

University of Groningen

Social networks and intergroup conflict

Takács, Károly

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2002

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Takács, K. (2002). *Social networks and intergroup conflict*. [Thesis fully internal (DIV), University of Groningen]. [S.n.].

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

“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 CRITICALNESS, REINFORCEMENT AND RECIPROCITY: | |
| Experimental evidence of behavior in repeated situations | 141 |
| 5.1 Introduction | 143 |
| 5.2 Design of the repeated IPG game experiments | 144 |
| 5.2.1 Subjects | 144 |
| 5.2.2 Procedure | 144 |
| 5.3 Results of the intergroup competition | 145 |
| 5.3.1 Results by structural conditions | 145 |
| 5.3.2 Conflict over time | 148 |
| 5.3.3 Contribution rates over time | 152 |
| 5.4 Conditional responses | 157 |
| 5.4.1 Criticalness, reinforcement learning, and intergroup reciprocity | 157 |
| 5.4.2 Local reciprocity | 162 |
| 5.4.3 Interactions of structure and time | 164 |
| 5.5 Questionnaire data | 169 |
| 5.6 Results of the multilevel analyses | 172 |
| 5.6.1 The effect of decision heuristics | 173 |
| 5.6.2 The effect of personal characteristics and other control variables | 177 |
| 5.6.3 Interaction effects | 184 |
| 5.7 How beliefs coincide with reality | 190 |
| 5.8 Summary of results | 197 |

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

| | <i>Structural embeddedness</i> | <i>Temporal embeddedness</i> | <i>Interactions of structure and time</i> |
|---|--------------------------------|----------------------------------|---|
| Macro hypotheses <i>dependent variable:</i> outcome(s) of the game | Section 5.3.1 | Sections 5.3.2, 5.3.3 | Section 5.3.3 |
| Micro hypotheses <i>dependent variable:</i> 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

| <i>clustering condition in the experiment</i> | <i>outcome of the decision round</i> | | Total |
|---|--------------------------------------|---------------|--------------|
| | peace | conflict | |
| 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

| <i>incentives introduced first</i> | low clustering | medium clustering | high clustering | Total |
|--------------------------------------|-----------------------|--------------------------|------------------------|---------------|
| <i>Part I (control condition)*</i> | 45.57 (937) | 50.31 (1270) | 46.89 (1286) | 47.48 (3493) |
| <i>Part II (minimal contact)</i> | 43.93 (972) | 58.50 (1323) | 53.23 (1255) | 52.65 (3550) |
| <i>Part III b (confirmation pr.)</i> | - | 59.51 (699) | 51.15 (610) | 55.61 (1309) |
| <i>Part III s (normative press.)</i> | - | 67.68 (659) | 75.53 (617) | 71.47 (1276) |
| <i>Part III total</i> | 33.08 (925) | 63.48 (1358) | 63.41 (1227) | 55.44 (3510) |
| <i>Part IV b (confirmation pr.)</i> | - | 61.55 (632) | 71.65 (589) | 66.42 (1221) |
| <i>Part IV s (normative press.)</i> | - | 73.08 (650) | 88.62 (668) | 80.96 (1318) |
| <i>Part IV total</i> | 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.

