

WORKING PAPERS

Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon

Wieland Müller
Fangfang Tan

September 2011

Working Paper No: 1111



DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

All our working papers are available at: <http://mailbox.univie.ac.at/papers.econ>

Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon*

Wieland Müller[†]

VCEE & University of Vienna,

CentER, TILEC & Tilburg University

Fangfang Tan[‡]

Max Planck Institute

for Tax Law and Public Finance

September 30, 2011

Abstract

Previous experimental results on one-shot sequential two-player games show that group decisions are *closer* to the subgame-perfect Nash equilibrium than individual decisions. We extend the analysis of inter-group versus inter-individual decision making to a Stackelberg market game, by running both one-shot and repeated markets. Whereas in the one-shot markets we find no significant differences in the behavior of groups and individuals, we find that the behavior of groups is *further away* from the subgame-perfect equilibrium of the stage game than that of individuals. To a large extent, this result is independent of the method of eliciting choices (sequential or strategy method) and the method used to account for observed first- and second-mover behavior. We provide evidence on followers' response functions and electronic chats to offer an explanation for the differential effect that the time horizon of interaction has on the extent of individual and group players' (non)conformity with subgame perfectness.

JEL Classification numbers: C72, C92, L13.

Keywords: Stackelberg market, groups versus individuals, discontinuity effect, experiment.

*We are grateful to David Vonka for technical consulting and help. We thank Marco Castillo, Guillaume Fréchette, Daniel Houser, Rudolf Kerschbamer, Martin Kocher, Charles Noussair, Ragan Petrie, Jan Potters, Arno Riedl, Andrew Schotter, Jean-Robert Tyran and seminar participants at Innsbruck University, Tilburg University, New York University, George Mason University, the 3rd Maastricht Behavioral and Experimental Economics Symposium, the ESA world meeting 2010 in Copenhagen, and the Symposium on Industrial Organization and Management Strategy 2011 in Chengdu for helpful comments. Furthermore, we thank James C. Cox, Daniel Friedman, and Steven Gjerstad as well as Sau-Him Paul Lau and Felix Leung for making available estimation codes. Wieland Müller acknowledges financial support from the Netherlands Organisation for Scientific Research (NWO) through a VIDI grant.

[†]Department of Economics, University of Vienna, Brünnerstrasse 72, 1210 Vienna, Austria, Email: wieland.mueller@univie.ac.at.

[‡]Max Planck Institute for Tax Law and Public Finance, Marstallplatz 1, 80539, Munich, Germany, Email: fangfang.tan@tax.mpg.de.

1 Introduction

Many decisions in private, public, and business life are not taken by individuals, but by groups of individuals. Think, for instance, of households, public authorities, court juries, boards of directors, or management teams.¹ However, much of economic theory does not distinguish between decisions taken by individuals or groups. Also, until recently, experimental economists were mainly concerned with testing economic models using individuals as decision makers. Various authors rightly point out that in the presence of systematic differences in decisions made by individuals and groups, it would be risky to export results observed in interindividual decision making to domains where groups interact with each other (see, e.g., Cooper and Kagel, 2005).

Recently growing experimental research on interindividual-intergroup comparisons has so far derived the result that, indeed, often there are differences in the behavior of individuals and groups. More precisely, although there are exceptions, one result that emerges from the literature is that often groups appear to be more selfish than individuals. This has mainly been shown in the context of two classes of games. The first class consists of simple, sequential-move, two-player games such as the ultimatum game (Bornstein and Yaniv 1998, and Robert and Carnevale 1997), the trust game (Cox 2002, and Kugler et al. 2007), the centipede game (Bornstein et al. 2004), and the gift-exchange game (Kocher and Sutter 2007).² Bornstein (2008, p. 30) summarizes much of this literature by stating that

“Groups, it seems, are more selfish and more sophisticated players than individuals, and, as a result, interactions between two unitary groups are closer to the rational, game-theoretical solution than interactions between two individuals.”

Note that the literature Bornstein summarizes in this quote is based on experimental games in which individuals and groups interact only once. The second class consists of games that authors characterize as having a “Eureka” component, meaning that once the solution or equilibrium is found, it is recognized as a clear solution of the game. Based on results from, e.g., signaling games

¹For example, the chairman’s office of the News Corporation is a group of five persons meeting every week to consider “every acquisition and item of capital expenditure” (FT May 20th, 2003). More generally, the organization literature has a long tradition of analyzing the role of management teams in firms. As Finkelstein and Hambrick (1996) point out, decision makers are informed, influenced and sometimes constrained by others, both inside and outside the organization.

²One exception is provided by Cason and Mui’s (1997) dictator games where, in some cases, group dictators give more than individual dictators. In their re-examination, Luhan et al. (2009) report team dictators to be more selfish than individual dictators.

(Cooper and Kagel, 2005) and beauty contests (Kocher and Sutter, 2005), Sutter et al. (2009, p. 391) state that

“It can be considered a stylized fact in the literature that teams are generally closer to game-theoretic predictions than individuals in (interactive) games in which rationality and correct reasoning are the predominant task characteristics.”

Moreover, to the extent that groups and individuals converge to the same equilibrium in these repeated “Eureka”-type games, groups are found to do so much faster than individuals.

In this paper we contribute to the literature on interindividual-intergroup comparisons by studying a Stackelberg market game which, arguably, belongs to the first class of games above. A particular aim is to study the effect the time horizon of interaction has on the behavior of individuals and groups—a topic that has not yet been thoroughly studied in this class of games. Our results are in (partial) contrast to the quotes above. In fact, in our one-shot Stackelberg markets we find no significant differences in the behavior of groups and individuals, and in our repeated Stackelberg markets we find that the behavior of groups is *further away* from the subgame-perfect equilibrium than that of individuals. That is, we show that once a simple sequential-move game (belonging to the class of games summarized by Bornstein, 2008) is repeated, the behavior of groups relative to that of individuals goes in the opposite direction to what is stated in Bornstein’s summary. In particular, group play *diverges* from the (refined) game-theoretic solution.

The Stackelberg (1934) model is among the most frequently applied models of oligopolistic competition. In a Stackelberg duopoly market game, one firm (the first mover) makes its quantity decision first. Then, knowing the first mover’s choice, the other firm (the second mover) decides on its quantity, before the market clears. In case of linear market demand and symmetric and constant marginal costs, in the subgame perfect equilibrium the first mover produces and earns twice as much as the second mover. Moreover, the second mover’s best response is a linear and downward sloping function of the leader’s quantity choice.³ We chose a Stackelberg game because it has a very attractive feature: For each of the first mover’s quantity choice, a second mover can, by its own quantity choice, express a wide range of preferences over own and the other player’s income.⁴

³Experimental evidence on individual-player Stackelberg duopoly markets and how they compare to simultaneous-move Cournot duopoly markets is reported in Huck et al. (2001).

⁴This feature distinguishes the Stackelberg game from other sequential games such as the ultimatum game or the trust game.

We implement this market game both as one-period and as multiple-period games by having either individuals or groups of three subjects act in the role of the first and the second mover. Subjects acting in groups have to agree unanimously on the quantity produced. The decision-making process within groups is aided by access to a chat tool. The members of a group are able to exchange written messages until they reach a joint decision.

Comparing first mover quantities across treatments is straightforward. In the one-period games we find that although the average group leader quantity is somewhat higher than the average individual leader quantity, the difference is insignificant. In the multiple-period games, in contrast, we find that average leader quantities chosen by groups are significantly lower than average leader quantities chosen by individuals. Comparing second mover behavior across treatments is less straightforward as we observe followers' choices in response to varying first mover choices. Nevertheless, for the one-period games we find that, if anything, the observed average response function of groups is closer to the best-response function than that of individuals, which is in line with earlier experimental results. But, again, we fail to detect statistical differences. In the multiple-period game treatments, average observed reaction functions of followers display a specific non-monotonic pattern not predicted by standard theory. However, this pattern is predicted and can be accounted for by models of other-regarding preferences. We use maximum-likelihood techniques to estimate average follower response functions for the multiple-period treatments, using either Lau and Leung's (2010) implementation of the Fehr and Schmidt (1999) model of inequality aversion or the Cox et al. (2007) model of emotion-driven reciprocity. As the standard rational best response function of followers is nested in both of these models, we have a clear and unambiguous method to test which of two observed average response functions is closer to the prediction of subgame perfectness. Irrespective of which of the two models we use to account for followers' reaction functions, we find that the one employed by groups is further away from the rational best response function than that of individuals.

Since individuals and groups partly choose markedly different quantities as first movers, differences we observe in individual and group second-mover decisions might be driven by different experiences second movers make in the individual and the relevant group-player treatments. We control for this by also eliciting choices in four additional treatments employing the strategy method (Selten, 1967) in which, simultaneously with the first movers making their decision, the second movers have to indicate how they would react to each of the first movers' quantities. Thus, this method gives us the complete response function of second movers. The results of the control

treatments largely confirm the results obtained in the main treatments with truly sequential play. In the one-shot sessions, behavior appears to be in line with results reported in the literature as group leaders and followers are closer to the prediction of subgame perfectness, although the differences are insignificant. In the multiple-period treatments, we find, again, that in comparison to individuals, groups choose lower leader quantities and employ response functions that are further away from the rational best response function.

Our paper makes two main contributions. The literature reports so far that in the class of simple, two-player, sequential-move games groups often appear to be *closer* to the game-theoretic prediction than individuals if the game is played only once. We show for a game belonging to this class of games that once the game is repeated, the result is turned around in the sense that groups are shown to be *further away* from the game-theoretic prediction. The Stackelberg market game is, arguably, not a “Eureka”-type problem that has a clear solution, which, once found, is clearly seen as such by players. Instead, a Stackelberg duopoly market is a game that, like the other games summarized by Bornstein (2008), leave more room for other-regarding preferences. In these games, the presence of profit-maximizing and other-regarding motives might play out differently depending on whether the game is played by groups or by individuals and depending on the time horizon of interaction. In fact, to explain our results, in the discussion section we provide evidence that there is heterogeneity in subjects’ types. Concentrating on second movers, we find that they are often either *myopic profit maximizers* (who always best respond to a first mover’s quantity), *strategic rewarders and punishers*, or *preference-driven rewarders and punishers*.⁵ The latter two types’ behavior is indistinguishable until the last period (until which both types employ a reward-and-punishment scheme). In the last period, however, strategic punishers and rewarders play rational best response, while preference-driven punishers and rewarders continue to employ a reward-and-punishment scheme. These varying types of subjects play largely unaffected by each other in the individual treatments, but do influence each other via group discussions in the group treatments. We illustrate how this can lead to different results depending on the different time horizons adopted in our and earlier experiments. Our results suggest that the apparent consensus in the literature regarding sequential two-player games, as summarized by the Bornstein (2008) quote above, needs to be modified to accommodate for differential effects of the time horizon of interaction and possibly other design features—a point we discuss in more detail in the concluding

⁵This categorization is reminiscent of types in public-good games identified in Fischbacher et al. (2001) or, more recently, Reuben and Suetens (in press).

section. In any case, the answer to the question of who behaves more like a game theorist, groups or individuals, is not independent of the time horizon of interaction.

