jr. (1965): *The Logic of Collective Action*. Cambridge (Mass.), Harvard University Press.

Olson, M. Jr. (1982): *The Rise and Decline of Nations: Economic Growth, Stagnation, and Social Rigidities*. New Haven (Conn.), Yale University Press.

Olzak, S. (1986): A Competition Model of Ethnic Collective Action in American Cities, 1877-1889. In: Olzak, S. and Nagel, J. (eds.): *Competitive Ethnic Relations*. Orlando (Florida), Academic Press.

Olzak, S. (1992): *The Dynamics of Ethnic Competition and Conflict*. Stanford, CA, Stanford University Press.

Olzak, S.; Shanahan, S., and McEneaney, E. H. (1996): Poverty, Segregation, and Race Riots: 1960 to 1993. *American Sociological Review,* 61: 590-613.

Opp, K.-D. (1989): *The Rationality of Political Protest: A Comparative Analysis of Rational Choice Theory*. Boulder, CO, Westwiew.

Opp, K-D. (1991): Processes of Collective Political Action: A Dynamic Model and the Results of a Computer Simulation. *Rationality and Society*, 3(2): 215-251.

Opp, K-D. and Gern, C. (1993): Dissident Groups, Personal Networks, and Spontaneous Cooperation - The East German Revolution of 1989. *American Sociological Review*, 58: 659-680.

Orbell J.M. and Dawes, R. M. (1991): A "Cognitive Miser" Theory of Cooperators' Advantage. *American Political Science Review*, 85: 515-528.

Ortmann, A. and Tichy, L. K. (1999): Gender Differences in the Laboratory: Evidence from Prisoner's Dilemma Games. *Journal of Economic Behavior and Organization*, 39: 327-339.

Palfrey, T. and Rosenthal, H. (1983): A Strategic Calculus of Voting. *Public Choice*, 41: 7-53.

Parsons, T. (1937): *The Structure of Social Action.* Glencoe, Free Press.

Pingle, M. (1995): Imitation versus Rationality: An Experimental Perspective on Decision Making. *Journal of Socio-Economics*, 24(2): 281-316.

Poundstone, W. (1992): *Prisoner's Dilemma.* New York, Doubleday.

Rabbie, J. M. (1982): The Effects of Intergroup Competition and Cooperation on Intragroup and Intergroup Relationships. In: Derlega, V. J. and Grzelak, J. (eds.): *Cooperation and Helping Behavior*. New York, Academic Press.

Ragin, C. C. (1986): The Impact of Celtic Nationalism on Class Politics in Scotland and Wales. In: Olzak, S. and Nagel, J. (eds.): *Competitive Ethnic Relations*, Orlando, FL: Academic Press.

Rapoport, Amnon (1987): Research Paradigms and Expected Utility Models for the Provision of Step-Level Public Goods. *Psychological Review*, 94: 74-83.

Rapoport, Amnon and Amaldoss, W. (1999): Social Dilemmas Embedded in Between-Group Competitions: Effects of Contest and Distribution Rules. In: Foddy, M.; Smithson, M.; Schneider, S., and Hogg, M. (eds.): *Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects.* Philadelphia, Psychology Press / Taylor & Francis.

Rapoport, Amnon and Bornstein, G. (1987): Intergroup Competition for the Provision of Binary Public Goods. *Psychological Review*, 94: 291-299.

Rapoport, Amnon; Bornstein, G., and Erev, I. (1989): Intergroup Competition for Public Goods: Effects of Unequal Resources and Relative Group Size. *Journal of Personality and Social Psychology*, 56: 748-756.

Rapoport, Anatol and Chammah, A. M. (1965): *Prisoner's Dilemma: A Study in Conflict and Cooperation.* Ann Arbor, University of Michigan Press.

Rasbash, J.; Healy, M.; Browne, W., and Cameron, B. (1998): *MLwiN 1.02.0002*. Multilevel Models Project, Institute of Education.

Raub, W. and Snijders, C. (1997): Gains, Losses, and Cooperation in Social Dilemmas and Collective Action: The Effects of Risk Preferences. *Journal of Mathematical Sociology*, 22(3): 263-302.

Raub, W. and Weesie, J. (1990): Reputation and Efficiency in Social Interaction: an Example of Network Effects. *American Journal of Sociology*, 96: 626-654.

Richardson, L. F. (1948a): War-Moods: I. *Psychometrika,* 13: 147-174.

Richardson, L. F. (1948b): War-Moods: II. *Psychometrika,* 13: 197-232.

Richardson, L. F. (1951): Could an Arms-Race End Without Fighting? *Nature,* 168 (September 29): 567; 920.

Richardson, L. F. (1960): *Arms and Insecurity.* (Ed. by Rashevsky, N. and Trucco, E.) Chicago, Quadrangle Press.

Roth, A. E. (1995): Introduction to Experimental Economics. In: Kagel, J. H. and Roth, A. E. (eds.): *The Handbook of Experimental Economics.* Princeton NJ, Princeton University Press.

Roth, A. E. and Erev, I. (1995): Learning in Extensive-form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior*, 10: 6-38.

Rubin, J. Z. (1980): Experimental Research on Third Party Intervention in Conflict: Toward Some Generalizations. *Psychological Bulletin*, 87: 379-391.

Sandell, R. and Stern, C. (1998): Group Size and the Logic of Collective Action: A Network Analysis of a Swedish Temperance Movement 1896-1937. *Rationality and Society*, 10(3): 327-345.

Sandler, T. (1992): *Collective Action: Theory and Applications*. Ann Arbor, University of Michigan Press.

Schelling, T. C. (1960): *The Strategy of Conflict.* Cambridge (Mass.), Harvard University Press.

Schelling, T. C. (1971): Dynamic Models of Segregation. *Journal of Mathematical Sociology,* 1: 143-186.

Schelling, T. C. (1978): *Micromotives and Macrobehavior*. New York, W. W. Norton.

Schopler, J. and Insko, C. A. (1992): The Discontinuity Effect in Interpersonal and Intergroup Relations: Generality and Mediation. *European Review of Social Psychology,* 3: 121-151.

Schulz, U. and May, T. (1989): The Recording of Social Orientations with Ranking and Pair Comparison Procedures. *European Journal of Social Psychology*, 19(1): 41-59.

Schüssler, R. (1989): Exit Threats and Cooperation Under Anonymity. *Journal of Conflict Resolution,* 33(4): 728-749.

Scott, J. F. (1971): *Internalization of Norms.* Englewood Cliffs (NJ), Prentice-Hall.

Selten, R. (1975): Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory*, 4: 25-55.

Sen, A. K. (1977): Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs*, 6: 317-344.

Sherif, M. (1966a): *In Common Predicament. Social Psychology of Intergroup Conflict and Cooperation.* Boston, Houghton Mifflin.

Sherif, M. (1966b): *Group Conflict and Co-operation*. London, Routledge and Kegan Paul.

Sherif, M. and Sherif, C. W. (1969): *Social Psychology.* New York, Harper and Row.

Simmel, G. (1955) [1908]: Conflict. In: *Conflict. The Web of Group-Affiliations*. Glencoe, Illinois, Free Press.