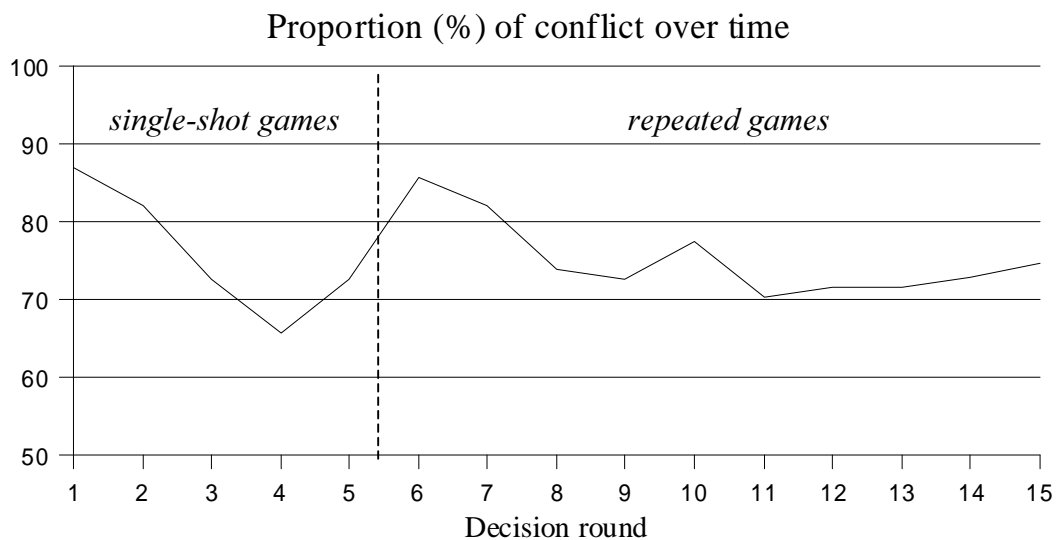


Figure 5.3.2.1 The likelihood of intergroup conflict over time aggregating all experimental parts and all experimental conditions

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.

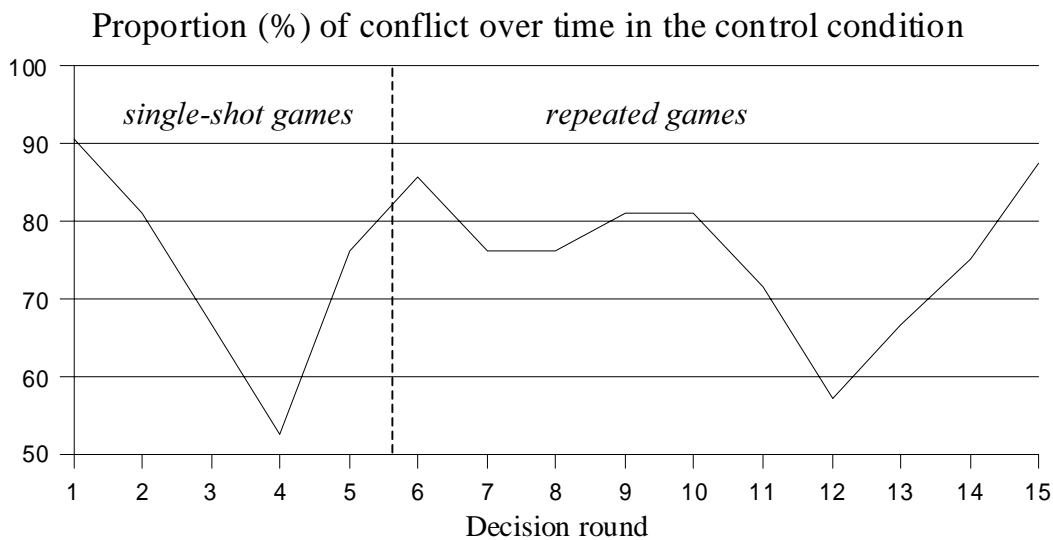


Figure 5.3.2.2 The likelihood of intergroup conflict in the control condition (Part I)

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 (<i>p</i>) | victory of the red team | victory of the green team | clash (<i>c</i>) | <i>conflict</i> <i>total</i> | Total |
|----------------------------------|---------|-----------------------|-------------------------------|---------------------------------|-----------------------|---------------------------------|----------------|
| peace (<i>p</i>) | | 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 (<i>c</i>) | | 23 (11.6%) | 60 (30.2%) | 70 (35.2%) | 46 (23.1%) | 176 (88.4%) | 199 (100%) |
| <i>conflict total</i> | | 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 (<i>p</i>) | victory of the red team | victory of the green team | clash (<i>c</i>) | <i>conflict total</i> | Total |
|----------------------------------|---------|-----------------------|-------------------------|---------------------------|-----------------------|-----------------------|---------------|
| peace (<i>p</i>) | | 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 (<i>c</i>) | | 6 (21.4%) | 6 (21.4%) | 13 (46.4%) | 3 (10.7%) | 22 (78.6%) | 28 (100%) |
| <i>conflict total</i> | | 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.