Our second main contribution is on a methodological level. We run both one-period and multiple-period games and employ the strategy method for the first time in a “group” experiment and in a repeated Stackelberg market game.⁶ Doing so not only enables us to control for different first-mover actions across treatments, but also to uncover the shape of complete response functions in (repeated) individual and group Stackelberg markets. The heterogeneity in followers’ behavior mentioned above implies that *average* response functions in both the individual and the team treatments show a somewhat surprising pattern: they slope downward for low leader quantities, slope upward for intermediate leader quantities (around the Cournot quantity), and slope downward again for higher leader quantities. This result suggests that it is appropriate to account for response functions in e.g. sequential market games by running simple linear regressions. As other authors and we demonstrate, structural estimation of other-regarding preference models are able to account for the shape of average and complete individual response functions and thus offer theory-driven alternatives to account for follower behavior.⁷

The remainder of the paper is organized as follows. Section 2 gives a brief overview of the related literature, concentrating mainly on earlier studies of interindividual and intergroup decision making in sequential two-player games. Section 3 introduces the experimental design and the main hypothesis. In Section 4 we report our results and present the estimations of structural models accounting for second-mover behavior. In Section 5 we discuss our results and Section 6 provides a summary and offers some concluding remarks.

2 Related literature

There are now a considerable number of studies comparing behavior of individuals and groups in experimental games. We mainly confine our overview to the papers most relevant for our purposes, that is, to sequential two-player games and market games. Doing so, we only very briefly describe the main results of these studies while providing design details of the most relevant studies in Table 8 in Section A of the Web Appendix. Bornstein (2008) and Engel (2010) provide more complete

⁶Huck and Wallace (2002) elicit complete-response functions in a one-shot Stackelberg experiment. However, we will show that the behavior these authors and we elicit in one-shot games does not (fully) reflect the behavior of subjects and groups who are given the opportunity to learn over the course of various rounds of play.

⁷Note that observed behavior is in line with that predicted by social-preference models, despite the fact that we use non-neutral “firm” language in the instructions and employ random-matching in the multiple-period treatments to weaken other-regarding motives.

overviews of the experimental literature on the behavior of groups.

The early studies on group decision making focus on the ultimatum game. Bornstein and Yaniv (1998) find that groups in the role of the proposer offer less than individuals, and groups in the role of the responder show a willingness to accept less. Robert and Carnevale (1997) also analyzed an ultimatum game, in which, however, no responders were present. These authors find similar results as Bornstein and Yaniv (1998) with respect to proposers.

Subsequent studies replicate this finding in other games. Cox (2002) analyze a trust game (Berg et al. 1995) and reports no differences between groups and individuals playing the role of the trustor. However, groups in the role of the trustee are reported to return significantly less than individuals. Kugler et al. (2007), on the other hand, find that groups are less trusting than individuals, but just as trustworthy. However, if there are differences, both studies point in the direction of more selfish behavior on the part of groups. Kocher and Sutter (2007) analyze a gift-exchange game and find that groups acting in the role of the employer and of the employee choose lower wages and, in return, lower effort levels, respectively, than individuals. Bornstein et al. (2004) have both individuals and groups play two centipede games and report that groups exit the game significantly earlier than individuals. One exception is reported by Cason and Mui (1997) in a dictator game. They note that in some cases, group dictators give more than individual dictators. A recent re-examination by Luhan et al. (2009) indicates that group dictators are more selfish than individuals, possibly caused by replacing the face-to-face discussion among group members with an electronic chat. Bosman et al. (2006) study a power-to-take game where first movers can claim any part of the second movers' income. Then, second movers decide how much of the income to destroy. The authors do not find any differences between groups and individuals both in terms of the first-mover take rates and the income destroyed.

Some studies compare the behavior of groups and individuals in market settings. Bornstein et al. (2008), building on work by Bornstein and Gneezy (2002), analyze Bertrand price competition between individuals and between groups. They find that winning prices were significantly lower in competition between two- or three-person groups than in competition between individuals. In contrast to the results of Bornstein et al. (2008), Raab and Schipper (2009) find no differences in behavior of individuals or groups in Cournot competition. Note, however, that earlier studies show that the Nash equilibrium is a good predictor in individual-player Cournot markets (see, e.g., Huck et al. 2004). Cooper and Kagel (2005) analyze limit-pricing games (Milgrom and Roberts 1982) and report that teams consistently play more strategically and learn faster than individuals. A similar

finding is reported in Kocher and Sutter (2005) in a beauty-contest game. Feri et al. (2010) report that groups can coordinate more efficiently than individuals.

In sum, it seems fair to say that most studies that find differences in interindividual and intergroup comparison find that groups tend to behave more in line with game theoretic predictions, appear more selfish, and show less regard for others, leading Bornstein (2008) and Sutter et al. (2009) to the summaries stated in the Introduction.

3 Experimental design, procedures, and hypotheses

3.1 The Stackelberg duopoly game and its predictions

In our Stackelberg duopoly game, two firms face inverse demand function $p = \max\{30 - Q, 0\}$ where Q denotes total quantity. Both players have constant unit costs of $c = 6$ and no fixed cost. Firms choose their quantities sequentially. First, the Stackelberg leader (L) decides on its quantity q_L , then, knowing q_L , the Stackelberg follower (F) decides on its quantity q_F . The subgame perfect equilibrium is given by $q_L = 12$ and the follower's best-reply function $q_F(q_L) = 12 - 0.5q_L$, yielding $q_F = 6$ in equilibrium. Joint profits are maximized if $q_L + q_F = 12$ and the Nash equilibrium of the simultaneous-move game (Cournot market) predicts $q_L = q_F = 8$.

The following two motivations lead us to choose a Stackelberg game. First, in contrast to other sequential two-player games, a second mover in a Stackelberg game has a much richer strategy space. For instance, in an ultimatum game the choice set of the responder is a binary set containing just two alternatives, "accept" and "reject". By contrast, a second mover in a Stackelberg game has much more room to react to a leader's action, both positively and negatively. As Cox et al. (2008, p. 33) point out "The [Stackelberg] duopoly games are especially useful because the follower's opportunity sets [...] have a parabolic space that enables the follower to reveal a wide range of positive and negative trade-offs between her own income and the leader's income." The second motivation concerns potential results. Huck et al. (2001), who use the same market specification as introduced above, find in their individual-player Stackelberg games that, on average, first movers produce less and second movers produce more than predicted by theory. Hence, there is room for groups to be either closer or further away from the subgame-perfect equilibrium prediction than individuals.

Treatment Name	Sequential Method	Strategy Method	Individual Players	Team Players	# Periods	# Subjects	# Matching Groups
SEQ-IND-1	Yes	No	Yes	No	1	18	9
SEQ-TEAM-1	Yes	No	No	Yes	1	36	6
SM-IND-1	No	Yes	Yes	No	1	18	9
SM-TEAM-1	No	Yes	No	Yes	1	36	6
SEQ-IND-15	Yes	No	Yes	No	15	36	6
SEQ-TEAM-15	Yes	No	No	Yes	15	72	4
SM-IND-15	No	Yes	Yes	No	15	36	6
SM-TEAM-15	No	Yes	No	Yes	15	72	4

Table 1: Experimental design

3.2 Treatment design

Our experiment is based on a $2 \times 2 \times 2$ factorial design, varying the number of periods of interaction (1 period or 15 periods), varying who acts in the two player positions of the Stackelberg game (individuals or groups), and varying the method of eliciting choices (truly sequential play or strategy method). We refer to the eight treatments as follows. The one-shot individual and group treatments with truly sequential play are called “SEQ-IND-1” and “SEQ-TEAM-1”, while the one-shot individual and group treatments which employ the strategy method are called “SM-IND-1” and “SM-TEAM-1”. The corresponding multiple-period treatments are, respectively, called, “SEQ-IND-15”, “SEQ-TEAM-15”, “SM-IND-15”, and “SM-TEAM-15”. Table 1 gives an overview of the design. Information about profits was given in the form of a payoff table (see Table 10 in the Web Appendix). Next, we describe the setting in each of the four treatments in detail.

Treatments SEQ-IND: These are baseline treatments that are similar to the Stackelberg experiment in Huck et al. (2001). In each period, the first mover chose a quantity (selected a row in the payoff table). Knowing the quantity chosen by the first mover, the second mover then decided about his own quantity (selected a column in the table).

Treatments SEQ-TEAM: These are the team baseline treatments which were, with respect to timing, identical to the SEQ-IND treatments except that players were teams (consisting of three participants each) instead of individuals. To reach a joint decision, members of a team could exchange messages within their team via an electronic chat box.⁸ There were no restrictions regarding the contents of messages sent, except that (a) the discussion must be in English; (b) the language used should be civil; and (c) subjects cannot identify themselves by revealing their names,

⁸Electronic chat helps maintain the anonymity among subjects. As this paper is about studying how groups differ from individuals, it is preferable to use electronic chat to exclude the possible influences of other attributes (say, physical attractiveness) on communication.

seat numbers, etc. Subjects could enter their quantity decisions into a box in the decision screen and were then able to submit them to the other group members. All submitted quantity decisions of own group members then appeared on the screen of each group member. As long as not all submitted quantity decisions were the same, the chat box remained open and group members could continue discussing their decision. When all submitted quantity decisions of a team's members were the same, the decision screen (including the chat box) disappeared and subjects had to wait until the experiment continued.

Treatments SM-IND: In these treatments, individual first and second movers made decisions according to the strategy method. That is, first movers decided about a single quantity, while second movers were, at the same time, asked to make a quantity decision for each of the 13 possible quantities the first mover could choose. Once all subjects had made their decisions, the computer randomly matched first and second movers, and selected the relevant quantity of the second mover (that is, the quantity the second mover chose for the quantity chosen by the first mover).

Treatments SM-TEAM: These treatments are similar to treatments SM-IND, except that players are groups instead of individuals. The same communication technology as in treatments SEQ-TEAM were employed to facilitate group decisions. In particular, each member of a second-mover group had to indicate an entire strategy consisting of how it would react to each of the 13 possible choices of a first-mover team. At any point in the process of entering this strategy, second-mover group members could submit their strategy (entered so far) to the other group members. Similar to the individual-player treatments, all entered quantities submitted so far appeared on the screen of each group member. There were no restrictions regarding the order in which follower quantities for the 13 possible first-mover choices had to be entered on the decision screen. Again, the chat box remained open as long as group members had not yet entered the same complete strategy.