Simmel, G. (1955) [1922]: The Web of Group-Affiliations. In: *Conflict. The Web of Group-Affiliations*. Glencoe, Illinois, Free Press.

Simon, H. A. (1976): From Substantive to Procedural Rationality. In: Latsis, S. J. (ed.): *Method and Appraisal in Economics*. Cambridge, Cambridge University Press.

Simon, H. A. (1981): *Models of Bounded Rationality*. Cambridge (Mass.), The MIT Press.

Skinner, B. F. (1938): *The Behavior of Organisms.* New York, Appleton-Century-Crofts.

Smelser, N. J. (1992): The Rational Choice Perspective. A Theoretical Assessment. *Rationality and Society*, 4: 381-410.

Smith, A. (1976)[1759]: *The Theory of Moral Sentiments*. Ed. by Raphael, D. D. and Macfie, A. L. Oxford, Clarendon Press.

Solnick, S. J. and Schweitzer, M. E. (1999): The Influence of Physical Attractiveness and Gender on Ultimatum Game Decisions. *Organizational Behavior and Human Decision Processes*. 79(3): 199-215.

Stokman, F. N. and Zeggelink, E. P. H. (1996): 'Self-Organizing' Friendship Networks. In: Liebrand, W. B. G. and Messick, D. M. (eds.): *Frontiers in Social Dilemma Research.* Berlin, Springer.

Suleiman, R. and Fischer, I. (1996): The Evolution of Cooperation in a Simulated Inter-Group Conflict. In: Liebrand, W. B. G. and Messick, D. M. (eds.): *Frontiers in Social Dilemma Research.* Berlin, Springer.

Suleiman, R. and Fischer, I. (2000): When One Decides for Many: The Effect of Delegation Methods on Cooperation in Simulated Inter-group Conflicts. *Journal of Artificial Societies and Social Simulation,* 3(4), http://www.soc.surrey.ac.uk/JASSS/3/4/1.html

Suleiman, R. and Or-Chen, K. (1999): Providing Step-Level Public Goods Under Uncertainty: The Case of Probable External Supply. In: Foddy, M.; Smithson, M.; Schneider, S., and Hogg, M. (eds.): *Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects*. Philadelphia, Psychology Press / Taylor & Francis.

Sumner, W. G. (1974) [1906]: *Folkways*. New York, Arno Press.

Suttles, G. R. (1967): *The Social Order of the Slum*. Chicago, IL, University of Chicago Press.

Tajfel, H. (1970): Experiments in Intergroup Discrimination. *Scientific American*, 223: 96-102.

Tajfel, H. (1981): *Human Groups and Social Categories. Studies in Social Psychology.* Cambridge, Cambridge University Press.

Tajfel, H. (1982): Social Psychology of Intergroup Relations. *Annual Review of Psychology,* 33: 1-39.

Tajfel, H. and Turner, J. C. (1986): The Social Identity Theory of Intergroup Behavior. In: Worchel, S. and Austin, W. G. (eds.): *The Social Psychology of Intergroup Relations*. Chicago, Nelson-Hall.

Tajfel, H.; Billig, M. G.; Bundy, H. P., and Flament, C. I. (1971): Social Categorization and Intergroup Behavior. *European Journal of Social Psychology*, 1: 149-178.

Takács, K. (2001): Structural Embeddedness and Intergroup Conflict. *Journal of Conflict Resolution*, 45(6): 743-769.

Taylor, M. (1987): *The Possibility of Cooperation*. Cambridge, MA, Cambridge University Press.

Thorndike, E. L. (1898): *Animal Intelligence: An Experimental Study of the Associative Processes in Animals*. Psychological Monographs, 2.

Tickle-Degnen, L. and Rosenthal, R. (1990): The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry*, 1: 285-293.

Tingsten, H. (1963) [1937]: *Political Behavior: Studies in Election Statistics*. Totowa, NJ, Bedminster.

Turner, J. C. (1982): Intergroup Conflict and Cooperation. In: Colman, A. (ed.): *Cooperation and Competition in Humans and Animals.* Berkshire, Nostrand Reinhold.

Tversky, A. and Kahneman, D. (1986): Rational Choice and the Framing of Decisions. In: Hogarth, R. M. and Reder, M. W. (eds.): *Rational Choice: The Contrast between Economics and Psychology*. Chicago, The University of Chicago Press.

Tversky, A. and Kahneman, D. (1992): Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5: 297-323.

van Assen, M. A. L. M. (2001): *Essays on Actor Models in Exchange Networks and Social Dilemmas*. Amsterdam, Thesis Publishers.

van Assen, M. A. L. M. and Snijders, C. (2002): Effects of Risk Preferences in Social Dilemmas: A Game-Theoretical Analysis and Evidence from Two Experiments. In: Suleiman, R.; Budescu, D., Messick, D. M., and Fischer, I. (eds.): *Contemporary Approaches to Social Dilemma Research.* Cambridge, Cambridge University Press.

Vanberg, V. J. (2002): Rational Choice vs. Program-Based Behavior. *Rationality and Society*, 14(1): 7-54.

Vanberg, V. and Congleton, R. (1992): Rationality, Morality and Exit. *American Political Science Review*, 86: 418-431.

van de Kragt, A. J. C.; Dawes, R. M., and Orbell, J. M. (1988): Are People who Cooperate 'Rational Altruists'? *Public Choice*, 56: 233-247.

van de Kragt, A. J. C.; Orbell, J. M., and Dawes, R. M. (1983): The Minimal Contributing Set as a Solution to Public Goods Problems. *American Political Science Review*, 77: 112-122.

van Duijn, M. A. J.; van Busschbach, J. T., and Snijders, T. A. B. (1999): Multilevel Analysis of Personal Networks as Dependent Variables. *Social Networks*, 21: 187-209.

van Lange, P. A. M. (1999): The Pursuit of Joint Outcomes and Equality in Outcomes: an Integrative Model of Social Value Orientation. *Journal of Personality and Social Psychology*. 77(2): 337-349.

van Lange, P. A. M.; Liebrand, W. B. G.; Messick, D. M., and Wilke, H. A. M. (1992): Social Dilemmas: The State of the Art. Introduction and Literature Review. In: Liebrand, W. B. G.; Messick, D. M., and Wilke, H. A. M. (eds.): *Social Dilemmas. Theoretical Issues and Research Findings.* Oxford, Pergamon Press.

van Lange, P. A. M.; Otten, W.; De Bruin, E. M. N., and Joireman, J. A. (1997): Development of Prosocial, Individualistic, and Competitive Orientations: Theory and Preliminary Evidence. *Journal of Personality and Social Psychology*, 73(4): 733-746.

von Neumann, J. (1966): *Theory of Self-Reproducing Automata* (edited and completed by Burks, A. W.). Urbana, University of Illinois Press.

von Neumann, J. and Morgenstern, O. (1944): *Theory of Games and Economic Behavior.* Princeton (NJ), Princeton University Press.

Wagner, T. (1998): Reciprocity and Efficiency. *Rationality and Society*, 10(3): 347-375.

Wasserman, S. and Faust, K. (1994): *Social Network Analysis. Methods and Implications.* Cambridge, Cambridge University Press.

