



# Who acts more like a game theorist? Group and individual play in a sequential market game and the effect of the time horizon <sup>☆</sup>

Wieland Müller <sup>a,b,\*</sup>, Fangfang Tan <sup>c</sup>

<sup>a</sup> VCEE & University of Vienna, Vienna, Austria

<sup>b</sup> CentER, TILEC & Tilburg University, Tilburg, The Netherlands

<sup>c</sup> Max Planck Institute for Tax Law and Public Finance, Munich, Germany

## ARTICLE INFO

### Article history:

Received 19 May 2012

Available online 15 October 2013

### JEL classification:

C72

C92

L13

### Keywords:

Stackelberg market

Groups versus individuals

Discontinuity effect

Experiment

## 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 intergroup versus interindividual decision-making by running both one-shot and repeated sessions of a simple two-player sequential market game (Stackelberg duopoly). Whereas in one-shot markets we find no significant differences in the behavior of groups and individuals, in repeated markets 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), the matching protocol (random- or fixed-matching), and the econometric method used to account for observed first- and second-mover behavior. We discuss various possible explanations for the differential effect that the time horizon of interaction has on the extent of individual and group players' (non)conformity with subgame perfectness.

© 2013 Elsevier Inc. All rights reserved.

## 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. However, much of economic theory does not distinguish between decisions taken by individuals or groups. Moreover, until recently, experimental economists were mainly concerned with testing economic models by employing 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).

<sup>☆</sup> We are grateful to the associate editor and two anonymous referees for helpful feedback on our paper. We thank Sau-Him Paul Lau and Felix Leung, as well as James C. Cox, Daniel Friedman and Steven Gjerstad for making available estimation codes, and David Vonka for technical consulting and help. We also thank Marco Castillo, Dirk Engelmann, 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. Wieland Müller acknowledges the financial support from the Netherlands Organisation for Scientific Research (NWO) through a VIDI grant.

\* Corresponding author at: Department of Economics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria.

E-mail addresses: [wieland.mueller@univie.ac.at](mailto:wieland.mueller@univie.ac.at) (W. Müller), [fangfang.tan@tax.mpg.de](mailto:fangfang.tan@tax.mpg.de) (F. Tan).

One result that emerges from the recent experimental literature on interindividual–intergroup comparisons is that often, groups appear to be more selfish than individuals. This has mainly been shown in the context 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; Song, 2006, and Kugler et al., 2007), the centipede game (Bornstein et al., 2004), and the gift-exchange game (Kocher and Sutter, 2007) and the Stackelberg game (Cardella and Chiu, 2012). 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.” Similarly, in their more recent review, Kugler et al. (2012) stated “Our review suggests that results are quite consistent in revealing that groups behave closer to the game-theoretical assumption of rationality and selfishness than individuals.”<sup>1</sup>

Note that the literature Bornstein (2008) summarizes in his quote (and, to a lesser extent the study reviewed in Kugler et al., 2012) is based on experimental two-stage games in which individuals and groups interact only once. But what if such a game is played repeatedly? Will interacting groups still have a tendency towards more selfish behavior in comparison with interacting individuals, as suggested by the earlier literature? Or will there be a trend towards more cooperation in intergroup interaction, as this, in the longer run, promises higher profits? That is, will groups be better than individuals at achieving higher payoffs through cooperation in repeated interactions, as suggested, for instance, in the psychological literature (see, e.g., Rabbie, 1998; Lodewijkx et al., 2006, or Meier and Hinsz, 2004)?

In this paper, we study the behavior of groups and individuals in a simple two-stage market game (a Stackelberg duopoly) in both one-period and multiple-period experiments. Our results are in (partial) contrast to the quotes given above. In fact, in our one-shot markets we find no significant differences between the behavior of groups and individuals. However, and more importantly, in our repeated markets we find that the behavior of groups is *further away* from the subgame-perfect equilibrium than that of individuals. More precisely, in the repeated markets, the average leader quantities chosen by groups are often significantly lower than the average leader quantities chosen by individuals, and follower groups punish leader groups who are “greedy” harder, and reward leader groups who behave collusively more than individual followers. That is, we show that once a simple sequential-move game is repeated, the behavior of groups relative to that of individuals goes in the opposite direction to that stated in Bornstein’s summary. Moreover, in our repeated markets, group play *diverges* from the (refined) game-theoretic solution.