3.3 Experimental procedures

The experiment with 18 sessions was conducted at CentER Lab of Tilburg University in April, May, October 2009, and September 2010. Each session consisted of 18 subjects. A total number of 324 Tilburg University students participated in the study. Each subject took part in only one session. Each session consisted of either 1 period or 15 periods. In the repeated sessions, all 15 periods of play counted toward final earnings. There were no practice periods at the beginning of any session. On average, a one-shot session lasted about 45 minutes, whereas the repeated sessions

lasted about 1 hour and 45 minutes (including the time to read the instructions and payment of the subjects). On average, a subject in a one-shot (repeated) session earned €7.29 (€18.51). The experiment was programmed and conducted with the z-Tree software (Fischbacher, 2007).

At the beginning of each session, subjects were randomly assigned to be either a first or second mover, and these roles remained fixed throughout the entire session. In the team treatments, a team was formed by three randomly selected players who belonged to the same team for the entire experiment.⁹ Hence, a team-treatment session consisted of three first-mover teams and three second-mover teams. First-mover and second-mover teams were randomly rematched with each other in each of the 15 periods of the repeated game treatments. In order to control for the size of the random matching group, the 18 subjects in an individual-player session were divided into three cohorts of six subjects (three first and three second movers), and matching happened only within cohorts. This is explained in the instructions.

The instructions used non-neutral language, referring, e.g., to “firms,” “product,” or “profits.” With the instructions, subjects received a payoff table (see the Web Appendix) which, to ease comparison, was the same as used in Huck et al. (2001). The payoff table showed all possible combinations of quantity choices and the corresponding profits. The numbers given in the payoff table were measured in a fictitious currency unit called “Points”. Each firm could choose a quantity from the set $\{3, 4, \dots, 15\}$. The payoff table was generated according to the demand and cost functions given above.¹⁰ In each period, each individual first- or second-mover earned the amount indicated in the table for the selected quantity combination of both firms. In the team treatments, each member of a first- or second-mover firm also earned the amount indicated in the table for the selected quantity combination of both firms.

In the 15-period treatments, first and second movers (individuals or teams) were randomly rematched with each other in each period.¹¹ In the repeated game treatments and starting from

⁹There is a large body of social psychology literature on the size of a small group. The majority stipulate that the lower bound should be three people, for “a dyad (that is, two persons) is a much simpler social system” (see Fisher, 1980).

¹⁰Due to the discreteness of the strategy space, such a payoff table typically induces multiple equilibria (see Holt, 1985). To avoid this, the bi-matrix representing the payoff table was slightly manipulated. By subtracting one Point in 14 of the 169 entries we ensured uniqueness of both the Cournot–Nash equilibrium and the subgame perfect Stackelberg equilibrium.

¹¹Random matching across repetitions was also employed in the team versus individual play signaling games reported in Cooper and Kagel (2005). Note that, given the choice of a multiple-period treatment, random matching across periods constitutes a minimal change compared to a one-shot treatment. It is left for further research to analyze the effect of fixed matching across periods on interindividual and intergroup comparisons in our Stackelberg market game. In the light of our results, we hypothesize that the behavior of groups might be even further away from subgame perfect behavior than that of individuals if one employs fixed matching across multiple rounds of play.

the second period, subjects were informed about the results of the previous round in their own market, including the quantity of the first mover, the (relevant) quantity of the second mover, and own profits.

3.4 Hypothesis

Recall that the Stackelberg market game has a unique subgame perfect equilibrium. Hence, the unique subgame perfect equilibrium of a repeated Stackelberg market game is to play the unique subgame perfect equilibrium of the stage game in each period of interaction. This implies that the rational behavior in each period is described by the subgame perfect equilibrium of the stage game, even if our subjects in the 15-period treatments viewed the experiment as a finitely repeated game, despite the fact that we employed random-matching across periods. However, in the experimental economics literature it is known that play in finitely repeated interactions might be more cooperative even if the stage-game equilibrium is unique and subjects are randomly rematched across rounds within relatively small groups (see, e.g., Selten and Stoecker 1986, or Andreoni and Miller 1993). Yet, in repeated interactions it is a priori not clear how groups will behave in comparison to individuals. Will groups have a tendency towards more selfish behavior in comparison to interindividual interaction as suggested by the earlier literature reviewed in Section 2? Or will there be a trend towards more cooperation in intergroup interaction as this, in the long run, promises higher profits? The few studies reported in the economics literature find that groups in *repeated* interactions play more strategically and converge more quickly to the stage game equilibrium than individuals (Cooper and Kagel, 2005 and Kocher and Sutter, 2005). Hence, based on these earlier results and those reviewed in Section 2, we should expect groups to behave more in accordance with the prediction of subgame perfectness than individuals in both the one-period and the multiple-period treatments. More precisely:

Hypothesis: Group first movers will choose quantities closer to the Stackelberg leader quantity than individual first movers, and group second movers' response functions will be closer to the standard best response function than that of individual second movers, independent of the duration of the interaction.

4 Experimental results

We report the results in two sections with the purpose of comparing behavior of individuals and groups in related treatments. The first section briefly presents summary statistics of our treatments, formal tests for differences in first mover behavior, and visual evidence of second mover behavior. In the second section we concentrate exclusively on second-mover behavior in the 15-period treatments, as accounting for it and formally testing for differences across treatments is much less straightforward than in the case of first movers. In fact, to account for the observed non-monotonic second-mover behavior, we are led to estimate two social preference models: the (simplified) inequality-aversion model by Fehr and Schmidt (1999) as put forward by Lau and Leung (2010), and the parametric model of emotion-driven reciprocity by Cox et al. (2007). To purge the data of learning effects at the beginning of the 15-period sessions (especially in the strategy-method treatments) and, at the same time, preserve sufficient power for maximum-likelihood estimations, in the results section we report and use data from periods 3-15, if not otherwise indicated.

4.1 A first look at the data

Table 2 presents summary statistics of average quantity choices and payoffs for each treatment. The results of the 1-period (15-periods) treatments are presented in the upper (lower) half of this Table. For the strategy-method treatments, only the relevant quantities of the second movers are taken into account (i.e., only quantity choices of second movers at quantities actually chosen by first movers).

In all treatments, we note that average first-mover quantities are clearly smaller and average second-mover quantities clearly larger than the predictions along the subgame perfect equilibrium path, which predicts quantity 12 for first and quantity 6 for second movers. To facilitate comparison, note that the average first (second) mover quantity observed in the 10-period random-matching Stackelberg game of Huck et al. (2001) was 10.19 (8.32). Hence, average quantities of 10.37 (7.77) chosen in our treatment SEQ-IND-15 (which comes closest in terms of design features to this earlier study) are similar to those reported in Huck et al. (2001).

4.1.1 First-mover behavior

In the 1-period experiments we observe that average leader quantities in the individual treatments are slightly *lower* than in the corresponding group treatments. By contrast, in the 15-period

Prediction		Sequential Play				Strategy Method				
		SEQ-IND		SEQ-TEAM		SM-IND		SM-TEAM		
Leader	Follower	Leader	Follower	Leader	Follower	Leader	Follower	Leader	Follower	
1-period treatments										
Individual Quantities	12	6	9.11 (2.32)	8.11 (2.09)	9.33 (2.42)	7.67 (1.03)	9.67 (3.00)	7.11 (1.36)	10.67 (2.07)	7.00 (1.10)
Total Quantities	18		17.22 (1.86)		17.00 (1.79)		16.78 (1.64)		17.67 (1.51)	
Individual Payoffs	72	36	59.78 (13.91)	53.89 (16.31)	62.00 (9.30)	54.33 (18.77)	65.44 (12.60)	49.56 (28.88)	69.33 (14.46)	45.33 (14.46)
Total Payoffs	108		113.67 (18.19)		116.33 (20.09)		115.00 (18.86)		114.67 (10.33)	
15-periods treatments										
Individual Quantities	12	6	10.37 (1.90)	7.77 (1.62)	8.41 (1.71)	7.99 (1.02)	9.29 (2.33)	7.99 (1.87)	8.73 (1.69)	8.16 (1.39)
Total Quantities	18		18.14 (0.90)		16.4 (0.90)		17.28 (0.29)		16.89 (1.17)	
Individual Payoffs	72	36	57.59 (18.64)	43.32 (17.67)	60.54 (11.29)	59.23 (15.29)	57.90 (19.60)	51.16 (22.23)	59.09 (15.06)	55.97 (16.45)
Total Payoffs	108		100.91 (6.89)		119.77 (2.27)		109.06 (1.17)		115.06 (3.76)	

Notes: Standard errors of the mean in parentheses.

Table 2: Summary of experimental results: Average quantities and payoffs

experiments we observe that average leader quantities in the individual treatments are *higher* than in the corresponding group treatments. To test for significance of differences in first-mover data, we ran regressions of the form $q_{ijt}^L = \beta_0 + \beta_1 \times TREATM + \varepsilon_{ijt}$ where q_{ijt}^L is the quantity chosen by leader subject/group i in session j in period t , and $TREATM$ is the dummy to code the two treatments that are included in the regression. The coefficient β_1 measures the difference in average first-mover quantities in the two treatments included in the regression. A test of the hypothesis $H_0: \beta_1 = 0$ will show whether or not the difference is significant. In order to account for possible non-independence of observations in the 15-period treatments, we ran the regressions clustering data by subject or group and by session and using general linear latent and mixed models, GLLAMM (Rabe-Hesketh and Skrondal, 2005). The results are reported in Table 3,¹² where the main comparisons

¹²Using Tobit regression techniques delivers very similar results.

Estimates for the coefficient β_1 . $H_0: \beta_1 = 0$.			
Comparison based on player types		Comparison based on elicitation method	
SEQ-IND	SM-IND	SEQ-IND	SEQ-TEAM
versus	versus	versus	versus
SEQ-TEAM	SM-TEAM	SM-IND	SM-TEAM
1-period treatments			
-0.222 (1.4511)	-1.000 (1.315)	-0.556 (1.191)	-1.333* (0.694)
15-period treatments			
2.132*** (0.399)	0.485 (0.607)	1.209*** (0.470)	-0.373 (0.302)

Notes: Estimated equation: $q_{ijt}^L = \beta_0 + \beta_1 \times TREATM + \varepsilon_{ijt}$, where q_{ijt}^L is the quantity chosen by first-mover subject/group i in session j in period t and $TREATM$ is a dummy used to code the treatments included in the regressions. In all regressions, the dummy variable $TREATM$ is coded such that it is equal to 1 for the treatment mentioned in the upper entry in each column of this table and it is equal to 0 for the treatment mentioned in the lower entry in each column. We report as p -levels $P > |t|$. ***, * indicates significance at the 1%, 10% level. Standard errors in parentheses.

Table 3: Results of statistical tests for differences in first-mover choices

between related individual and group player treatments are presented in the first two columns.