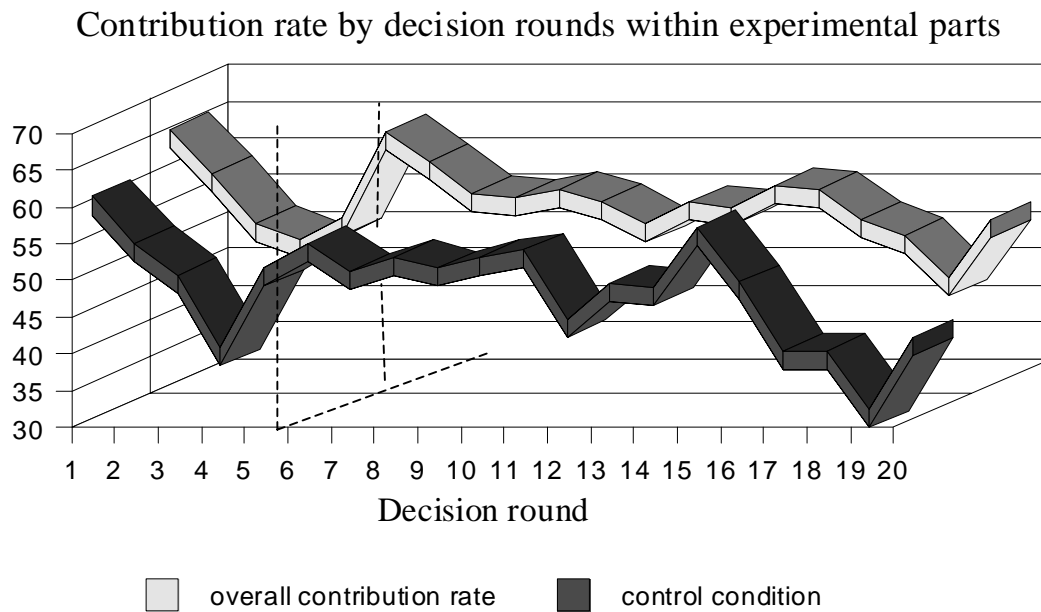


Figure 5.3.3.1 Contribution rates by decision rounds within parts

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.

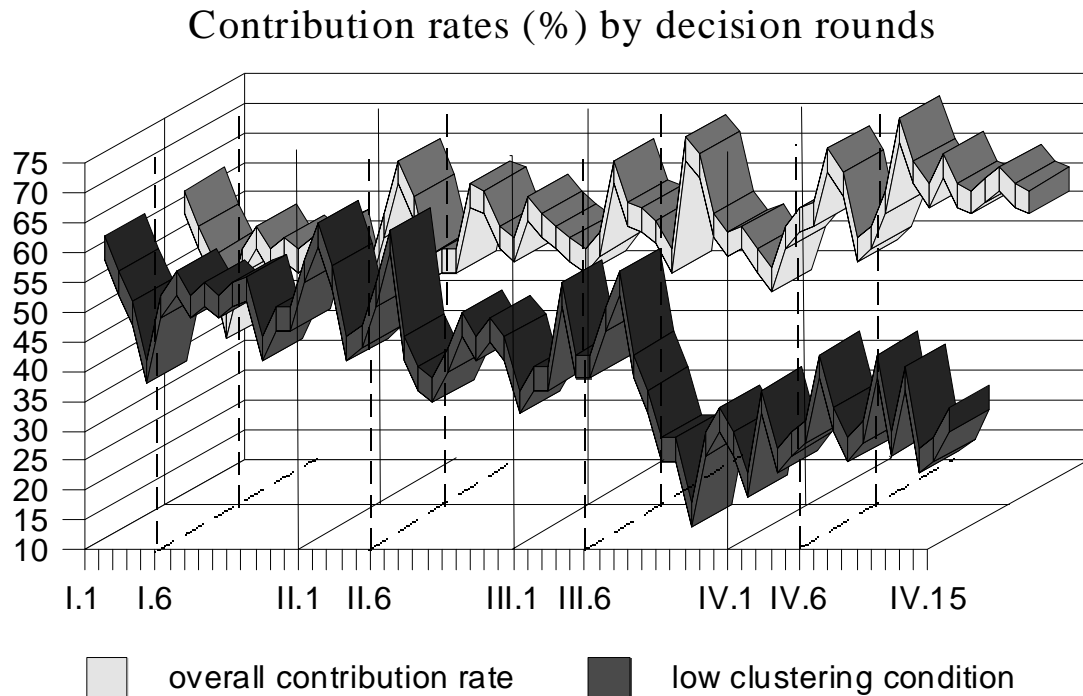


Figure 5.3.3.2 Contribution rates by decision rounds

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 parts*. 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