The Stackelberg (1934) model is among the most frequently applied models of oligopolistic competition, featuring a first- and a second-mover who compete in quantities.<sup>2</sup> 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 their own and the other player’s income (Cox et al., 2007).<sup>3</sup> 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 second-movers. Subjects acting in groups have to agree unanimously on the quantity produced. The decision-making process within the 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.

Since individuals and groups partly choose markedly different quantities as first-movers, the 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 decisions, 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 the 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, the behavior appears to be in line with the 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 best-response function.

Our paper makes two main contributions. The literature has reported, 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 it is repeated, the result is turned around in the sense that groups are shown to be *further away* from the game-theoretic prediction. Regarding a first possible explanation of our results and those reported in the literature, we note that the Stackelberg market game, like other sequential

<sup>1</sup> Moreover, Charness and Sutter (2012, p. 173) state that “The existing literature that compares group and individual decision-making provides considerable evidence that groups make choices that are more rational in a standard game-theoretic sense than those of individuals.” Additional evidence comes from 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 (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.

<sup>2</sup> 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. Experimental evidence on individual-player Stackelberg duopoly markets and how they compare to simultaneous-move Cournot duopoly markets is reported in Huck et al. (2001).

<sup>3</sup> This feature distinguishes the Stackelberg game from other sequential games such as the ultimatum game or the trust game.

games, leaves 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, in the discussion section we provide evidence that there is heterogeneity in the subjects' types (e.g., *myopic profit maximizers* or (*non*)*strategic rewarders and punishers*).<sup>4</sup> 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 own and earlier experiments.

Other important issues that we discuss to help explain our experimental results are the possible differences in the belief formation of individuals and groups and their interaction with the time horizon of play, and differences in the ability to employ repeated-game strategies for individuals and groups. To support the latter explanation, we also report the results of four additional repeated fixed-matching treatments in Section 5. Overall, our results suggest that the apparent consensus in the literature regarding sequential two-player games, as summarized above, might need modification to accommodate the differential effects of the time horizon of interaction and, possibly, other design features – a point we discuss in more detail in the concluding section. In any case, the answer to the question of who behaves more like a game theorist – groups or individuals – seems not be 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 (both with random- and fixed-matching) and employ the strategy method for the first time in a “group” experiment and in a repeated Stackelberg market game.<sup>5</sup> 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 the *average* response functions in both the individual and the team treatments exhibit 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 might be inappropriate to account for response functions in, e.g., sequential market games by running simple linear regressions. As other authors and ourselves demonstrate, the structural estimations 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.<sup>6</sup> As the standard myopic best-response function of followers is nested in both of the other-regarding preference models that we structurally estimate (Fehr and Schmidt, 1999; Cox et al., 2007), we have a clear and unambiguous method for formally testing which of the 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 best-response function than that of individuals.

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 hypotheses. 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 report on additional fixed-matching treatments. Section 6 provides a summary and offers some concluding remarks.

## 2. Related literature

There is now a considerable number of studies comparing the 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. In doing so, we only very briefly describe the main results of these studies while providing the design details of the most relevant studies in Table 6 in Section A of the Web Appendix. Bornstein (2008), Engel (2010) and Kugler et al. (2012) provide more detailed 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, while 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 to Bornstein and Yaniv (1998) with respect to proposers.

Subsequent studies replicate this finding in other games. Cox (2002) analyzes 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. Song (2006) reports similar findings. 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 roles of employer or employee choose lower wages and, in return,

<sup>4</sup> These types are reminiscent of the types in public-good games identified in Fischbacher et al. (2001) or, more recently, Reuben and Suetens (2012).

<sup>5</sup> Huck and Wallace (2002) elicit complete-response functions in a one-shot Stackelberg experiment. However, we will show that the behavior these authors and ourselves 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.

<sup>6</sup> Note that the 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 main treatments to weaken other-regarding motives.

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 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. [Cardella and Chiu \(2012\)](#) compare the decisions of groups and individuals in a one-shot Stackelberg game. They find that groups choose smaller quantities than individuals as leaders, but similar quantities as followers. Note that in contrast to our design, their design features one-shot interaction only, and their groups consist of two subjects who communicate face-to-face.