The test results in Table 3 indicate that none of the differences in first-mover behavior between individual and group player treatments are significant in the 1-period treatments. However, first movers in treatment SEQ-IND-15 choose significantly higher quantities than first movers in the corresponding team treatment SEQ-TEAM-15. This contradicts our hypothesis. Note that average first-mover choices in treatment SM-IND-15 and the corresponding team treatment SM-TEAM-15 do not differ significantly.

For completeness, in column 3 and 4 of Table 3 we also report results across the two individual and the two team treatments, for both sets of experiments.

4.1.2 Second-mover behavior

Let us first consider second-mover behavior in the 1-period treatments. Figure 1 shows the average response function observed in the 1-period treatments (for the sequential-play treatments in the left and for the strategy-method treatments in the right panel). As the sequential-play treatments only deliver a few data points, no clear picture emerges in the left panel of Figure 1. If anything, the average response function of team players seems to be closer to the rational response function than that of individual players in the sequential play treatments. A clearer picture emerges in the right panel showing the average response functions in the strategy-method treatments. We make

two observations. First, for leader quantities smaller than the Cournot quantity of 8, the average team response function coincides exactly with the best-response function, whereas the average response function of individuals runs slightly below the rational response function. The latter implies that individuals on average have a slight tendency to reward what could be interpreted as “nice” first-mover behavior. Second, for leader quantities larger than the Cournot quantity of 8, the average response functions of individuals and teams are very similar and both run above the best response function, implying that both individuals and teams slightly punish what could be interpreted as “greedy” first-mover behavior. Although there is weak visual evidence indicating that the observed response functions of teams are closer to the rational best response function than that of individual players in the 1-period treatments (which is in line with earlier results in the literature and our hypothesis), the estimation of simple linear response functions do not deliver any statistically significant differences, neither with respect to the intercept nor to the slope.

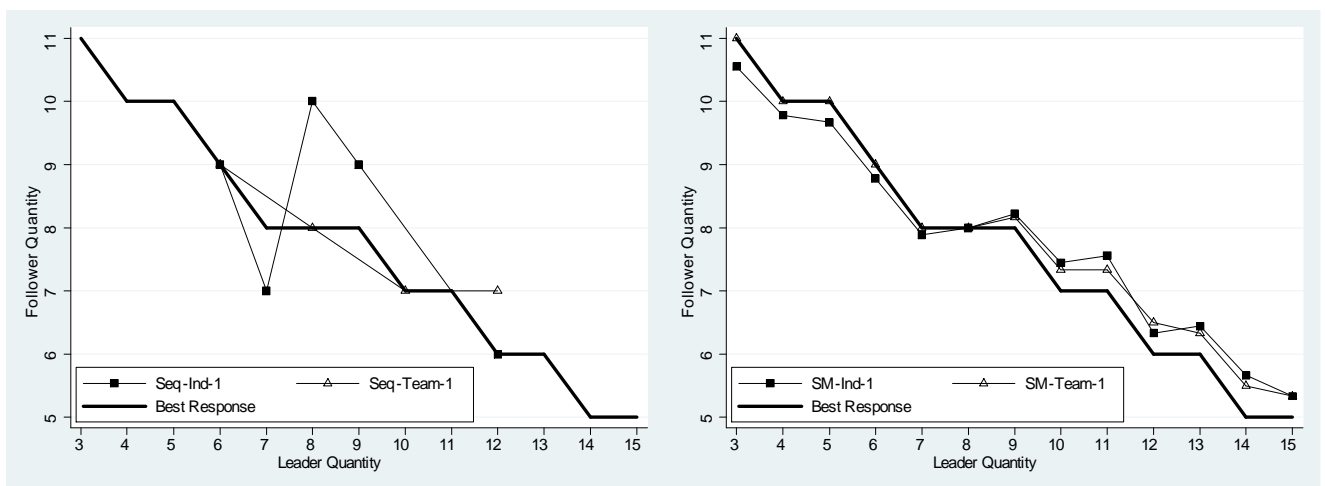


Figure 1: Average response functions observed in the one-period sequential treatments (left) and the one-period strategy-method treatments (right).

Note: There are no observations for leader quantities 10 and 11 in treatment SEQ-IND-1.

Next we turn to second-mover behavior in the 15-period treatments. The two panels in Figure 2 show the average response functions in the 15-periods truly sequential (left panel) and the 15-periods strategy-method treatments (right panel). Inspecting the two panels of Figure 2, it seems fair to state that the average observed response functions of team second movers are further away from the best-response function than that of individual second movers in the 15-period treatments. Importantly, the two panels in Figure 2 as well as simple diagnostic tools suggest that

team second movers reward more and punish harder than individual followers.¹³ Interestingly, all observed response functions show a particular and perhaps somewhat surprising “first slope downward, then slope upward, then slope downward” pattern. This is most evident in the strategy-method treatments. More precisely, the response functions in the strategy-method treatments are downward sloping for leader choices between 3 and 7, upward-sloping for leader choices between 7 and 11/12, and then slope downward again for higher leader choices. Due to the more limited number of different choices of first movers in the sequential treatments, this pattern is less clear in the left panel of Figure 2.¹⁴

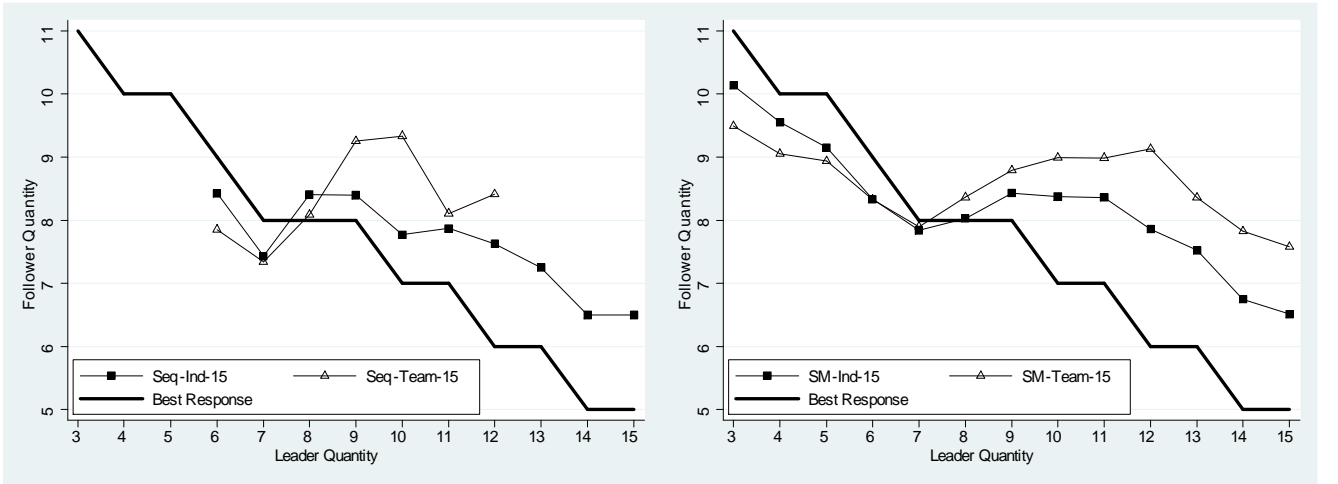


Figure 2: Average response functions observed in the 15-period sequential treatments (left) and the 15-period strategy-method treatments (right).

While estimation of linear and monotonic response functions may serve as a quick diagnostic tool (see footnote 13), from the preceding discussion we conclude that simple linear estimations are inappropriate and incapable of accounting for patterns observed in the average and individual response functions. Furthermore, although basic patterns are easily identifiable on the individual and

¹³Recall that the theoretical response function of followers is given by $q_F(q_L) = 12 - 0.5q_L$. Estimating such response functions as a quick diagnostic tool for our data and comparing the results of the relevant 15-period treatments delivers the following results (details are provided in Section B of the Web Appendix.): First, both the intercept and the slope of the response function employed in the individual-player treatment SEQ-IND-15 are significantly closer to the ones of the rational best response function than the intercept and slope of the response function in the team-player treatment SEQ-TEAM-15. This suggests that individual second movers behave more selfishly than team second movers. Second, the reaction function in treatment SEQ-IND is *downward*-sloping, while the reaction function in treatment SEQ-TEAM is *upward*-sloping. This suggests that team followers reward more and punish harder than individual followers. Third, repeating this exercise for the “relevant” data (i.e., only second-movers’ reactions at quantities actually chosen by first movers) in the 15-period strategy-method treatments confirms the result obtained for the truly sequential treatments.

¹⁴Reaction functions on the individual and group level show heterogeneity ranging from best-response behavior to flat response functions (reflecting a basic reward-and-punishment scheme) to response functions that resemble the shape of those shown in the right panel of Figure 2. We return to this issue in Section 5.

team level in the repeated strategy-method treatments, this is not easy in the repeated sequential treatments as in the latter treatments we observe second-mover behavior only for a possibly small subset of first-mover quantities, which leads to identification and categorization issues. This raises two problems. First, how can we appropriately account for (average) observed response functions in the various treatments? Second, how can we formally compare second-mover behavior across relevant treatments?

We can solve these two problems for the 15-period treatments by employing two recently suggested structural models. In fact, it turns out that the patterns observed in Figure 2 (and at the individual and group level) are consistent with the predictions of models of other-regarding preferences, especially the model by Fehr and Schmidt (1999). Therefore, in the next section we will account for followers' observed response functions by structural estimation of the Fehr and Schmidt (1999) model of inequality aversion as suggested in Lau and Leung (2010). Furthermore, we also estimate and discuss Cox et al.'s (2007) model of emotion-driven reciprocity. Doing so, we will ignore what other-regarding motive drives the results. The important point is that irrespective of the model we estimate, individuals appear to be more selfish than teams. We are able to make this statement as the standard selfish best response function is nested in both of the social preference models we estimate. Therefore, we have a clear and unambiguous method to test which of two observed average response functions is closer to the prediction of subgame perfectness.¹⁵

4.1.3 A closer look at follower data: structural estimations

Estimating a model of inequality aversion: Lau and Leung (2010) suggest that the experimental results of the Stackelberg markets reported in Huck et al. (2001) can be accounted for using a simplified version of the inequality-aversion model by Fehr and Schmidt (1999). In particular, Lau and Leung suggest that the population of second movers consists of a mixture of “standard” and “non-standard” preference types. Standard types are assumed to use the theoretical best-response function, whereas non-standard types are assumed to act as if maximizing a utility function of the Fehr and Schmidt type. In their paper, Lau and Leung first derive the response function of

¹⁵Clearly, in the group treatments it is the group decision-making process that maps individual member's preferences into a decision of the group. Hence, in estimating these models also for the group treatments we maintain an as-if assumption, according to which a group's decision is a reflection of this “group's preferences.” (See also Kocher and Sutter, 2007, p. 71) Given the specific non-monotonic shape of the observed response functions of groups and individuals, we employ these other-regarding preference models as a technical device in order to more adequately estimate and compare response functions.

non-standard types. Interestingly, it turns out that this response functions accurately predicts the shape of the average response functions we observe in our 15-period sessions (see Figure 2). Lau and Leung then develop a maximum-likelihood model in which a share ϕ_{ns} of second movers are non-standard types and a share of $1 - \phi_{ns}$ of second movers are standard types. Estimating this model, using the random-matching Stackelberg data of Huck et al. (2001), they show that a substantial share (about 40%) of the second movers in Huck et al. (2001) appear to have preferences of the Fehr-Schmidt type. The fact that in our 15-period strategy-method data we directly observe individual response functions that are consistent with either those of standard or non-standard types is a rationale to apply Lau and Leung's model to our data to account for follower behavior. In the following we will briefly introduce the model put forward by Lau and Leung, closely following their exposition. We will then estimate it for our four treatments.