Watanabe, Y. and Yamagishi, T. (1999): Emergence of Strategies in a Selective-Play Environment with Geographic Mobility: A Computer Simulation. In: Foddy, M. et al. (eds.): *Resolving Social Dilemmas: Dynamic, Structural, and Intergroup Aspects*. Philadelphia, Psychology Press / Taylor & Francis.

Weber, M. (1976) [1913]: *Economy and Society. An Outline of Interpretive Sociology.* (Ed. by Roth, G. and Wittich, C.) Berkeley, University of California Press.

Wedekind, C. and Milinski, M. (2000): Cooperation Through Image Scoring in Humans. *Science*, 288(5 May): 850-852.

Whatley, M. A.; Webster, M. J.; Smith, R. H., and Rhodes, A. (1999): The Effect of a Favor on Public and Private Compliance: How Internalized Is the Norm of Reciprocity? *Basic and Applied Social Psychology*, 21(3): 251-259.

Whyte, J. H. (1986): How is the Boundary Maintained between the Two Communities in Northern Ireland? *Ethnic and Racial Studies,* 9: 219-234.

Williams, R. (1947): *The Reduction of Intergroup Tensions. A Survey of Research on Problems of Ethnic, Racial, and Religious Group Relations.* New York, Social Science Research Council.

Willms, J. D. and Paterson, L. (1995): A Multilevel Model for Community Segregation. *Journal of Mathematical Sociology*, 20 (1): 23-40.

Wintrobe, R. (1995): Some Economics of Ethnic Capital Formation and Conflict. In: Breton, A.; Galeotti, G.; Salmon, P., and Wintrobe, R. (eds.): *Nationalism and Rationality.* Cambridge, Cambridge University Press.

Wu, J. and Axelrod, R. (1995): How to Cope with Noise in the Iterated Prisoner's Dilemma. *Journal of Conflict Resolution*, 39(1): 183-189.

Young, H. P. (1993): The Evolution of Conventions. *Econometrica,* 61: 57-84.

Zeggelink, E. P. H. (1993): *Strangers into Friends. The Evolution of Friendship Networks Using an Individual Oriented Modeling Approach.* Amsterdam, Thesis Publishers.

Samenvatting

Probleemstelling

Intergroepsconflict, vooral als er geweld bij speelt, behoort tot de ernstigste problemen waar de mensheid momenteel mee geconfronteerd wordt. Tal van factoren beïnvloeden het ontstaan van dergelijke conflicten. Intergroepsconflict in het dagelijks leven is erg ingewikkeld, daarom kan niet worden ingegaan op alle facetten ervan. Het zijn zowel de ernst als deze complexiteit die conflict oplossing tot een uitdagende taak maken. Deze studie levert een bijdrage aan het onderzoek naar een aantal factoren die van invloed zijn op het ontstaan van ongunstige groepsconflicten. *Intergroepsconflict ontstaat als groepen concurreren om materiële en immateriële hulpbronnen*. Teneinde te onderzoeken waardoor competitie tot slechte uitkomsten leidt, richt deze studie zich op de effecten van *structurele en tijdelijke inbedding*. In het bijzonder is het doel de condities vast te stellen waaronder *segregatie* samengaat met conflict en welke de onderliggende mechanismen hierbij zijn. Omdat de meeste intergroepsrelaties niet gescheiden kunnen worden van hun historische context wordt ook onderzocht hoe conflict in de loop van de tijd verandert en wat de endogene determinanten zijn die leiden tot positieve en negatieve veranderingen.

Groepen vormen geen homogene eenheid, omdat ze bestaan uit bewust handelende individuen die op basis van hun eigen vrije wil beslissingen nemen. Hierdoor kan een analyse van intergroepsconflict niet beperkt blijven tot het groepsniveau. Handelen op basis van vrijwilligheid komt bijvoorbeeld voor in tijden van oorlog als burgers zich aanmelden om in dienst te gaan, soldaten zich vrijwillig aanmelden om aan het front te vechten en frontsoldaten zich aanbieden om speciale opdrachten uit te voeren (Coleman, 1990). Een ander voorbeeld wordt gevormd door zelfmoordcommando's; zij riskeren niet alleen hun leven maar offeren het in het geloof dat het bij zal dragen aan het bereiken van een groepsdoel, zoals onafhankelijkheid. Over het algemeen geldt dat individuele participatie in intergroepsconflict zelden een grote investering vereist, wel is het zo dat aan het helpen van de groep altijd een bepaalde hoeveelheid kosten is verbonden. De centrale vraag in deze studie is *waarom mensen bereid zijn zulke offers te brengen, zelfs als het niet van hen verlangd wordt en hun gedrag fatale gevolgen kan hebben*. Om intergroepsconflict te kunnen verklaren is het nodig te achterhalen waarop individuele participatie gebaseerd is.

Deze studie richt zich op de invloed van factoren die voortkomen uit *groepscompetitie*, *het sociale netwerk van de groepsleden* of de *tijdsperiode* waarin de intergroep- en interpersoonlijke relaties spelen. Intergroepscompetitie en -vergelijking stimuleren deelname in conflict via *publieke goederen* die verdeeld worden onder de leden van de winnende partij, zoals territorium, trots of sociale identiteit. Sociale relaties met relevante anderen leggen beperkingen op aan individuele beslissingen via verschillende vormen van *sociale controle*, waardoor participatie zowel kan worden bevorderd als verminderd. De historische context is van belang omdat mensen zich gedragen op basis van bepaalde *besluitvormingsheuristieken*. Deze heuristieken zijn gedeeltelijk gebaseerd op de *ervaring* die individuen in de loop van de tijd opdoen. Zo kan men van het verleden hebben geleerd of eerdere beledigingen vergolden worden of niet. Op vergelijkbare wijze worden individuele handelingen sterk beïnvloed door *verwachtingen* over toekomstige ontmoetingen.

Structurele inbedding en intergroepsconflict

Het voornaamste theoretische doel van deze studie is de *constructie van een model van intergroepscompetitie* dat inzicht geeft in het ontstaan van nadelige conflicten als gevolg van structurele en temporele inbedding. Het *Intergroup Public Goods (IPG) game* model voor competitieve intergroepsrelaties vormt de basis van dit model (Rapoport en Bornstein, 1987). Het belangrijkste voordeel van het IPG model is de coherente presentatie van enerzijds de duale onderlinge afhankelijkheid bij competitie tussen groepen en anderzijds van de negatieve interdependentie tussen groepen en het *free-rider* probleem binnen groepen. In zijn originele vorm kan het IPG model echter niet omgaan met *effecten van structurele inbedding*. Structurele inbedding beïnvloedt individuele beslissingen via *mechanismen van sociale controle*. Groepsleden worden beloond met *sociale selectieve prikkels*, zoals respect, wanneer zij offers brengen voor de groep. Groepsleden die geen inspanningen hebben geleverd worden gestraft met dergelijke prikkels. Daarnaast worden mensen beloond met *gedragsbevestiging* als zij zich net zo gedragen als andere groepsleden. Zij die de eigen groep afvallen krijgen *beloningen van de tegenpartij*. Een belangrijk onderdeel van deze studie is dat deze vormen van sociale controle zijn toegevoegd aan het IPG model voor competitieve intergroep relaties en dat het model daarmee in een structuur van individuele relaties is ingebed.