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 the Bertrand price competition between individuals and between groups. They find that on average 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 the behavior of individuals or groups in the Cournot competition. Note 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 all, 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 \(2008\)](#), [Sutter et al. \(2009\)](#), and [Kugler et al. \(2012\)](#) 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 the inverse demand function  $p = \max\{30 - Q, 0\}$  where  $Q$  denotes the total quantity. Both players have constant unit costs of  $c = 6$  and no fixed costs. 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 contains 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.

#### 3.2. Experimental design

Our main experiments are 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 that employ the strategy method are called "SM-IND-1" and "SM-TEAM-1". The corresponding 15-period treatments are, respectively, called, "SEQ-IND-15-RM", "SEQ-TEAM-15-RM", "SM-IND-15-RM", and "SM-TEAM-15-RM", where "RM" indicates that we employed random-matching of the first- and second-movers across periods. [Table 1](#) gives an overview of the design. This table also lists 15-period treatments with the suffix "FM" standing for fixed-matching. These are additional control treatments that we introduce later in [Section 5](#). Information about profits was given in the form of a payoff table (see [Table 17](#) in the Web Appendix). Next, we describe the setting in each of the four basic experimental situations 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).

**Table 1**  
Experimental design.

Treatment name	Sequential method	Strategy method	Individual players	Team players	Matching protocol	# 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-RM	Yes	No	Yes	No	Random	15	36	6
SEQ-TEAM-15-RM	Yes	No	No	Yes	Random	15	72	4
SM-IND-15-RM	No	Yes	Yes	No	Random	15	36	6
SM-TEAM-15-RM	No	Yes	No	Yes	Random	15	72	4
SEQ-IND-15-FM	Yes	No	Yes	No	Fixed	15	18	9
SEQ-TEAM-15-FM	Yes	No	No	Yes	Fixed	15	36	6
SM-IND-15-FM	No	Yes	Yes	No	Fixed	15	18	9
SM-TEAM-15-FM	No	Yes	No	Yes	Fixed	15	36	6

*Treatments SEQ-TEAM:* These are the team baseline treatments which were, with respect to timing, identical to the SEQ-IND treatments except that the players were teams consisting of three participants each instead of individuals. To reach a joint decision, members of a team could exchange messages among themselves via an electronic chat box.<sup>7</sup> There were no restrictions regarding the content of the messages sent, except that: (a) the discussion must be in English; (b) the language used should be civil; and (c) the subjects cannot identify themselves by revealing their names, seat numbers, etc. The 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 the submitted quantity decisions of each group member then appeared on the screen of every other member of the same group. As long as not all the submitted quantity decisions were the same, the chat box remained open and the group members could continue discussing their decision. When all the submitted quantity decisions of a team's members were the same, the decision screen (including the chat box) disappeared and the 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 the subjects had made their decisions, the computer matched first- and second-movers, and selected the relevant quantity of the second-mover (i.e., 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 the 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 the group members had not yet entered the same complete strategy.<sup>8</sup>

### 3.3. Experimental procedures

The experiment with 24 sessions was conducted at CentERlab of Tilburg University in April, May, October 2009, September 2010, and October 2012. Each session consisted of 18 subjects. A total of 432 Tilburg University students participated in the study. Each subject took part in just one session. Each session consisted of either 1 period or 15 periods. In the repeated sessions, all 15 periods of play counted towards the final earnings. There were no practice periods at the beginning of any session. On average, a one-shot session lasted around 45 minutes, whereas the repeated sessions lasted around one 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, the 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 subjects, who then belonged to the same team for the entire experiment. Hence, a team-treatment session consisted of three first-mover teams and three second-mover teams, which were randomly re-matched with each other during each of

<sup>7</sup> Electronic chat helps maintain 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 influence of other attributes (e.g., physical attractiveness) on communication.