Denote player i and j 's payoffs by π_i and π_j , respectively. Then, Fehr and Schmidt preferences are given by

$$u_i = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\} \quad (1)$$

where $0 \leq \beta_i < 1$, $\beta_i \leq \alpha_i$, $i, j = L, F$ with $i \neq j$. The parameter α_i measures player i 's aversion towards disadvantageous inequality, whereas the parameter β_i measures player i 's aversion towards advantageous inequality. For estimation purposes, Lau and Leung make two assumptions. First, they assume that there are two types of second movers. The first type of second movers have standard selfish preferences and hence play according to the standard best response. These second movers are referred to as standard types (S). The second type of players have Fehr-Schmidt preferences and maximize utility as given in 1. These second movers are referred to as non-standard types (NS). Second, Lau and Leung assume that all non-standard types have the same (dis)advantageous inequality parameter. Hence, $\alpha_i = a$ and $\beta_i = b$ for all non-standard players. Lau and Leung assume that the share of non-standard types in the population is given by $\phi_{ns} \in [0, 1]$ where ϕ_{ns} is to be estimated from the data. Hence, the basic assumptions of Lau and Leung's simplified version of the Fehr-Schmidt model are as follows: $\Pr(\alpha_i = a \ \& \ \beta_i = b) = \phi_{ns}$, $\Pr(\alpha_i = \beta_i = 0) = 1 - \phi_{ns}$, where $0 \leq \phi_{ns} < 1$, $0 \leq b < 1$, $b \leq a$.

Recall from above that a standard-type follower reacts according to the best response function given by $q_F^S(q_L) = 12 - \frac{1}{2}q_L$. Regarding the response function of non-standard followers, Lau

and Leung show that it is given by

$$q_F^{NS}(q_L) = \begin{cases} 12 - \frac{q_L}{2(1-b)} & \text{if } q_L \in A \\ q_L & \text{if } q_L \in B \\ 12 - \frac{q_L}{2(1+a)} & \text{if } q_L \in C, \end{cases}$$

where $A = \left[3, 12 \left(\frac{1-b}{3-2b}\right)\right]$, $B = \left[12 \left(\frac{1-b}{3-2b}\right), 12 \left(\frac{1+a}{3+2a}\right)\right]$, and $C = \left[12 \left(\frac{1+a}{3+2a}\right), 15\right]$. Note that the best-response function is piecewise linear and that the standard best response is obtained when $a = b = 0$. Note also that it slopes downward for low, slopes upward for intermediate, and slopes downward again for high first-mover quantities. Hence, it predicts the pattern observed in Figure 2. To briefly gain some intuition, consider the case of $q_L \in A$. Best responding to such a quantity choice maximizes a second mover's profit but reduces the utility of a non-standard type due to advantageous inequality. If q_L is small enough, the non-standard second mover finds it preferable to reduce quantity below the best response, which reduces advantageous inequality by more than that it decreases own profits.

To derive the likelihood function, let x_i and y_i represent the i th observed tuple of observed leader and follower choices. Lau and Leung assume that a follower with standard [non-standard] preferences chooses according to $y_i = q_F^S(x_i) + \varepsilon_i$ [$y_i = q_F^{NS}(x_i) + \varepsilon_i$], where ε_i is *iid* according to a normal distribution $N(0, \sigma^2)$ and $q_F^S(x_i)$ and $q_F^{NS}(x_i)$ are as given above. Since Lau and Leung assume a share ϕ_{ns} of non-standard and a share of $1 - \phi_{ns}$ standard second movers, the probability density of observing y_i is given by

$$(1 - \phi_{ns}) \times f_S(y_i | x_i; \sigma) + \phi_{ns} \times f_{NS}(y_i | x_i; a, b, \sigma),$$

where $f_S(y_i | x_i; \sigma)$ and $f_{NS}(y_i | x_i; a, b, \sigma)$, respectively, are the probability densities of observing y_i when the second mover has, respectively, standard and non-standard preferences.¹⁶ The log likelihood function of observing the sample $(x_i, y_i)_{i=1}^{N_{\text{Treatm}}}$ of leader and follower choices is then

¹⁶Using the definition of the normal distribution, one obtains $f_S(y_i | x_i; \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - 12 + \frac{x_i}{2})^2}{2\sigma^2}\right)$ and $f_{NS}(y_i) = f_A(y_i)^{1-D_B(x_i)-D_C(x_i)} \times f_B(y_i)^{D_B(x_i)} \times f_C(y_i)^{D_C(x_i)}$ where $f_A(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - 12 + \frac{x_i}{2(1-b)})^2}{2\sigma^2}\right)$, $f_B(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - x_i)^2}{2\sigma^2}\right)$, and $f_C(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - 12 + \frac{x_i}{2(1+a)})^2}{2\sigma^2}\right)$, with the indicator variables defined as $D_B(x_i) = \begin{cases} 1 & \text{if } 12 \left(\frac{1-b}{3-2b}\right) < x_i \leq 12 \left(\frac{1+a}{3+2a}\right) \\ 0 & \text{otherwise} \end{cases}$ and $D_C(x_i) = \begin{cases} 1 & \text{if } 12 \left(\frac{1+a}{3+2a}\right) < x_i \\ 0 & \text{otherwise} \end{cases}$.

	Truly Sequential Play		Strategy Method			
			All Data		Relevant Data	
	SEQ-IND-15	SEQ-TEAM-15	SM-IND-15	SM-TEAM-15	SM-IND	SM-TEAM-15
ϕ_{ns}	1.000	1.000	1.000	1.000	1.000	1.000
a	0.303*** (0.085)	0.629*** (0.129)	0.279*** (0.068)	0.641** (0.273)	0.358** (0.151)	0.500*** (0.160)
b	0.216*** (0.029)	0.252*** (0.017)	0.192*** (0.062)	0.215** (0.090)	0.215*** (0.035)	0.4000*** (0.001)
σ	1.506*** (0.164)	0.862*** (0.190)	1.485*** (0.200)	1.740*** (0.423)	1.658*** (0.263)	1.322*** (0.224)
LL	-427.864	-198.222	-4599.221	-3334.489	-436.902	-262.242
N	234	156	2535	1690	173	156
Hypothesis	$a^{\text{SEQ-IND-15}} = a^{\text{SEQ-TEAM-15}}$		$a^{\text{SM-IND-15}} = a^{\text{SM-TEAM-15}}$		$a^{\text{SM-IND-15}} = a^{\text{SM-TEAM-15}}$	
Testing	&		&		&	
	$b^{\text{SEQ-IND-15}} = b^{\text{SEQ-TEAM-15}}$		$b^{\text{SM-IND-15}} = b^{\text{SM-TEAM-15}}$		$b^{\text{SM-IND-15}} = b^{\text{SM-TEAM-15}}$	
	$p = 0.075$ ($\chi^2_{(2)} = 5.17$)		$p = 0.077$ ($\chi^2_{(2)} = 5.14$)		$p = 0.000$ ($\chi^2_{(2)} = 28.72$)	

Table 4: Estimation results for Lau-Leung’s (2010) implementation of the Fehr and Schmidt model when the share of non-standard types is restricted to be equal to 1.

given by

$$\begin{aligned} & \ln L \left(a, b, \phi_{ns}, \sigma; (x_i, y_i)_{i=1}^{N_{\text{Treatm}}} \right) \\ &= \sum_{i=1}^{N_{\text{Treatm}}} \ln \left\{ (1 - \phi_{ns}) f_S(y_i) + \phi_{ns} \left[f_A(y_i)^{1-D_B(x_i)-D_C(x_i)} \times f_B(y_i)^{D_B(x_i)} \times f_C(y_i)^{D_C(x_i)} \right] \right\}, \end{aligned}$$

where N_{Treatm} is the number of observations in the treatment under consideration. To control for non-independence of observations, we cluster standard errors on individuals or groups.

In an effort to first estimate the average response functions as shown in Figure 2, we set $\phi_{ns} = 1$, that is, in a first step we assume that there are only non-standard types. The estimation results are given in Table 4.¹⁷

We note that the parameter estimates of the inequality-aversion parameters a and b are significantly different from 0 in all treatments and data sets. Note also that the parameter estimates of a and b are in line with the restrictions $0 \leq b < 1$ and $b \leq a$ imposed by the Fehr and Schmidt

¹⁷10 out of 4998 choice pairs result in negative payoffs to both players (1 in SEQ-TEAM-15; 3 in SM-IND-15 relevant data; 5 in SM-TEAM-15 all data and 1 in SM-IND-15 all data). Since the utility function in (1) is defined only for non-negative payoffs, we truncate these observations at $q_F = 24 - q_L$ which implies zero payoffs for both players. Furthermore, in treatment SM-IND-15, seven second movers reacted with quantities above the best-response to first-mover quantities smaller than 8. A possible explanation is that individual second movers exposed to the strategy method are likely to make more errors, especially at first mover quantities they do not actually observe very often in the course of the experiment. In the SM treatments (all data), observations from three individuals and two teams were dropped due to extreme responses to leader quantity 3 and 15, causing difficulties in finding convergence.

model. Most importantly for the purpose of deciding which observed average response function is closer to the rational best-response function (characterized by $a = b = 0$), we observe that both the disadvantageous inequality parameter a and the advantageous inequality parameter b are larger in the team treatment than in the relevant individual treatment. For instance, while in SEQ-TEAM-15 the parameter a is estimated as 0.629, it is only 0.303 in treatment SEQ-IND-15. This is in contrast to the main hypothesis according to which the observed response function of teams should be closer to the rational best-response function than the one of individuals. The test results reported at the bottom of Table 4 indicate that we can (weakly) reject the hypothesis that, in each of two relevant treatments comparisons, the parameters a and b are the same.