Omdat sociale controle via interpersoonlijke relaties wordt uitgeoefend, heeft de *netwerk structuur van de individuele relaties invloed op de uitkomst van intergroupscompetitie*. Een centrale onderzoeksvraag is: *onder welke structurele condities is de kans op intergroepsconflict groter en onder welke condities kan een vreedzame samenwerking verwacht worden*.

In het ingebedde IPG model kan sterke sociale controle, onder bepaalde structurele condities, tot een ongewoon sociaal dilemma leiden. In dit sociale dilemma is iedereen beter af door aan zijn groep *bij te dragen*, maar pakt algehele bijdrage slechter uit voor een ieder dan vreedzame coëxistentie. In deze structuur van onderlinge afhankelijkheid is conflict het onbedoelde gevolg van bedoelde individuele acties. Het model voorspelt dat een dergelijk sociaal dilemma optreedt als hechte contacten binnen groepen meer gebruikt worden voor verspreiding van sociaal selectieve prikkels (beloning voor je inzetten voor de groep) dan voor verspreiding van gedragsbevestiging (beloning voor 'hetzelfde doen als relevante anderen'). Een extra voorwaarde voor het ontstaan van een sub-optimale conflictsituatie is dat er weinig interpersoonlijke relaties tussen strijdende groepen zijn of dat de beloningen van de tegenpartij van gering belang zijn.

Doel is niet enkel het aantonen van de condities waaronder algeheel conflict zich ontwikkelt als een sociaal dilemma, maar om te laten zien welke structurele condities de kans op intergroepsconflict vergroten. Meer in het bijzonder is het doel van deze studie te kijken waarom en onder welke omstandigheden *segregatie* tot intergroepsconflict aanzet. Simulaties en experimenten worden gebruikt voor het beantwoorden van deze vraag.

Simulaties worden gebruikt om de algemene theoretische relatie tussen segregatie en intergroepsconflict te exploreren. In de simulaties worden de structureel ingebedde IPG games gespeeld door agenten binnen gesimplificeerde netwerkstructuren. Het effect van de segregatie op de kans op intergroepsconflict kan worden uitgedrukt in een S-vormige functie. Dit betekent dat in het algemeen *segregatie de kans op groepsconflict vergroot*. Echter, dit effect is gering bij zeer lage en bij hoge waarden van segregatie, afhankelijk van waarden op andere parameters. Dit zijn *vloer- en plafond effecten*. Verder is het *segregatie effect sterker als de locale selectieve prikkels belangrijker zijn dan gedragsbevestiging van de buren*.

De hypothesen zijn gebaseerd op de resultaten van de simulaties. Ze zijn vervolgens middels experimenten getoetst. Teneinde de structurele condities in de toetsing mee te kunnen nemen is het volgende *experimentele model* ontwikkeld: verdeeld over twee groepen van vijf hebben proefpersonen de opdracht gekregen IPG games te spelen zonder daarbij met elkaar te communiceren. Teneinde te toetsen op de aanwezigheid van segregatie effecten en onderliggende controle mechanismen zijn de zitplaatsen gevarieerd en zijn de zichtbaarheidcondities gemanipuleerd. Vier structurele condities zijn toegepast, één waarin proefpersonen uit de twee groepen volledig van elkaar gescheiden zijn (volledige separatie), één waarin de twee groepen in een sterk gescheiden opstelling zitten (hoge segregatie), één waarin zij gemiddeld gescheiden zitten (gemiddelde segregatie), en één waarin de twee groepen in een gemengde opstelling zitten (lage segregatie). Het effect van de segregatie op de kans op intergroepsconflict is getoetst door de spelresultaten van de verschillende opstellingen met elkaar te vergelijken.

We verwachten dat ook *geïnternaliseerde* sociale controle individuele beslissingen kan beïnvloeden. Teneinde geïnternaliseerde vormen van sociale controle waar te kunnen nemen, hebben de proefpersonen in een onderdeel van het experiment de mogelijkheid oogcontact met hun buren tot stand brengen. In latere delen van het experiment worden ook *directe vormen van sociale controle* geïntroduceerd als financiële beloningen. Dit gebeurt om de relatieve invloed van geïnternaliseerde en directe vormen van sociale controle op de individuele beslissingen te kunnen toetsen.

De resultaten van de experimenten *bevestigen de segregatie hypothese gedeeltelijk*. Zoals voorspeld, is intergroepsconflict het minst waarschijnlijk wanneer er sprake is van een laag segregatieniveau van de leden van beide groepen. Echter, bij gemiddelde segregatie is conflict niet minder waarschijnlijk dan bij sterke segregatie. Deels kan dit verklaard worden door een plafond effect, deels door de hoge contributiebereidheid in de gemiddelde gesegregeerde setting.

Over het algemeen laten proefpersonen zich sterk door financiële sociale controle beïnvloeden. Interessanter is echter de bevinding dat proefpersonen hun beslissingen aanpassen aan de verwachting van wat hun buren zullen doen, als die tot de eigen groep behoren, zelfs al is er enkel oogcontact. Dit onderschrijft het bestaan van geïnternaliseerde gedragsbevestiging. Er is ook enige ondersteuning voor het effect van geïnternaliseerde selectieve prikkels. Hoewel het effect kleiner is dan dat van de financiële sociale controle, blijkt de geïnternaliseerde sociale controle van buren van de eigen groep een belangrijke voorspeller van de individuele geneigdheid om bij te dragen. Geïnternaliseerde beloningen van de tegenpartij worden enkel onder bepaalde condities aangetroffen. Proefpersonen ondersteunen hun groepsgenoten niet als zij omringd worden door leden van de andere groep die behoren tot de andere sekse.

Samenvattend kan gesteld worden dat dit nieuwe experimentele model in belangrijke mate bijdraagt aan het inzicht in het effect van structuur en sociale controle op intergoepssituaties. De resultaten tonen met name hoe belangrijk oogcontact tussen proefpersonen in een laboratorium is voor het activeren van geïnternaliseerde sociale controle mechanismen.

De belangrijkste *onderzoeksvragen* (V), hypotheses (H), en **experimentele resultaten** (R) over de effecten van segregatie op intergroepsconflict worden hieronder kort samengevat.

V *Waarom en onder welke condities versterkt segregatie de kans op intergroepsconflict? Is de kans op intergroepsconflict in een experimentele setting groter wanneer groepsleden in een sterk gesegregeerde opstelling zitten? Heeft segregatie ook effect als proefpersonen alleen oogcontact met elkaar hebben en niet het gedrag kunnen observeren?*

H Netwerkstructuren zijn van invloed op het ontstaan van conflicten omdat zij de sociale controle mechanismen kanaliseren. In een gesegregeerde structuur is de kans op intergroepsconflict groter omdat hechte connecties groepsleden in staat stellen selectieve prikkels te verspreiden ten behoeve van mobilisatie. Afwezigheid van relaties tussen groepen vermindert de effect van beloningen van de tegenpartij. Omdat sociale controle vaak geïnternaliseerd is, vergroot segregatie de kans op intergroepsconflict ook als individuen enkel oogcontact met elkaar hebben.