| <i>expectation of subject</i> | <i>previous outcome</i> | <i>criticalness hypothesis</i> | <i>contribution rate (from N cases)</i> |
|-------------------------------|-------------------------|--------------------------------|---|
| peace (p) | any | - | 9.52% (2919) |
| defeat (d) | any | - | 25.55% (1926) |
| victory (v) | <i>p, d, c</i> | + | 77.03% (4166) |
| | victory (v) | ? | 78.38% (2433) |
| | no previous round | ? | 70.85% (590) |
| <i>v total</i> | | | 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

| <i>prev. decision</i> | hypotheses | | contribution rates | | N | | total (C+D) | |
|-----------------------|------------|----|--------------------|--------|------|------|-------------|-------|
| | C | D | C | D | C | D | C rate | N |
| <i>p</i> | - ? | -- | 51.15% | 29.57% | 954 | 2831 | 35.01% | 3785 |
| <i>previous</i> | <i>d</i> | +- | 61.62% | 53.79% | 1597 | 2043 | 57.23% | 3640 |
| <i>outcome</i> | <i>v</i> | ?+ | 71.64% | 43.84% | 2870 | 771 | 65.75% | 3641 |
| | <i>c</i> | +- | 69.24% | 52.54% | 1453 | 512 | 64.89% | 1965 |
| <i>total</i> | | | 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, *pC*) 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 outcomes²

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

| | <i>Number of defecting neighbors in r-1</i> | | | <i>Total</i> |
|--|---|--------------|--------------|--------------|
| | <i>0</i> | <i>1</i> | <i>2</i> | |
| Number of contributing neighbors in <i>r-1</i> | <i>0</i> 38.33 (120) | 37.03 (721) | 37.72 (1853) | 37.56 (2694) |
| | <i>1</i> 60.65 (958) | 56.27 (3160) | - | 57.29 (4118) |
| | <i>2</i> 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

| | <i>Number of defecting fellow nbs in r-1</i> | | | <i>Total</i> |
|---|--|--------------|-------------|--------------|
| | <i>0</i> | <i>1</i> | <i>2</i> | |
| Number of contributing fellow nbs in <i>r-1</i> | <i>0</i> 35.05 (3347) | 55.52 (1403) | 42.73 (344) | 41.21 (5094) |
| | <i>1</i> 71.65 (2744) | 64.91 (778) | - | 70.16 (3522) |
| | <i>2</i> 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

| | <i>N of defecting other group neighbors in r-1</i> | | | <i>Total</i> |
|--|--|--------------|--------------|--------------|
| | <i>0</i> | <i>1</i> | <i>2</i> | |
| N of contributing neighbors from the other group in <i>r-1</i> | <i>0</i> 66.98 (2895) | 56.30 (1746) | 28.98 (1049) | 56.70 (5690) |
| | <i>1</i> 64.69 (2634) | 31.35 (874) | - | 56.39 (3508) |
| | <i>2</i> 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 *dC*, *vC*, and *cC*), 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.

| prev. dec. | subject's expectation | | | | | | | | | | |
|-------------------------|-----------------------|----------------|---------------------|----------------|----------------------|-----------------|--------------------|----------------|----------------|-----------------|-----------------|
| | peace (<i>p</i>) | | defeat (<i>d</i>) | | victory (<i>v</i>) | | clash (<i>c</i>) | | total | | |
| | C | D | C | D | C | D | C | D | C | D | |
| <i>previous outcome</i> | <i>p</i> | 18.75 (64) | 4.15 (241) | 13.33 (15) | 10.53 (38) | 71.60 (81) | 68.51 (181) | 57.89 (19) | 30.00 (40) | 46.37 (179) | 30.00 (500) |
| | <i>d</i> | 0.00 (23) | 18.60 (43) | 27.03 (74) | 15.07 (73) | 75.22 (113) | 67.70 (226) | 63.33 (30) | 56.10 (82) | 51.67 (240) | 51.42 (424) |
| | <i>v</i> | 17.78 (45) | 4.76 (21) | 11.36 (44) | 6.67 (15) | 69.14 (324) | 53.21 (109) | 51.43 (70) | 40.63 (32) | 56.52 (483) | 41.24 (177) |
| | <i>c</i> | 8.70 (23) | 20.00 (10) | 25.71 (35) | 0.00 (9) | 72.45 (98) | 59.32 (59) | 62.50 (24) | 20.00 (15) | 53.89 (180) | 43.01 (93) |
| <i>total</i> | | 14.19 (155) | 6.67 (315) | 21.43 (168) | 11.85 (135) | 71.10 (616) | 64.35 (575) | 56.64 (143) | 43.79 (169) | 53.33 (1082) | 40.28 (1194) |
| | | 9.15 (470) | | 17.16 (303) | | 67.84 (1191) | | 49.68 (312) | | 46.49 (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.

