Social networks and intergroup conflict

Takács, Károly

"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
previous outcome	*peace* (*p*)	- - ?	- - -	- - ?	- - -	+ - ?	+ - -	+ - ?	+ - -
	defeat (*d*)	- + -	- + +	- + -	- + +	+ + -	+ + +	+ + -	+ + +
	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

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

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:

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

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:

$$P_{rix} = \alpha_0 + \beta'_S S_{rix} + \varphi_x + \varepsilon_{ix} + \xi_{rix}. \qquad (4.A.3)$$

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:

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

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:

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

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:

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

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

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

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.