R **Een vergelijking van resultaten van experimenten met verschillende structurele configuraties laat zien dat intergroepsconflict het minst waarschijnlijk is als groepsleden in een gemengde opstelling zitten. Echter,**

conflict is niet minder waarschijnlijk bij gemiddelde segregatie dan bij hoge segregatie, ook als buren slechts oogcontact met elkaar hebben.

Voor het beantwoorden van de volgende vragen is uitleg over individueel gedrag nodig dat deze macro uitkomsten verklaart.

V *Waarom en onder welke omstandigheden nemen individuen deel aan collectieve actie die de belangen van de andere groep schaadt en ook de eigen groep zou kunnen schaden? Wat zijn de onderliggende mechanismen van netwerk effecten op het niveau van interpersoonlijke relaties?*

Hoe dragen verschillende vormen van sociale controle, namelijk selectieve prikkels, gedragsbevestiging en beloningen van de tegenpartij bij aan intergroepsconflict en wat is hun invloed? Beïnvloeden deze vormen van sociale controle individuele beslissingen direct en indirect (*in geïnternaliseerde vormen*) *in een experimentele omgeving?*

H Individuen kunnen tot participatie aangezet worden als directe en geïnternaliseerde vormen van sociale controle de kosten van participatie compenseren. Sociale controle resulteert in een verdeling van positieve en negatieve prikkels die de individuele beslissingen beperken. De verdeling van die prikkels is afhankelijk van de verwachtingen en het gedrag van het individu en van relevante anderen. Afhankelijk van het type sociale controle en de samenstelling van het ego netwerk van het individu kan sociale controle participatie vergemakkelijken of bemoeilijken. Sociale controle mechanismen op het niveau van interpersoonlijke relaties bewerkstelligen de macro structurele effecten op conflict.

Sociale selectieve prikkels motiveren groepsleden tot participatie. Beloningen van de tegenpartij verhinderen participatie. Deze vormen van sociale controle zijn dus verantwoordelijk voor het effect dat segregatie heeft op intergroepsconflict. Gedragsbevestiging, echter, heeft twee kanten; het kan bijdragen aan totstandkoming van wijd verbreid activisme of juist bijdragen aan de verspreiding van vreedzaam gedrag. Hoe sterker de normatieve druk (selectieve prikkels) in verhouding tot de confirmerende druk (gedragsbevestiging), hoe sterker het effect van segregatie op intergroepsconflict.

R **De proefpersonen in het experiment worden sterk door directe vormen van sociale controle beïnvloedt. De uitkomsten onderschrijven de aanwezigheid van geïnternaliseerde gedragsbevestiging en, hoewel minder overtuigend, van geïnternaliseerde selectieve prikkels. Geïnternaliseerde beloningen van de tegenpartij komen enkel onder specifieke condities voor. De effecten van de drie vormen van sociale controle zijn voldoende groot om binnen eenzelfde segregatieconditie toch tot andere kansen op conflict te leiden. Het segregatieeffect is sterker bij normatieve druk dan bij confirmerende druk.**

Naast deze centrale onderzoeksvragen is in de simulaties onderzocht of de effecten en macro consequenties van sociale controle mechanismen al dan niet afhangen van assumpties over *individuele rationaliteit* en *toegang tot informatie*. Het effect van sociale controle mechanismen en structurele configuraties op intergroepsconflict is voor vier gedragsmodellen onderzocht. Deze modellen verschillen wat betreft het niveau van calculerende rationaliteit van de individuen en de hoeveelheid informatie waarover individuen beschikken. We zijn met name geïnteresseerd in het effect van segregatie op de kans van conflict onder de verschillende modelspecificaties. De simulaties laten zien dat, onder bepaalde structurele condities, rationele individuen met meer informatie meer in een schadelijk conflict terecht komen dan minder rationele actoren. Verder versterken zwaardere aannames over de individuele rationaliteit het effect van segregatie op intergroepsconflict enigszins.

Daarnaast tonen de simulaties dat niet alleen segregatie, maar ook *andere kenmerken van het sociale netwerk* van invloed zijn op de kans op intergroepsconflict. Zo kunnen minderheden omringd door de meerderheid mobilisatie onderdrukken, kunnen mensen die twee subgroepen verbinden mobilisatie bevorderen en kunnen zwak gerelateerde subgroepen, afhankelijk van de gedragsassumpties, intergroepsconflict stimuleren dan wel ontmoedigen.

Tijdelijke inbedding en intergroepsconflict

Naast de structurele inbedding van gedrag speelt ook *temporele inbedding* een cruciale rol in intergroepsrelaties. Ervaringen opgedaan bij vorige ontmoetingen en verwachtingen ten aanzien van toekomstige relaties hebben een grote invloed op huidige houdingen en gedrag in de intergroepscontext. De dagelijkse praktijk toont ons vele voorbeelden van uit de hand lopende confrontaties en duurzame conflicten tussen groepen. Hoe kan het ontstaan van zulke conflicten verklaard worden? Intergroepsrelaties zijn (gelukkig) vaker vreedzaam dan gewelddadig. Vandaar dat de nadruk ligt op de vraag onder welke omstandigheden vreedzaam samenleven mogelijk is.

In de gebruikte verklaring over de dynamiek van intergroepsrelaties ligt de nadruk op mechanismen op het individuele niveau. Individuele beslissingen worden sterk beïnvloed door enerzijds ervaringen uit het verleden en anderzijds verwachtingen ten aanzien van de toekomst. Met betrekking tot de temporele inbedding van gedrag is in dit onderzoek gebruik gemaakt van het idee van *beperkte rationaliteit*. Er is vanuit gegaan dat beslissingen van individuen niet gebaseerd zijn op een volledig overzicht van gebeurtenissen uit het verleden gecombineerd met ingewikkelde berekeningen over de lange-termijn gevolgen van die beslissingen. In plaats daarvan maken personen gebruik van *eenvoudige gedragsheuristieken*.

Met betrekking tot de vraag welke beslissingsregels gebruikt worden zijn op basis van de literatuur hypotheses geformuleerd over het bestaan van drie mechanismen, te weten vermeende invloed, leren door beloningen en wederkerigheid. Het eerste mechanisme zorgt ervoor dat een individu investeert in het collectieve goed als hij verwacht dat zijn beslissing *invloed* heeft op de uitkomst van de competitie. Het *leren door beloningen mechanisme* zorgt ervoor dat individuen vasthouden aan een beslissing die eerder heeft geleid tot een bevredigend resultaat, en van strategie veranderen in het geval van een onbevredigend resultaat. *Wederkerigheidsmechanismen* lokken vreedzame reacties uit op geobserveerd vreedzaam gedrag van anderen en vergeldingsgedrag als reactie op investeringen in conflict. Individuen kunnen zowel reageren op het collectieve gedrag van de andere groep als op het gedrag van relevante individuen. In dit onderzoek zijn alle voornoemde mechanismen in beschouwing genomen. Het belangrijkste argument voor deze benadering is dat gedragsmechanismen kunnen verschillen tussen personen, maar ook binnen personen.