| prev. dec. | subject's expectation | | | | | | | | | | |
|-------------------------|-----------------------|----------------|---------------------|----------------|----------------------|-----------------|--------------------|----------------|----------------|-----------------|-----------------|
| | peace (<i>p</i>) | | defeat (<i>d</i>) | | victory (<i>v</i>) | | clash (<i>c</i>) | | total | | |
| | C | D | C | D | C | D | C | D | C | D | |
| <i>previous outcome</i> | <i>p</i> | 11.83 (169) | 3.32 (905) | 11.54 (26) | 11.11 (72) | 72.46 (138) | 69.06 (320) | 44.00 (25) | 40.66 (91) | 37.43 (358) | 21.33 (1388) |
| | <i>d</i> | 5.41 (37) | 5.97 (67) | 26.00 (50) | 16.28 (86) | 56.52 (69) | 76.24 (181) | 63.16 (19) | 52.70 (74) | 37.71 (175) | 47.79 (408) |
| | <i>v</i> | 7.46 (67) | 2.78 (36) | 30.43 (46) | 6.67 (15) | 71.91 (235) | 69.74 (76) | 65.43 (81) | 43.48 (23) | 56.18 (429) | 43.33 (150) |
| | <i>c</i> | 0.00 (15) | 0.00 (10) | 18.18 (11) | 0.00 (4) | 65.12 (43) | 68.18 (22) | 36.36 (11) | 57.14 (7) | 42.50 (80) | 44.19 (43) |
| <i>total</i> | | 9.38 (288) | 3.44 (1018) | 24.06 (133) | 12.99 (177) | 69.28 (485) | 71.29 (599) | 58.82 (136) | 46.15 (195) | 45.59 (1042) | 28.91 (1989) |
| | | 4.75 (1306) | | 17.74 (310) | | 70.39 (1084) | | 51.36 (331) | | 34.64 (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.

| prev. dec. | | subject's expectation | | | | | | | | | |
|------------------|----------|-----------------------|---------------|---------------------|----------------|----------------------|----------------|--------------------|----------------|-----------------|-----------------|
| | | peace (<i>p</i>) | | defeat (<i>d</i>) | | victory (<i>v</i>) | | clash (<i>c</i>) | | total | |
| | | C | D | C | D | C | D | C | D | C | D |
| previous outcome | <i>p</i> | 28.26 (46) | 5.48 (146) | 41.67 (12) | 22.22 (72) | 79.56 (137) | 62.79 (215) | 57.14 (28) | 49.15 (59) | 64.13 (223) | 38.21 (492) |
| | <i>d</i> | 7.50 (40) | 15.00 (60) | 43.90 (82) | 23.36 (107) | 85.62 (146) | 79.90 (209) | 68.75 (48) | 44.00 (50) | 62.34 (316) | 52.35 (426) |
| | <i>v</i> | 20.00 (45) | 13.33 (30) | 16.67 (48) | 9.38 (32) | 76.87 (402) | 65.66 (99) | 60.71 (56) | 72.41 (29) | 65.34 (551) | 48.95 (190) |
| | <i>c</i> | 18.18 (22) | 16.67 (12) | 17.14 (35) | 37.50 (16) | 79.61 (152) | 78.57 (70) | 68.29 (41) | 61.54 (13) | 63.60 (250) | 63.96 (111) |
| <i>total</i> | | 18.95 (153) | 9.27 (248) | 31.07 (177) | 22.03 (227) | 79.33 (837) | 71.16 (593) | 64.16 (173) | 52.98 (151) | 64.10 (1340) | 47.17 (1219) |
| | | 12.97 (401) | | 25.99 (404) | | 75.94 (1430) | | 58.95 (324) | | 56.04 (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.