We next estimate the full model, dropping the restriction $\phi_{ns} = 1$, and concentrating on the estimated share of standard and non-standard types in two related treatments. The results are shown in Table 5. With the exception of treatment SEQ-IND-15, the share ϕ_{ns} of nonstandard types is estimated to be significantly larger than 0 in all treatments and range from about 0.27 in the individual treatments to 0.773 in treatment SEQ-TEAM-15. More importantly for our purposes, the share of non-standard types is estimated to be consistently higher in the group treatments than in the corresponding individual treatments. These differences are highly significant in all treatments (and data sets), as indicated by the test results presented at the bottom of Table 5.¹⁸ This again is strong evidence against our main hypothesis according to which groups are expected to be more in line with the predictions of subgame perfectness.

Estimating a model of reciprocity: Recently, the behavior of second-movers in Stackelberg markets was also accounted for by a model of emotion-driven reciprocity (Cox, Friedman, and Gjerstad 2007). Clearly, in addition to or besides inequality aversion, reciprocity is a possible motivational force for second-mover behavior. Furthermore, the response function of the Cox-Friedman-Gjerstad model is flexible enough, in principle, to rationalize the shape of the observed average response functions shown in Figure 2. Therefore, as a robustness check of our finding that team second-movers are less (myopically) rational than individual second-movers in the 15-period treatments, we also estimated the model put forward by Cox, Friedman, and Gjerstad. We present the details in Section C of the Web Appendix, but note here that the estimation results show that the “emotional state” of groups is more pronounced (both positively and negatively) than that of individuals. In particular, an estimated reciprocity parameter is significantly larger in

¹⁸We apply Wald test for testing parameter significance. We first accommodate data from different treatments into a large, unrestricted model. Then we put restrictions on coefficients to see whether they are equal to zero.

	Truly Sequential Play		Strategy Method			
			All Data		Relevant Data	
	SEQ-IND-15	SEQ-TEAM-15	SM-IND-15	SM-TEAM-15	SM-IND-15	SM-TEAM-15
ϕ_{ns}	0.277 (0.175)	0.773*** (0.077)	0.276*** (0.011)	0.418*** (0.015)	0.279*** (0.092)	0.543*** (0.132)
a	1.713 (1.513)	0.949*** (0.256)	1.035*** (0.029)	1.479*** (0.054)	15.584 (15.727)	3.327** (1.147)
b	0.383 (0.673)	0.470*** (0.034)	0.823*** (0.004)	0.828*** (0.005)	0.605*** (0.020)	0.476*** (0.024)
σ	1.094*** (0.117)	0.717*** (0.200)	0.818*** (0.013)	1.034*** (0.020)	0.968*** (0.140)	0.931*** (0.190)
LL	-426.523	-196.848	-4149.451	-3232.943	-377.509	-241.571
N	234	156	2535	1690	227	156
Hypothesis	$\phi_{ns}^{\text{SEQ-IND-15}} = \phi_{ns}^{\text{SEQ-TEAM-15}}$		$\phi_{ns}^{\text{SM-IND-15}} = \phi_{ns}^{\text{SM-TEAM-15}}$		$\phi_{ns}^{\text{SM-IND-15}} = \phi_{ns}^{\text{SM-TEAM-15}}$	
Testing	$p < 0.001$	$(\chi^2_{(1)} = 40.93)$	$p = 0.006$	$(\chi^2_{(1)} = 7.634)$	$p = 0.002$	$(\chi^2_{(1)} = 9.82)$

Table 5: Estimation results for Lau-Leung’s implementation of the Fehr and Schmidt model

the groups treatments than in the corresponding individual-player treatments. Hence, the results of this robustness exercise show that team followers appear to behave more reciprocally (or less selfishly) than individual followers. This is, again, not in line with our main hypothesis.

5 Discussion

5.1 A potential explanation of the results

Summarizing our results derived so far, we can state the following. In the one-shot treatments we find weak evidence that is in line with previous results reported in the literature according to which groups are closer to the subgame perfect equilibrium prediction than individuals (although the differences we find are small and not significant). In our 15-period treatments, by contrast, we find that in comparison to individuals, groups choose lower quantities as first movers and reward more and punish harder as second movers. In other words, groups in our repeated game treatments appear to be *less* selfish than individuals. This raises the question of how to explain the different results in our and the earlier experiments. We believe that a possible explanation rests on the observation that there is substantial *heterogeneity in subjects’ types*, and on the fact that different *time horizons* were used in our and the earlier experiments.

Regarding heterogeneity of subjects’ types, further below we present substantial evidence suggesting that most subjects belong to one of three categories: (myopic) profit maximizer (“PM”);

strategic rewarder and punisher (“*Strat-R&P*”); and other-regarding preference-driven rewarder and punisher (“*Pref-R&P*”) (where the other-regarding preference can be, e.g., inequality aversion or reciprocity). We will identify these types by concentrating on second-mover behavior, which is easily interpretable. *PMs* always maximize their payoff in response to any first-mover choice, independent of the time horizon of interaction. *Strat-R&Ps* reward “nice” low leader quantities and punish “greedy” high leader quantities during all but the final period, where they revert to rational best response. These types arguably want to strategically “educate” leaders to choose lower quantities, until the final round where they revert to opportunistic behavior. Hence, *PMs* and *Strat-R&Ps* are indistinguishable in one-shot games. *Pref-R&Ps* behave like *Strat-R&Ps* in all but the last period. Since *Pref-R&Ps* do not revert to payoff-maximizing behavior even in the final round, their reward and punishment behavior can be interpreted as stemming from other-regarding preferences. Note that the existence of such or similar types has been reported in other studies in the literature (see, e.g., Fischbacher et al. 2001 and especially Reuben and Suetens (in press) for the existence of *Strat-R&Ps* and *Pref-R&Ps*).

Many earlier experiments that report groups to be more selfish than individuals (see Section 2) employ one-shot interaction between subjects. By contrast, we have subjects interact both one-shot *and* repeatedly over 15 periods (using random re-matching of individuals and teams). We believe that the heterogeneity in subjects’ types and the different time horizons could explain the different results in our and the earlier experiments. For this purpose, let us first consider the case of one-shot interactions. Assume that subjects are one of the three types mentioned above. Of those, *PMs* and *Strat-R&Ps* will behave according to subgame perfect behavior while *Pref-R&Ps* will deviate from this behavior by displaying other-regarding concerns. Hence, behavior in inter-individual one-shot treatments is likely to be a mixture of rational and other-regarding behavior. However, in the one-shot team treatments it is conceivable that both *PMs* and *Strat-R&Ps* convince the potentially present *Pref-R&Ps* that deviation from subgame-perfect behavior is not meaningful in a one-shot interaction. For instance, given the first mover quantity they might convince a group member who is an emotion-driven reciprocator to control feelings and to also vote for myopic best-response behavior. Hence, behavior in inter-group one-shot treatments is likely to be more homogeneous and more in line with the prediction of standard game theory. This would explain why in earlier experiments (and to a lesser extent in our experiment) groups were on average found to be more selfish than individuals.

Consider now the case of multiple-period interactions. In the inter-individual treatments,

average behavior will be a mixture of other-regarding behavior (displayed by both *Pref-R&P*s and *Strat-R&P*s) and *PM*s. However, in the multiple-period team treatments it is conceivable that *Strat-R&P* now side with *Pref-R&P*s in an effort to convince the potentially present *PM*s that more cooperative behavior (established by reward and punishment) is the better thing to do in the sense of achieving higher overall payoffs when the game is repeated multiple times (even with random-matching across periods). Hence, behavior in inter-group multiple-round treatments is likely to be more homogenous and more in line with cooperative behavior. This would explain why in our repeated game treatments, groups were on average found to be less (myopically) “rational” than individuals.¹⁹ We believe that the mechanisms we describe here are applicable to simultaneous-move dilemma games (such as prisoner’s dilemma) and to sequential games that allow for competitive and cooperative outcomes (such as dictator, ultimatum, trust, or Stackelberg games). It is presumably less applicable to so-called “Eureka”-type problems that have a “clear” solution that, once discovered, is recognized as such (e.g., limit-pricing or beauty-contest game). In the remainder of this section we provide evidence for the existence of the different types of subjects mentioned above.

5.2 Evidence for the explanation of the results

The first kind of evidence is provided by the estimation results of the Lau and Leung (2010) model presented in Section 4.1.3. There, the term $1 - \phi_{ns}$ measures the share of “standard” or best-response subjects. As this share is estimated to be significantly larger than 0, no matter which of the individual-treatment data sets we use, this provides evidence for the existence of myopic profit maximizers.

The second, more direct evidence is delivered by the inspection of the individual response functions in Figures 6 and 7 in Section D of the Web Appendix. These Figures show the individual response functions of second movers in round 14 and 15, respectively, in treatment SM-IND-15. Inspecting the response function in Figures 6 and 7, we find the following categorization. *PM*s: Subjects 18, 21, 26, 28, and 30 are pure myopic profit maximizers. Moreover, subjects 15 and 29 also play mostly best response, and could therefore also be classified as myopic profit maximizer.²⁰

¹⁹Note that the mechanism we propose here, where some subjects in a group try to convince other subjects of what is the “right” thing to do depending on the time horizon, is in line with “Persuasive Argument Theory” (PAT) put forward in the psychological literature (see, e.g., Stoner, 1961; Teger and Pruitt, 1967; Levine and Moreland, 1998). PAT suggests that if the mean response of the individuals exhibits a preference towards a particular position, it is likely that the subjects will be exposed to more persuasive arguments in favor of this position during the discussion. Therefore, the ex-post group outcome will shift towards that particular initial position.

²⁰Note that in treatment SM-TEAM-15 there are only 2 pure profit maximizers (teams 2 and 9), as shown in Figures

Strat-R& P: In period 14, subject 24 (25) basically plays best response for quantities smaller than 8 (9). In round 15, however, both subjects choose best response behavior for all first-mover quantities. Hence, these two subjects can clearly be identified as strategic players. To a lesser extent, the same is true for subjects 22 and 23. The remaining subjects consist of those that can be classified as *Pref-R& P* and “*Others*.”²¹ Figures 4 and 5 in Section D of the Web Appendix show the complete response function in the one-shot strategy-method treatments. Most of the response functions we observe in these Figures show best-reply behavior, which is compatible with behavior described for *PMs* and *Strat-R&Ps*. As some of these observed one-shot response functions also reflect a taste for rewarding and punishment, we also have evidence for *Pref-R&Ps* in these treatments.²²

The third kind of evidence is provided by the analysis of follower chat protocols. We do this in view of illustrating two things: that statements made during the group discussions can be (albeit not exclusively) assigned to subject types mentioned above, and that many of the discussions can be easily characterized as a conflict between the types of subjects mentioned above. To economize on space, we only concentrate on followers in the group treatments SEQ-TEAM-15 and SEQ-TEAM-1. Again, followers’ discussions simply provide “richer” material.