Een nieuw gezichtspunt in dit onderzoek is dat gelijktijdig getoetst is op de invloed van verschillende gedragsheuristieken op beslissingen van individuen. Toetsing van de invloed van temporele inbedding is mogelijk doordat een gedeelte van het experiment bestaat uit het spelen van *herhaalde* IPG spelen. In tegenstelling tot de *enkelvoudige* spelen, waarin proefpersonen geen feedback krijgen over eerdere uitkomsten, krijgen ze in de herhaalde spelen informatie over de uitkomst van het vorige spel en over de beslissingen van hun buren. Met het introduceren van structurele condities in de herhaalde spelen is het ook mogelijk om *interactie-effecten van structurele en temporele aspecten* te toetsen.

Hieronder volgt een korte samenvatting van de belangrijkste *onderzoeksvragen* **(V)** over de invloed van temporele inbedding en over de interacties van structuur en tijd. Tevens worden de hypotheses **(H)** gemeld die zijn geformuleerd voor deze onderzoeksvragen en de conclusies die zijn getrokken op basis van de **experimentele resultaten (R)**.

V *Hoe veranderen intergroepsrelaties door de tijd heen? Bestaan er typische ontwikkelingen, zoals een eindeloze regressie van conflict of een "spiraal van vrede"? Verschillen deze ontwikkelingen in termen van structurele condities?*

H Voorspeld is dat er enkele typische ontwikkelingen voorkomen, zoals stabiele vrede, duurzaam conflict, spiraal van vrede en spiraal van conflict. Stabiele vrede en spiraal van vrede zijn waarschijnlijker in een opstelling waarbij de zitplaatsen gemengd zijn. Duurzaam conflict en spiraal van conflict zijn waarschijnlijker in een opstelling met hoge segregatie.

R **De resultaten laten zien dat de ontwikkelingen van stabiele vrede en spiraal van vrede vaker voorkomen als leden van de teams in een gemengde opstelling zitten. In sessies met hoge segregatie ontstaat duurzaam conflict, maar met afwisselende voordelen voor de twee groepen en niet gradueel, zoals de spiraal van conflict hypothese voorspelt. De hoogte van de**

investeringen neemt toe met het introduceren van sociale controle tijdens het experiment.

V *Wat zijn de eenvoudige gedragsheuristieken die de beslissingen van individuen leiden en bijgevolg verantwoordelijk zijn voor het ontstaan van macro ontwikkelingen? Zijn vermeende invloed, leren door beloningen, en wederkerigheid belangrijke determinanten van individuele acties in herhaalde competities tussen groepen?*

H De verwachting is dat beslissingen van proefpersonen geleid worden door eenvoudige gedragsregels die voor een deel gebaseerd zijn op ervaringen van ontmoetingen in het verleden en voor een deel gebaseerd zijn op verwachtingen over toekomstige interacties. De combinatie van vermeende invloed, leren door beloningen, intergroepswederkerigheid en lokale wederkerigheid is van invloed op de huidige beslissingen van proefpersonen.

R De resultaten rechtvaardigen het toepassen van een combinatie van verschillende gedragsregels voor het verklaren van het investeringsgedrag van de proefpersonen. De hypothese over vermeende invloed is bevestigd, maar met die kanttekening dat proefpersonen die verwachten dat hun groep het volgende spel zal winnen een opvallend hoge geneigdheid tot investeren vertonen. Leren door beloningen blijkt echter niet symmetrisch te werken; er is geen bewijs gevonden voor het belonen van de keuze om te investeren. Verder is er weinig steun voor het bestaan van intergroepswederkerigheid. Alleen een nederlaag lokt significante vergeldingsreacties uit, maar dit kan ook het gevolg zijn van simpele agressieve houdingen. Lokale wederkerigheid hangt af van persoonseigenschappen van de proefpersoon en het geslacht en de groepsidentiteit van de buurman of buurvrouw.

ICS dissertation series

The ICS-series presents dissertations of the Interuniversity Center for Social Science Theory and Methodology. Each of these studies aims at integrating explicit theory formation with state-of-the-art empirical research or at the development of advanced methods for empirical research. The ICS was founded in 1986 as a cooperative effort of the universities of Groningen and Utrecht. Since 1992, the ICS expanded to the University of Nijmegen. Most of the projects are financed by the participating universities or by the Netherlands Organization for Scientific Research (NWO). The international composition of the ICS graduate students is mirrored in the increasing international orientation of the projects and thus of the ICS-series itself.

1) C. van Liere (1990), *Lastige Leerlingen. Een empirisch onderzoek naar sociale oorzaken van probleemgedrag op basisscholen.* Amsterdam: Thesis Publishers.
2) Marco H.D. van Leeuwen (1990), *Bijstand in Amsterdam, ca. 1800 - 1850. Armenzorg als beheersings- en overlevingsstrategie.* ICS dissertation, Utrecht
3) I. Maas (1990), *Deelname aan podiumkunsten via de podia, de media en actieve beoefening. Substitutie of leereffecten?* Amsterdam: Thesis Publishers.
4) M.I. Broese van Groenou (1991), *Gescheiden Netwerken. De relaties met vrienden en verwanten na echtscheiding.* Amsterdam: Thesis Publishers.
5) Jan M.M. van den Bos (1991), *Dutch EC Policy Making. A Model-Guided Approach to Coordination and Negotiation.* Amsterdam: Thesis Publishers.
6) Karin Sanders (1991), *Vrouwelijke Pioniers. Vrouwen en mannen met een 'mannelijke' hogere beroepsopleiding aan het begin van hun loopbaan.* Amsterdam: Thesis Publishers.
7) Sjerp de Vries (1991), *Egoism, Altruism, and Social Justice. Theory and Experiments on Cooperation in Social Dilemmas.* Amsterdam: Thesis Publishers.
8) Ronald S. Batenburg (1991), *Automatisering in bedrijf.* Amsterdam: Thesis Publishers.
9) Rudi Wielers (1991), *Selectie en allocatie op de arbeidsmarkt. Een uitwerking voor de informele en geïnstitutionaliseerde kinderopvang.* Amsterdam: Thesis Publishers.
10) Gert P. Westert (1991), *Verschillen in ziekenhuisgebruik.* ICS dissertation, Groningen
11) Hanneke Hermsen (1992), *Votes and Policy Preferences. Equilibria in Party Systems.* Amsterdam: Thesis Publishers.
12) Cora J.M. Maas (1992), *Probleemleerlingen in het basisonderwijs.* Amsterdam: Thesis Publishers.
13) Ed A.W. Boxman (1992), *Contacten en carrière. Een empirisch-theoretisch onderzoek naar de relatie tussen sociale netwerken en arbeidsmarktposities.* Amsterdam: Thesis Publishers.
14) Conny G.J. Taes (1992), *Kijken naar banen. Een onderzoek naar de inschatting van arbeidsmarktkansen bij schoolverlaters uit het middelbaar beroepsonderwijs.* Amsterdam: Thesis Publishers.