| prev. dec. | | subject's expectation | | | | | | | | | |
|------------------|----------|-----------------------|----------------|---------------------|----------------|----------------------|----------------|--------------------|----------------|-----------------|-----------------|
| | | peace (<i>p</i>) | | defeat (<i>d</i>) | | victory (<i>v</i>) | | clash (<i>c</i>) | | total | |
| | | C | D | C | D | C | D | C | D | C | D |
| previous outcome | <i>p</i> | 21.87 (32) | 11.19 (143) | 17.65 (17) | 16.67 (54) | 86.09 (115) | 78.13 (192) | 69.57 (23) | 52.08 (48) | 66.84 (187) | 45.77 (437) |
| | <i>d</i> | 23.88 (67) | 14.29 (91) | 42.48 (226) | 27.19 (217) | 84.83 (402) | 85.27 (353) | 86.34 (161) | 73.64 (110) | 69.16 (856) | 58.88 (771) |
| | <i>v</i> | 21.57 (51) | 11.11 (27) | 43.28 (67) | 12.50 (32) | 90.95 (1039) | 57.14 (147) | 78.30 (235) | 30.56 (36) | 83.98 (1392) | 42.15 (242) |
| | <i>c</i> | 26.76 (71) | 17.86 (28) | 39.50 (119) | 24.24 (33) | 88.57 (490) | 68.18 (154) | 83.27 (251) | 40.82 (49) | 76.15 (931) | 52.27 (264) |
| <i>total</i> | | 23.98 (221) | 12.80 (289) | 40.79 (429) | 23.81 (336) | 88.91 (2046) | 75.65 (846) | 81.79 (670) | 56.38 (243) | 77.09 (3366) | 52.16 (1714) |
| | | 17.65 (510) | | 33.33 (765) | | 85.03 (2892) | | 75.03 (913) | | 68.68 (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.

| <i>Categories of intentions</i> (I wanted ...) | <i>Major intention</i> | | <i>Ranks of categories</i> | | <i>Factor loadings</i> | | |
|---|------------------------|------|----------------------------|--------|------------------------|-------|-------|
| | N | % | mean | median | 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 |
| <i>to initiate and break the deadlock</i> | 1 | 0.5 | - | - | - | - | - |
| <i>to protest against others' behavior</i> | 1 | 0.5 | - | - | - | - | - |
| <i>an outcome to fulfill my forecast</i> | 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 $\alpha=0.56$. 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 $\alpha=0.65$. 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.