We started the analysis by first listing all (interpretable) statements, proposals, motives, etc. that were voiced in any of the group chats in treatment SEQ-TEAM-15. Then we tried to assign each of these statements to a broader category which would also reflect the type categories introduced above. These categories were: *PM*, *Strat-R&P*, *Pref-R&P*, *Non-PM*, and “*Other*”. These categories are the column titles in Table 6, which summarizes our chat analysis of treatment SEQ-TEAM-15. The complete list of all statements collected under the respective broad category for treatment SEQ-TEAM-15 is provided in the first column of Table 11 in the Web Appendix. Statements summarized in category Non-PM are those that, arguably, belong to either category *Strat-R&P* or *Pref-R&P*. However, an assignment to either of these categories is not unambiguous which is why we summarize them in a separate category.

The next step of the analysis was to briefly summarize each group’s discussion in each round of treatment SEQ-TEAM-15. It turned out that each discussion can be summarized by one of eight headlines, which provide the row titles in the upper part of Table 6. Here “R” stands for

8 and 9 in Appendix D. Hence, we observe a lower share of profit maximizers in the team treatment than in the individual treatment. This is consistent with our explanation above according to which, through team discussions, PMs are likely to be convinced to abandon their behavior in favor of some sort of reward-and-punishment behavior.

²¹These results are confirmed by a hierarchical agglomerative cluster analysis (see Kaufman and Rousseeuw, 1990) of individual response functions. The details are available from the authors upon request.

²²Note that the standard deviation of responses to specific leader quantities is usually larger for individuals than for groups, which is in line with our explanation put forward in the previous section.

Overall characterization of a round's discussion	# Obs.	Categories of motives mentioned in group discussions				
		PM	Strat R&P	Pref R&P	Non- PM	Other
Quick agreement on R	23	—	1	5	30	6
Quick agreement on PM	90	94	—	3	5	3
Quick agreement on P	15	—	2	13	13	7
PM vs R, R “wins”	10	18	19	9	12	2
PM vs R, PM “wins”	5	9	9	3	8	2
PM vs P, P “wins”	20	23	10	13	16	8
PM vs P, PM “wins”	7	20	1	4	6	5
How much P?	10	3	1	3	10	6
Σ	180	167 (41.5%)	43 (10.7%)	53 (13.2%)	100 (24.9%)	39 (9.7%)
Leaders' Choices						
$q_L = 6$	12	18	10	10	9	4
$q_L = 7$	42	31	13	6	34	5
$q_L = 8$	79	79	2	7	6	4
$q_L = 9$	4	—	2	—	5	3
$q_L = 10$	14	11	3	9	15	8
$q_L = 11$	9	11	4	5	8	5
$q_L = 12$	20	17	9	16	23	10
Σ	180	167	43	53	100	39

Notes: Abbreviations used: R = Reward, PM = Profit maximization, P = Punishment. Percentages in row “ Σ ” refer to percentages of cases in the columns labeled “Categories of motives mentioned in group discussions”.

Table 6: Analysis of chat protocols in treatment SM-TEAM-15

reward, “PM” for profit maximization, and “P” for punishment, respectively. The upper half of Table 6 is a cross table of the short summaries of chat contents (column 1) and the broad categories of statements made during the chats (row 2). For instance, in the 23 cases that a round’s chat could be summarized as “quick agreement on R” in treatment SEQ-TEAM-15, there was 1 statement attributable to a *Strat-R&P* motive, 5 statements attributable to a *Pref-R&P* motive, 30 statements attributable to Non-PM motive, and 6 statements that could not be summarized under a common headline.²³ A different cross table is provided in the lower half of Table 6. Here we cross the leader groups’ quantity choices with the broad categories of statements made in treatment SEQ-TEAM-15. (A more detailed overview of the cross table is provided in Tables 11 and 12 in the Web Appendix). The understandably less extensive categorization for treatment SEQ-TEAM-1 is provided in Table 7, which has a similar structure as 6.

With these preparations in place, we can come back to the two points (see second paragraph

²³Note that the sum of these statements do not sum up to 23, the number of observations listed in column 2 in Table 6. This is so because typically many different statements were made during one group’s discussion in a single round of the experiment.

Overall characterization of a round's discussion	# Obs.	Categories of motives men- tioned in group discussions		
		PM	Non- PM	Other
PM vs R, PM “wins”	2	4	2	—
PM vs P, P “wins”	2	3	4	1
PM vs P, PM “wins”	2	6	2	5
Σ	6	13 (48.1%)	8 (29.6%)	6 22.2(%)
Leaders' Choices				
$q_L = 6$	1	2	1	—
$q_L = 8$	2	5	2	2
$q_L = 10$	1	1	1	1
$q_L = 12$	2	5	4	3
Σ	6	13	8	6

Notes: Abbreviations used: R = Reward, PM = Profit maximization, P = Punishment. Percentages in row “ Σ ” refer to percentages of cases in the columns labeled “Categories of motives mentioned in group discussions”.

Table 7: Analysis of chat protocols in treatment SM-TEAM-1

of this subsection) we want to illustrate with the help of the chat protocols. Let us concentrate on Table 6, which shows the results for treatment SEQ-TEAM-15. First, we observe that also in the chat protocols we find ample evidence for various types of subjects. In fact, the column sums in the upper (or lower) part of Table 6 suggest that respectively 41.5%, 10.7%, and 13.2% of all interpretable statements made in treatment SEQ-TEAM-15 stem from subjects who can, respectively, be classified as (myopic) profit maximizers, strategic teachers, and other-regarding subjects. Second, row-wise inspection of Table 6 illustrates the conflicts that are carried out in group discussions. Surely, and almost tautologically, in cases in which there is quick agreement on an action we typically observe only one kind of argument. For instance, if there is quick agreement on best response (which typically happens in response to leader quantity 7 or 8, see the lower part of Table 6) there are almost no statements made in favor of a different action. On the other hand, if there is quick agreement on either reward or punishment, no statement is made in favor of best response. The more interesting cases arise, of course, when a group's discussion can be characterized as a conflict between best response and a rewarding or a punitive action. In these cases we typically observe arguments and statements that can be attributed to all kinds of motives ranging from myopic profit maximization to strategic teaching to other-regarding and non-profit maximizing behavior. For instance, in the 10 group discussions that revolve around the question whether the leader group should be best responded to or be rewarded, and rewarding is the result

(see the row labeled “PM vs R, R “wins”” in Table 6), we observe 18 statements made in favor of profit maximization, and, respectively, 19, 9, and 12 statements in favor of strategic teaching, other-regarding motives, and non-profit maximization behavior. Not surprisingly, as there are many more statements made against best response, in these cases a response is chosen that rewards the leader’s action. Similar patterns can be observed in the other discussions that are characterized by conflicts among group members. Note the fact that in conflict-laden group discussions it is typically the case that all kinds of arguments are exchanged, which can be seen by reading row-wise the lower part of Table 6. For instance in response to the collusive leader quantity $q_L = 6$, we see statements coming from all “camps.” This applies likewise for higher leader quantities (≥ 10).

6 Summary and concluding remarks

In this study we compare the behavior of individuals and groups in a sequential market game in both one-period and multiple-period game treatments. Our main finding is the differential effect that the time horizon of interaction has on the extent of individual and group players’ (non)conformity with subgame perfectness. In the one-shot treatments we find that, although on average groups appear to be somewhat closer to subgame perfectness than individuals, none of the differences in behavior are statistically significant. However, in the repeated game treatments we find that groups are less selfish and more cooperative than individuals. These findings are to a large extent independent of the mode in which we elicit choices or the model we employ to account for second-mover behavior. Importantly, our main finding is in (stark) contrast to results in earlier studies reporting that groups appear to be more selfish than individuals. A possible explanation for the different results in our and earlier studies is that there is heterogeneity in subjects’ types, ranging from pure (myopic) profit maximization to either strategic or preference-driven reward-and-punishment behavior. Depending on the time horizon of the interaction, the exchange of persuasive arguments via discussions is likely to lead groups to (possibly) more selfish behavior in one-shot interactions and to more cooperative behavior in repeated interactions. Since subjects in inter-individual interactions cannot exchange arguments regarding what constitutes “meaningful” behavior in the face of different features of the interaction, it is conceivable that their behavior reacts to a lesser extent to the time horizon of interaction. Our main result implies that the statement “Groups, it seems, are more selfish and more sophisticated players than individuals, and, as a result, interactions between two unitary groups are closer to the rational, game-theoretical solution than interactions between two individuals.”

(Bornstein 2008, p. 30), which summarizes much of the previous literature on interindividual and intergroup comparisons in simple, sequential-move games, needs modification.

Our results show that the second part of the above statement does not generally apply to multiple-period game settings. In fact, for games that leave relatively more room for other-regarding preferences, the time horizon of interaction seems important, leading the play of groups either closer or further away from the game-theoretic prediction than that of individuals. In the light of our results, and to the extent that the explanation of our results is convincing, it might be worthwhile to revisit other simple sequential-move games (such as the ultimatum game, the trust game, the centipede game, and the gift-exchange game) to check for a possible differential effect of the time horizon of interaction. Whereas we concentrate on the effect of the time horizon of interaction in interindividual and intergroup comparisons, much more research is called for to analyze the effect of other design features such as the nature of communication within groups (e.g., face-to-face or anonymous chat) or the voting mechanism (e.g., majority or unanimity voting).²⁴

The Stackelberg market game is, arguably, not of the “Eureka” type, where a solution once found is recognized as such by players. Therefore the results of our repeated markets are not necessarily in contrast to the findings summarized by the second quote in the Introduction, which summarizes results from repeated interaction in games with a strong “Eureka” component. In these games, behavior of groups was shown to converge much faster to the (same) game-theoretic prediction than individuals. However, our repeated-game results show that neither groups nor individuals converge to a (refined) game-theoretic prediction, and, what is more, that groups clearly diverge further from it than individuals (see also Cox and Hayne, 2006 and Sutter et al. 2009).

It is one question to check who is closer to game-theoretic predictions in interindividual and intergroup comparisons; another is to check who earns higher profits. In particular and perhaps not surprisingly, there does not seem to be a simple relationship between higher conformity with game-theoretic predictions and higher profits. For instance, Feri et al. (2010) show that groups are significantly better at coordinating on more efficient outcomes and hence earn higher profits than individuals, while Bornstein et al. (2004) show that groups exit earlier in one-shot centipede games, leading to lower profits in comparison to individuals. On the other hand, Cox and Hayne (2006) and Sutter et al. (2009) show that in some auction formats, groups pay higher prices than individuals and are more often victim of the winner’s curse than individuals, and therefore groups make smaller

²⁴Some studies, such as Elbittar et al. (2004), Gillet et al. (2009, 2011), vary the nature of managerial decision-making processes within firms and analyze their impact on intergroup and interindividual firm behavior.

profits than individuals. In our repeated Stackelberg markets employing truly sequential play, however, we find that groups earn significantly higher total profits than individuals, although groups' behavior is further away from the (refined) game theoretic prediction. These results seem to suggest that more research is needed to explore when (type of game, etc.) and why (design features, ease of collusion, etc.) groups earn more than individuals. The answer to this question is important for a recommendation on when to entrust decision making to groups instead of to individuals in real-world settings.