15) Peter van Roozendaal (1992), *Cabinets in Multi-Party Democracies. The Effect of Dominant and Central Parties on Cabinet Composition and Durability.* Amsterdam: Thesis Publishers.
16) Marcel van Dam (1992), *Regio zonder regie. Verschillen in en effectiviteit van gemeentelijk arbeidsmarktbeleid.* Amsterdam: Thesis Publishers.
17) Tanja van der Lippe (1993), *Arbeidsverdeling tussen mannen en vrouwen.* Amsterdam: Thesis Publishers.
18) Marc A. Jacobs (1993), *Software: Kopen of Kopiëren? Een sociaal-wetenschappelijk onderzoek onder PC-gebruikers.* Amsterdam: Thesis Publishers.
19) Peter van der Meer (1993), *Verdringing op de Nederlandse arbeidsmarkt. sector- en sekseverschillen.* Amsterdam: Thesis Publishers.
20) Gerbert Kraaykamp (1993), *Over lezen gesproken. Een studie naar sociale differentiatie in leesgedrag.* Amsterdam: Thesis Publishers.
21) Evelien Zeggelink (1993), *Strangers into Friends. The Evolution of Friendship Networks Using an Individual Oriented Modeling Approach.* Amsterdam: Thesis Publishers.
22) Jaco Berveling (1994), *Het stempel op de besluitvorming. Macht, invloed en besluitvorming op twee Amsterdamse beleidsterreinen.* Amsterdam: Thesis Publishers.
23) Wim Bernasco (1994), *Coupled Careers. The Effects of Spouse's Resources on Success at Work.* Amsterdam: Thesis Publishers.
24) Liset van Dijk (1994), *Choices in Child Care. The Distribution of Child Care Among Mothers, Fathers and Non-Parental Care Providers.* Amsterdam: Thesis Publishers.
25) Jos de Haan (1994), *Research Groups in Dutch Sociology.* Amsterdam: Thesis Publishers.
26) K. Boahene (1995), *Innovation Adoption as a Socio-Economic Process. The Case of the Ghanaian Cocoa Industry.* Amsterdam: Thesis Publishers.
27) Paul E.M. Ligthart (1995), *Solidarity in Economic Transactions. An Experimental Study of Framing Effects in Bargaining and Contracting.* Amsterdam: Thesis Publishers.
28) Roger Th. A.J. Leenders (1995), *Structure and Influence. Statistical Models for the Dynamics of Actor Attributes, Network Structure, and their Interdependence.* Amsterdam: Thesis Publishers.
29) Beate Völker (1995), *Should Auld Acquaintance Be Forgot...? Institutions of Communism, the Transition to Capitalism and Personal Networks: the Case of East Germany.* Amsterdam: Thesis Publishers.
30) A. Cancrinus-Matthijsse (1995), *Tussen hulpverlening en ondernemerschap. Beroepsuitoefening en taakopvattingen van openbare apothekers in een aantal West-Europese landen.* Amsterdam: Thesis Publishers.
31) Nardi Steverink (1996), *Zo lang mogelijk zelfstandig. Naar een verklaring van verschillen in oriëntatie ten aanzien van opname in een verzorgingstehuis onder fysiek kwetsbare ouderen.* Amsterdam: Thesis Publishers.
32) Ellen Lindeman (1996), *Participatie in vrijwilligerswerk.* Amsterdam: Thesis Publishers.

33) Chris Snijders (1996), *Trust and Commitments.* Amsterdam: Thesis Publishers.
34) Koos Postma (1996), *Changing Prejudice in Hungary. A Study on the Collapse of State Socialism and Its Impact on Prejudice Against Gypsies and Jews.* Amsterdam: Thesis Publishers.
35) Jooske T. van Busschbach (1996), *Uit het oog, uit het hart? Stabiliteit en verandering in persoonlijke relaties.* Amsterdam: Thesis Publishers.
36) René Torenvlied (1996), *Besluiten in uitvoering. Theorieën over beleidsuitvoering modelmatig getoetst op sociale vernieuwing in drie gemeenten.* Amsterdam: Thesis Publishers.
37) Andreas Flache (1996), *The Double Edge of Networks. An Analysis of the Effect of Informal Networks on Cooperation in Social Dilemmas.* Amsterdam: Thesis Publishers.
38) Kees van Veen (1997), *Inside an Internal Labor Market: Formal Rules, Flexibility and Career Lines in a Dutch Manufacturing Company.* Amsterdam: Thesis Publishers.
39) Lucienne van Eijk (1997), *Activity and Well-being in the Elderly.* Amsterdam: Thesis Publishers.
40) Róbert Gál (1997), *Unreliability. Contract Discipline and Contract Governance under Economic Transition.* Amsterdam: Thesis Publishers
41) Anne-Geerte van de Goor (1997), *Effects of Regulation on Disability Duration.* ICS dissertation, Utrecht
42) Boris Blumberg (1997), *Das Management von Technologiekooperationen. Partnersuche und Verhandlungen mit dem Partner aus Empirisch-Theoretischer Perspektive.* ICS dissertation, Utrecht.
43) Marijke von Bergh (1997), *Loopbanen van oudere werknemers.* Amsterdam: Thesis Publishers.
44) Anna Petra Nieboer (1997), *Life-Events and Well-Being: A Prospective Study on Changes in Well-Being of Elderly People Due to a Serious Illness Event or Death of the Spouse.* Amsterdam: Thesis Publishers.
45) Jacques Niehof (1997), *Resources and Social Reproduction: The Effects of Cultural and Material Resources on Educational and Occupational Careers in Industrial Nations at the End of the Twentieth Century.* ICS dissertation, Nijmegen.
46) Ariana Need (1997), *The Kindred Vote. Individual and Family Effects of Social Class and Religion on Electoral Change in the Netherlands, 1956-1994.* ICS dissertation, Nijmegen.
47) Jim Allen (1997), *Sector Composition and the Effect of Education on Wages: an International Comparison.* Amsterdam: Thesis Publishers.
48) Jack B.F. Hutten (1998), *Workload and Provision of Care in General Practice. An Empirical Study of the Relation Between Workload of Dutch General Practitioners and the Content and Quality of their Care.* ICS dissertation, Utrecht.
49) Per B. Kropp (1998), *Berufserfolg im Transformationsprozeß. Eine theoretisch-empirische Studie über die Gewinner und Verlierer der Wende in Ostdeutschland.* ICS dissertation, Utrecht.
50) Maarten H.J. Wolbers (1998), *Diploma-inflatie en verdringing op de arbeidsmarkt. Een studie naar ontwikkelingen in de opbrengsten van diploma's in Nederland.* ICS