| <i>behavioral mechanism</i> | <i>conditional strategy asked</i> | <i>%</i> |
|-----------------------------|---|----------|
| 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 |
| intergroup reciprocity | defeat-contribute | 73.6 |
| | 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 |
|--|--|----------------------------|------------------------------|
| <i>FIXED EFFECTS</i> | | | |
| α_0 baseline contribution propensity | ? | .745*** (.082) | .749*** (.083) |
| <i>STRUCTURAL EMBEDDEDNESS</i> | | | |
| 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) |
| <i>TEMPORAL EMBEDDEDNESS</i> | | | |
| <i>Criticalness</i> | | | |
| 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) |
| <i>Intergroup reciprocity</i> | | | |
| 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) |
| <i>Reinforcement learning</i> | | | |
| predicted increase of contribution propensity | + | -.110 (.068) | -.135 (.073) |
| predicted decrease of contribution propensity | - | -.212*** (.059) | -.181** (.062) |
| <i>Local reciprocity</i> | | | |
| 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) |
| <i>RANDOM EFFECTS</i> | | | |
| 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 vs. previous model | 8037.7*** (17) | 160*** (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 |
|--|--|------------------------------------|--------------------------|
| <i>FIXED EFFECTS</i> | | | |
| α_0 baseline contribution propensity | ? | .886* (.361) | .737* (.375) |
| <i>STRUCTURAL EMBEDDEDNESS</i> | | | |
| 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) |
| <i>TEMPORAL EMBEDDEDNESS</i> | | | |
| <i>Criticalness</i> | | | |
| 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) |
| <i>Intergroup reciprocity</i> | | | |
| 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) |
| <i>Reinforcement learning</i> | | | |
| predicted increase of contribution propensity | + | -.114 (.068) | -.142 (.074) |
| predicted decrease of contribution propensity | - | -.215*** (.059) | -.186** (.062) |
| <i>Local reciprocity</i> | | | |
| 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 $r-l$) | | .193 ⁺⁺⁺ (.071) |
| variance of p peace in $r-l$ | | .000 (.000) |
| variance of d defeat in $r-l$ | | .000 (.000) |
| variance of c clash in $r-l$ | | .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 |
|--|--|------------------------------------|------------------------------|
| <i>FIXED EFFECTS</i> | | | |
| 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 | | | |
| <i>Criticalness</i> | | | |
| 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) |
| <i>Intergroup reciprocity</i> | | | |
| 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) |
| <i>Reinforcement learning</i> | | | |
| predicted increase of contribution propensity | + | -.050 (.072) | -.086 (.077) |
| predicted decrease of contribution propensity | - | -.167** (.061) | -.142* (.063) |
| <i>Local reciprocity</i> | | | |
| 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, <i>s</i> , medium clustering | -.006 (.015) | -.010 (.016) |
| within Part III trend, <i>s</i> , high clustering | -.018 (.019) | -.018 (.020) |
| within Part III trend, <i>b</i> , medium clustering | -.023 (.014) | -.034* (.015) |
| within Part III trend, <i>b</i> , 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_l direct selective incentives | | .117 (.068) |
| variance of b_0 internalized behavioral confirmation | | .070 ⁺⁺⁺ (.030) |
| variance of b_l 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 $r-l$) | | .195 ⁺⁺⁺ (.071) |
| variance of p peace in $r-l$ | | .000 (.000) |
| variance of d defeat in $r-l$ | | .003 (.055) |
| variance of c clash in $r-l$ | | .000 (.000) |
| variance of positive reinforcement | | .102 ⁺ (.064) |
| variance of negative reinforcement | | .029 ⁺⁺ (.044) |
| variance of local reciprocity t_l 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 proselves*. Subjects with prosocial orientation were strongly influenced by the

previous decision of fellow neighbors, not like proselves, 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 proselves 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 |
|--|--|--|---|
| <i>FIXED EFFECTS</i> | | | |
| α_0 baseline contribution propensity | ? | 1.217** (.413) | 1.133** (.434) |
| <i>STRUCTURAL EMBEDDEDNESS</i> | | | |
| 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) |
| <i>TEMPORAL EMBEDDEDNESS</i> | | | |
| <i>Criticalness</i> | | | |
| 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 - I) | + | .252** (.093) | .263** (.102) |
| <i>Intergroup reciprocity</i> | | | |
| p peace in r - I | - | -.012 (.111) | -.018 (.115) |
| d defeat in r - I | + | .224* (.107) | .216 (.112) |
| c clash in r - I | + | -.103 (.117) | -.115 (.123) |
| <i>Reinforcement learning</i> | | | |
| predicted increase of contribution propensity | + | -.062 (.072) | -.094 (.078) |
| predicted decrease of contribution propensity | - | -.168** (.061) | -.137* (.064) |
| <i>Local reciprocity</i> | | | |
| 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, <i>s</i> , medium clustering | -.014 (.016) | -.020 (.017) |
| within Part III trend, <i>s</i> , high clustering | -.025 (.020) | -.024 (.021) |
| within Part III trend, <i>b</i> , medium clustering | -.033* (.015) | -.045** (.016) |
| within Part III trend, <i>b</i> , 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 | | |
| <i>Interactions with gender</i> | | |
| 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 <i>b</i> is introduced | -.266* (.134) | -.183 (.140) |
| <i>Interactions with social control</i> | | |
| 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

| <i>Interactions with local reciprocity</i> | | |
|---|----------------------------|----------------------------|
| 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) |
| <i>RANDOM EFFECTS</i> | | |
| 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 $r-l$) | | .213 ⁺⁺⁺ (.074) |
| variance of p peace in $r-l$ | | .000 (.000) |
| variance of d defeat in $r-l$ | | .000 (.000) |
| variance of c clash in $r-l$ | | .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