Our results also speak to the extensive psychological literature on individual-versus-group decision making, especially regarding the so-called “discontinuity effect.” This effect, which so far largely rests on observations in one-shot prisoner’s dilemma games, refers to the finding of “intergroup interactions to be more competitive, or *less cooperative*, than interindividual relations” (Wildschut and Insko, 2007, p. 175, emphasis added). Clearly, the results of our 15-period treatments show that, indeed, there is a clear difference or discontinuity between inter-individual and inter-group interaction. However, our results also show that the “discontinuity” goes in the opposite direction than stated so far in the psychology literature. Hence, the definition of the discontinuity effect might need modification too, accommodating, among other things, the time horizon of interaction.²⁵

In this paper, we also make progress in terms of methodology regarding the comparison of interindividual and intergroup behavior. First, we study both one-shot and multiple-period treatments in a unified framework, whereas other studies either only implement one-period or only multiple-period games. Second, in an additional set of treatments we employ the strategy-method to control for the possibility that differences in second-mover behavior observed across interindividual and intergroup treatments are driven by the different experiences second movers have in the two environments. This also enables us to uncover the complete shape of the response function used by experienced Stackelberg followers. Independent of whether they were elicited from individual or group followers, average response functions in repeated Stackelberg markets display the same characteristic pattern. They slope downward for lower leader quantities, slope upward around the Cournot quantity and slope downward again for larger leader quantities. These results imply that it may not be warranted to just run linear regressions to estimate followers’ response functions in repeated games, as done for instance in Huck et al. (2001).²⁶ Interestingly, the specific shape of

²⁵Note that Lodewijckx et al. (2006) discuss the possibility that different time horizons may have differential effects on interindividual versus intergroup comparisons. However, they do not provide persuasive evidence for this claim.

²⁶It remains to be checked whether similar unexpected patterns can be observed in other sequential-move games,

followers' response functions is nicely predicted by models of other-regarding preferences, such as Fehr and Schmidt (1999). Building on earlier contributions by Lau and Leung (2010) and Cox et al. (2007), we demonstrate that experienced followers' response functions are more adequately accounted for by estimating structural models of other-regarding preferences rather than by simple linear regressions. This allows us to unambiguously test which of two response functions is closer to the best-reply function, which can be viewed as a third methodological contribution of our paper.

References

- [1] Andreoni, J. and J.H. Miller (1993): Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence, *Economic Journal* 103, 570-585.
- [2] Berg, J., J. Dickhaut and K. McCabe (1995): Trust, reciprocity, and social history, *Games and Economic Behavior* 10(1), 122-142.
- [3] Bornstein, G. (2008): A classification of games by player type, in: *New Issues and Paradigms in Research on Social Dilemmas*, A. Biel, D. Eek, T. Gärling, and M. Gustafsson (eds.), Springer US, 27-42.
- [4] Bornstein, G. and U. Gneezy (2002): Price competition between teams, *Experimental Economics* 5, 29-38.
- [5] Bornstein, G. T. Kugler and A. Ziegelmeyer (2004): Individual and group decisions in the centipede game: Are groups more "rational" players?, *Journal of Experimental Social Psychology* 40, 599-605
- [6] Bornstein, G. and I. Yaniv (1998): Individual and group behavior in the ultimatum game: Are groups more "rational" players?, *Experimental Economics* 1, 101-108.
- [7] Bornstein, G., D. Budescu, T. Kugler and R. Selten (2008): Repeated price competition between individuals and between teams. *Journal of Economic Behavior and Organization* 66, 808-821.
- [8] Bosman, R., H. Hennig-Schmidt and F. vanWinden (2006): Exploring group decision making in a power-to-take experiment, *Experimental Economics* 9(1), 35-51.

such as price leadership or models of endogenous timing in oligopoly markets.

- [9] Cason, T. and V. Mui (1997): A laboratory study of group polarization in the team dictator game, *Economic Journal* 107, 1465-1483.
- [10] Cooper, D. and J. Kagel (2005): Are two heads better than one? Team versus individual play in signaling games, *American Economic Review* 95, 477-509.
- [11] Cox, J. (2002): Trust, reciprocity and other-regarding preferences: Group vs. individuals and males vs. females, In Zwick, R., A. Rapoport eds., *Advances in Experimental Business Research*, 331-350, Kluwer Academic Publishers, Dordrecht.
- [12] Cox, J. and S. Hayne (2006) : Barking up the right tree: Are small groups rational agents?, *Experimental Economics* 9(3), 209-222.
- [13] Cox, J., D. Friedman and S. Gjerstad (2007): A tractable model of reciprocity and fairness, *Games and Economic Behavior* 59(1), 17-45.
- [14] Cox, J., D. Friedman and V. Sadiraj (2008): Revealed altruism, *Econometrica* 76 (1), 31-69.
- [15] Elbittar, A., A. Gomberg and L. Sour (2004): Group decision-making in ultimatum bargaining: An Experimental study, Working Papers 0407, Centro de Investigacion Economica, ITAM.
- [16] Engel, C. (2010): The behaviour of corporate actors: How much can we learn from the experimental literature?, *Journal of Institutional Economics*, 6: 445-475.
- [17] Fehr, E. and K.M. Schmidt (1999): A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics* 117(3):817-68.
- [18] Feri, F., B. Irlenbusch and M. Sutter (2010): Efficiency gains from team-based coordination – Large-scale experimental evidence, *American Economic Review* 100: 1892-1912.
- [19] Finkelstein, S. and D.C. Hambrick (1996): *Strategic Leadership: Top Executives and their Effects on Organizations*, West Publishing Company St. Paul, MN.
- [20] Fischbacher, U. (2007): Z-tree - Zurich toolbox for readymade economic experiments, *Experimental Economics* 10(2), 171-178.
- [21] Fischbacher, U., S. Gächter and E. Fehr (2001): Are people conditionally cooperative? Evidence from a public goods experiment, *Economics Letters* 71(3), 397-404.

- [22] Fisher, B. (1980): *Small Group Decision Making: Communication and the Group Process*, 2nd edition, McGraw-Hill Press, New York.
- [23] Gillet, J., A. Shram and J. Sonnemans (2009): The tragedy of the commons revisited: The importance of group decision-making, *Journal of Public Economics* 93, 785-797.
- [24] Gillet, J., A. Shram and J. Sonnemans (2011): Cartel formation and pricing: The effect of managerial decision making rules, *International Journal of Industrial Organization* 29, 126-133.
- [25] Holt, C., (1985): An experimental test of the consistent-conjectures hypothesis, *American Economic Review* 75(3), 314-325.
- [26] Huck, S., W. Müller and H. T. Normann (2001): Stackelberg beats Cournot: On collusion and efficiency in experimental markets, *Economic Journal* 111, 749-65.
- [27] Huck, S., H.-T. Normann and J. Oechssler (2004): Two are few and four are many: number effects in experimental oligopoly, *Journal of Economic Behavior & Organization* 53, 435-446.
- [28] Huck, S. and B. Wallace (2002): Reciprocal strategies and aspiration levels in a Cournot-Stackelberg experiment, *Economics Bulletin* 3(3), 1-7.
- [29] Kaufman, L. and P. J. Rousseeuw (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley.
- [30] Kocher, M. and M. Sutter (2005): The decision maker matters. Individual versus team behavior in experimental beauty-contest games, *Economic Journal* 115, 200-223.
- [31] Kocher, M. and M. Sutter (2007): Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments, *Empirica* 34, 63-88.
- [32] Sutter, M., M. Kocher and S. Strauss (2009): Individuals and teams in auctions, *Oxford Economic Papers* 61, 380-394.
- [33] Kugler, T., G. Bornstein, M. Kocher and M. Sutter (2007): Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy, *Journal of Economic Psychology* 28(6), 646-657.

- [34] Lau, S.-H. P and F. Leung (2010): Estimating a parsimonious model of inequality aversion in Stackelberg duopoly experiments, *Oxford Bulletin of Economics and Statistics*, 72(5), 669–686.
- [35] Levine, J. and R. L. Moreland (1998): Small groups. In: Fisher, G., Lindzey, G.(eds.), *The Handbook of Social Psychology*, Vol. II, 4th Edition, pp. 415–469, Oxford, University Press, Oxford.
- [36] Lodewijkx, H., J. Rabbie and L. Visser (2006): “Better to be safe than to be sorry”: Extinguishing the individual - group discontinuity effect in competition by cautious reciprocation, *European Review of Social Psychology* 17(1), 185 - 232.
- [37] Luhan,W., M. Kocher and M. Sutter (2009): Group polarization in the team dictator game reconsidered, *Experimental Economics* 12(1), 26-41.
- [38] Milgrom, P. and J. Roberts (1982): Limit pricing and entry under incomplete information: An equilibrium analysis, *Econometrica* 50, 443-59.
- [39] Raab, P. and B. Schipper (2009): Cournot competition between teams: An experimental study, *Journal of Economic Behavior & Organization* 72(2), 691-702.
- [40] Rabe-Hesketh, S. and Skrondal, A. (2005). *Multilevel and Longitudinal Modeling using STATA*, STATA Press, College Station, Texas.
- [41] Reuben, E. and S. Suetens (in press): Revisiting strategic versus non-strategic cooperation, *Experimental Economics*, forthcoming.
- [42] Robert, C. and P. Carnevale (1997): Group choice in ultimatum bargaining, *Organizational Behavior and Human Decision Processes* 72 (2), 256-279.
- [43] Selten, R. (1967): Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopolexperiments, in Heinz Sauer mann, ed., *Beiträge zur experimentellen Wirtschaftsforschung*. Tübingen, Germany: Mohr, 136-68.
- [44] Selten, R. and R. Stoecker (1986): End behavior in finite prisoner’s dilemma supergames, *Journal of Economic Behavior and Organization* 1, 47-70.
- [45] von Stackelberg, H. (1934): *Marktform und Gleichgewicht*, Vienna: Julius Springer.

- [46] Stoner, J. (1961): A comparison of individual and group decisions involving risk, Unpublished master's thesis, Massachusetts Institute of Technology, Boston.
- [47] Teger, A. and D. Pruitt (1967): Components of group risk taking, *Journal of Experimental Social Psychology* 3, 189-205.
- [48] Wildschut, T. and C. Insko (2007): Explanations of interindividual - intergroup discontinuity: A review of the evidence, *European Review of Social Psychology* 18(1): 175 - 211.