dissertation, Nijmegen.
51) Wilma Smeenk (1998), *Opportunity and Marriage. The Impact of Individual Resources and Marriage Market Structure on First Marriage Timing and Partner Choice in the Netherlands.* ICS dissertation, Nijmegen.
52) Marinus Spreen (1999), *Sampling Personal Network Structures: Statistical Inference in Ego-Graphs.* ICS dissertation, Groningen.
53) Vincent Buskens (1999), *Social Networks and Trust.* ICS dissertation, Utrecht.
54) Susanne Rijken (1999), *Educational Expansion and Status Attainment. A Cross-National and Over-Time Comparison.* ICS dissertation, Utrecht.
55) Mérove Gijsberts (1999), *The Legitimation of Inequality in State-Socialist and Market Societies, 1987-1996*. ICS dissertation, Utrecht.
56) Gerhard G. Van de Bunt (1999), *Friends by Choice. An Actor-Oriented Statistical Network Model for Friendship Networks Through Time*. ICS dissertation, Groningen.
57) Robert Thomson (1999), *The Party Mandate: Election Pledges and Government Actions in the Netherlands, 1986-1998.* Amsterdam: Thela Thesis.
58) Corine Baarda (1999), *Politieke besluiten en boeren beslissingen. Het draagvlak van het mestbeleid tot 2000.* ICS dissertation, Groningen.
59) Rafael Wittek (1999), *Interdependence and Informal Control in Organizations.* ICS dissertation, Groningen.
60) Diane Payne (1999), *Policy Making in the European Union: an Analysis of the Impact of the Reform of the Structural Funds in Ireland.* ICS dissertation, Groningen.
61) René Veenstra (1999), *Leerlingen - Klassen - Scholen. Prestaties en vorderingen van leerlingen in het voortgezet onderwijs*. Amsterdam, Thela Thesis.
62) Marjolein Achterkamp (1999), *Influence Strategies in Collective Decision Making. A Comparison of Two Models.* ICS dissertation, Groningen.
63) Peter Mühlau (2000), *The Governance of the Employment Relation. A Relational Signaling Perspective.* ICS dissertation, Groningen.
64) Agnes Akkerman (2000), *Verdeelde vakbeweging en stakingen. Concurrentie om leden*. ICS dissertation, Groningen.
65) Sandra van Thiel (2000), *Quangocratization: Trends, Causes and Consequences.* ICS dissertation, Utrecht.
66) Rudi Turksema (2000), *Supply of Day Care*. ICS dissertation, Utrecht.
67) Sylvia E. Korupp (2000), *Mothers and the Process of Social Stratification*. ICS dissertation, Utrecht.
68) Bernard A. Nijstad (2000), *How the Group Affects the Mind: Effects of Communication in Idea Generating Groups*. ICS dissertation, Utrecht.
69) Inge F. de Wolf (2000), *Opleidingsspecialisatie en arbeidsmarktsucces van sociale wetenschappers*. ICS dissertation, Utrecht.
70) Jan Kratzer (2001), *Communication and Performance: An Empirical Study in Innovation Teams*. ICS dissertation, Groningen.
71) Madelon Kroneman (2001), *Healthcare Systems and Hospital Bed Use*. ICS/NIVEL dissertation, Utrecht.
72) Herman van de Werfhorst (2001), *Field of Study and Social Inequality. Four Types of*

Educational Resources in the Process of Stratification in the Netherlands. ICS dissertation, Nijmegen.

73) Tamás Bartus (2001), *Social Capital and Earnings Inequalities. The Role of Informal Job Search in Hungary*. ICS dissertation, Groningen.
74) Hester Moerbeek (2001), *Friends and Foes in the Occupational Career. The Influence of Sweet and Sour Social Capital on the Labour Market*. ICS dissertation, Nijmegen.
75) Marcel van Assen (2001), *Essays on Actor Perspectives in Exchange Networks and Social Dilemmas.* ICS dissertation, Groningen.
76) Inge Sieben (2001), *Sibling Similarities and Social Stratification. The Impact of Family Background across Countries and Cohorts*. ICS dissertation, Nijmegen.
77) Alinda van Bruggen (2001), *Individual Production of Social Well-Being. An Exploratory Study.* ICS dissertation, Groningen.
78) Marcel Coenders (2001), *Nationalistic Attitudes and Ethnic Exclusionism in a Comparative Perspective: An Empirical Study of Attitudes Toward the Country and Ethnic Immigrants in 22 Countries*. ICS dissertation, Nijmegen.
79) Marcel Lubbers (2001), *Exclusionistic Electorates. Extreme Right-Wing Voting in Western Europe*, ICS dissertation, Nijmegen.
80) Uwe Matzat (2001), *Social Networks and Cooperation in Electronic Communities. A theoretical-empirical Analysis of Academic Communication and Internet Discussion Groups*, ICS dissertation, Groningen.
81) Jacques Janssen (2002), *Do Opposites Attract Divorce? Dimensions of Mixed Marriage and the Risk of Divorce in the Netherlands*, ICS dissertation, Nijmegen.
82) Miranda Jansen (2002), *Waardenoriënties en partnerrelaties. Een panelstudie naar wederzijdse invloeden*, ICS dissertation, Utrecht.
83) Anne Rigt Poortman (2002), *Socioeconomic Causes and Consequences of Divorce*. ICS dissertation, Utrecht.
84) Alexander Gattig (2002), *Intertemporal Decision Making*. ICS dissertation, Groningen.
85) Gerrit Rooks (2002), *Contract en Conflict. Strategisch Management van Inkooptransacties.* ICS dissertation, Utrecht.
86) Károly Takács (2002), *Social Networks and Intergroup Conflict.* ICS dissertation, Groningen.

Summary of notations

notation	meaning	special assumptions
payoff parameters		
v	victory payoff (temptation)	$v>e>0>c>d$
e	payoff for free-riding (bonus)	$v>e>0>c>d$
p	peace payoff	$p=0$
c	clash payoff (punishment, draw)	$v>e>0>c>d$
d	payoff for defeat (the sucker's payoff)	$v>e>0>c>d$
s	selective incentives	$s>0$
b	behavioral confirmation payoff	$b>0$
t	traitor's payoff	$t>0$
structural parameters and indexes		
$\{A, B\}$	the set of groups	$A\cap B=\emptyset$
n_A, n_B	group sizes	$n_A \geq 2, n_B \geq 2$
α, β	proportional group sizes	$\alpha=n_A/n_{A+B}, \beta= 1-\alpha$
f_i	number of fellow neighbors of individual i	
g_i	number of neighbors of i from the other group	
k_A^*, k_B^*	minimal contributing sets for group collective action	$0\leq k_A^* \leq n_A, 0\leq k_B^* \leq n_B$
simulations		
S	grid size	$S\geq n_{A+B}, S=RC\geq 6$
R, C	number of rows and columns in the grid	$R>0, C>0$
π	proportion of inhabited cells	$\pi=n_{A+B}/S$
T	total number of dyadic connections in the grid	
δ	the proportion of network relations (nonempty dyads)	
ϕ	the proportion of fellow ties (clustering, segregation)	
γ	the proportion of opposite ties (exposure)	$\gamma=1-\phi$
ρ_A	the proportion of ties between members of group A	$\rho_A+\rho_B=\phi$
decisions		
ω_A, ω_B	vector of strategy choices of group A and B (*dimensions* $n_A\times 1$ *and* $n_B\times 1$)	$\omega_i=1$ *if i contributes*
k_A, k_B	number of contributors in the groups	$0\leq k_A\leq n_A, 0\leq k_B \leq n_B$
notations in the experimental analysis		
r	round number	
$P_{rix}(\omega_i=1)$	probability of contribution in decision round r of actor i in experimental session x	
P_{rix}	contribution propensity in decision round r of actor i in experimental session x	$P_{rix} = \ln\left(\frac{P_{rix}(\omega_i = 1)}{P_{rix}(\omega_i = 0)}\right)$
α_0	baseline contribution propensity	
φ_x	session level error term	$\varphi_x \sim \mathrm{N}(0, \rho^2)$
ε_{ix}	subject level error term	$\varepsilon_{ix} \sim \mathrm{N}(0, \sigma^2)$
ξ_{rix}	intra-individual variation	