| | <i>prev.</i> <i>dec.</i> | subject's expectation | | | | total |
|---------------------|-----------------------------|-----------------------|---------------------|----------------------|--------------------|-----------------|
| | | peace (<i>p</i>) | defeat (<i>d</i>) | victory (<i>v</i>) | clash (<i>c</i>) | |
| <i>p</i> | 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) |
| <i>d</i> | 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) |
| <i>v</i> | 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) |
| <i>c</i> | 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) |
| <i>total</i> | 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) |
| <i>total</i> | | 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 (<i>p</i>) | defeat (<i>d</i>) | victory (<i>v</i>) | clash (<i>c</i>) | total |
|--|--------------------|---------------------|----------------------|--------------------|---------|
| <i>prosocial subjects</i> | 20.3% | 12.6% | 51.0% | 16.1% | 100% |
| | (1091) | (680) | (2745) | (867) | (5383) |
| <i>others (proselfs or not consistent)</i> | 20.9% | 14.1% | 50.8% | 14.2% | 100% |
| | (1808) | (1223) | (4407) | (1233) | (8671) |
| <i>total</i> | 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 |
|--|--|--|---|
| <i>FIXED EFFECTS</i> | | | |
| <i>Decision-level variables</i> | | | |
| (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) |
| <i>Personal characteristics and other subject-level variables</i> | | | |
| 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) |
| <i>RANDOM EFFECT</i> | | | |
| 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 | <i>fulfilling expectations</i> | <i>criticalness and bandwagon</i> |
|-----------------------|---|--------------------------------|-----------------------------------|
| peace (<i>p</i>) | any | - | - |
| defeat (<i>d</i>) | any | - | - |
| victory (<i>v</i>) | any | + | + |
| clash (<i>c</i>) | peace (<i>p</i>) or defeat (<i>d</i>) | + | + |
| | victory (<i>v</i>) | - | + |
| | clash and contribution (<i>cC</i>) | + | + |
| | clash and defection (<i>cD</i>) | - | + |

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

| previous decision | <i>fulfilling expectations</i> | | <i>criticalness and bandwagon</i> | | contribution rates | |
|----------------------------|--------------------------------|---|-----------------------------------|---|--------------------|-----------------|
| | C | D | C | D | C | D |
| <i>peace (p)</i> | + | + | + | + | 56.84% (95) | 43.28% (238) |
| <i>previous defeat (d)</i> | + | + | + | + | 78.68% (258) | 59.49% (316) |
| <i>outcome victory (v)</i> | - | - | + | + | 69.46% (442) | 45.83% (120) |
| <i>clash (c)</i> | + | - | + | + | 78.29% (327) | 41.67% (84) |
| <i>total</i> | | | | | 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 *vC* 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 *vD* 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 *cD* 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 unflinching 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*)

| | <i>Structural embeddedness</i> | <i>Temporal embeddedness</i> | <i>Interactions of structure and time</i> |
|--|---|--|---|
| Macro hypotheses <i>dependent variable:</i> outcome(s) of the game | <i>H:</i> Conflict is more likely in segregated structures. <i>R:</i> Partly confirmed. Conflict was least likely in the low clustering condition. Both medium and high clustering has high likelihood of conflict. | <i>H:</i> Stable peace. <i>R:</i> Confirmed. <i>H:</i> Durable conflict. <i>R:</i> Largely supported. <i>H:</i> Spiral of peace. <i>R:</i> Largely supported. <i>H:</i> Spiral of conflict. <i>R:</i> Not supported. | <i>H:</i> 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. <i>R:</i> Partly supported. |
| Micro hypotheses <i>dependent variable:</i> individual decision(s) | <i>H:</i> Subjects are influenced by internalized and direct forms of social control. <i>R:</i> Largely confirmed. Subjects are influenced by internalized and direct forms of social control from fellow neighbors and conditionally by traitor rewards. | <i>H:</i> Subjects follow (a combination of) simple behavioral rules. <i>R:</i> 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. | <i>H:</i> Subjects reciprocate previous behavior of their neighbors. <i>R:</i> Conditionally supported. <i>H:</i> There is no direct effect of structure on which behavioral rules are followed. <i>R:</i> 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 proselves. On the other hand, the introduction of monetary behavioral confirmation elicited less additional reciprocity from prosocials than from proselves. 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.