<sup>8</sup> Although agreeing on a complete-response function sounds like a formidable task, it usually took teams not more than five minutes to complete it.



the 15 periods of the main repeated-game treatments. In order to control for the size of the random-matching group, the 18 subjects in an individual-player, 15-period session were divided into three matching groups of six subjects each (three first- and three second-movers), and the random re-matching of first- and second-movers across periods happened only within matching groups. This was explained in the instructions.<sup>9</sup> The number of matching groups (i.e., independent observations) is indicated in Table 1.

The instructions (see Web Appendix, Section L) used non-neutral language, referring, e.g., to “firms”, “product”, or “profits”. With the instructions, subjects received a payoff table, 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.<sup>10</sup> 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 repeated-game treatments, the 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 their own profits.

### 3.4. Hypotheses

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 random-matching main treatments viewed the experiment as a finitely repeated game.

Earlier studies on simple sequential one-shot games, as reviewed in Section 2, typically found groups to be closer to the game-theoretic prediction than individuals. Also, in repeated “Eureka” kind problems, earlier studies report that groups 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, 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 1a.** Independent of the duration of the interaction, 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.

Note that psychologists’ research on interindividual versus intergroup comparison, preceding that of economists, finds that “In a series of previous studies, the authors have shown that intergroup interactions are dramatically more competitive and less cooperative than individual interactions. This phenomenon has been termed a *discontinuity effect*” (Schopler et al., 1991, p. 612, first emphasis added). Hypothesis 1a is in line with the “discontinuity effect”.<sup>11</sup>

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 the subjects are randomly re-matched 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? That is, will groups be better than individuals at achieving higher payoffs through cooperation?

Support for the latter line of reasoning can also be found in the psychology literature. In fact, some psychologists dispute the general validity of the “discontinuity effect” that predicts groups to be less cooperative than individuals. In fact, Rabbie (1998) and Lodewijckx et al. (2006) [henceforth LRV] put forward the “cautious reciprocation model”, suggesting that in the context of repeated interaction “group members will realize that “enlightened” long-term between-group cooperation is the best way to instrumentally achieve their group’s goal (maximizing monetary outcomes)” (LRV, p. 197). Moreover they argue that “stronger rationality/instrumentality would explain a discontinuity effect in non-iterated games, but it would also predict the opposite effect in iterated games, assuming that mutual cooperation is considered the rational and instrumental solution in such games” (LRV, p. 192). Lodewijckx et al. (2006) also point out that groups might be quicker in retaliating

<sup>9</sup> Random-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. As mentioned, in Section 5 we also discuss the results of four additional 15-period fixed-matching treatments.

<sup>10</sup> Due 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 the uniqueness of both the Cournot–Nash equilibrium and the subgame-perfect Stackelberg equilibrium.

<sup>11</sup> Wildschut and Insko (2007) review the evidence for the discontinuity effect.

norm violations. Indeed, they explain that “research, using mixed-motive situations such as the PDG, convincingly showed that group members retaliated more aggressively against another party (an outgroup), that deliberately transgressed the cooperative norm of reciprocity that regulated the exchange relationship” (p. 193).<sup>12</sup> This line of reasoning suggests:

**Hypothesis 1b.** In the multiple-period treatments, group first-movers will choose quantities that are *lower* than the Stackelberg leader quantity and *lower* than those of individual first-movers, and group second-movers’ response functions will be *further away* from the standard best-response function (more reward and punishment) than that of individual second-movers.

#### 4. Experimental results

We report the results in two sections with the purpose of comparing the 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 on formal tests for differences in second-mover behavior in the 15-period treatments. Here, we will deal with the 1-period and the 15-period random-matching treatments only. We report the results of the additional 15-period fixed-matching treatments in Section 5. To allow for 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, unless otherwise indicated.

##### 4.1. A first look at the data

Table 2 presents summary statistics of the average quantity choices, payoffs and a cooperation index for each treatment.<sup>13</sup>

In all the treatments, we note that the average first-mover quantities are clearly smaller and that the average second-mover quantities are 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-RM, 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 experiments we observe that average leader quantities in the individual treatments are *higher* than in the corresponding group treatments. To test for the significance of differences in the first-mover data, we ran regressions of the form  $q_{ijt}^L = \beta_0 + \beta_1 \times TREATM + \eta_i + \eta_{ij} + \varepsilon_{ijt}$  where  $q_{ijt}^L$  is the quantity chosen by leader subject/group  $i$  in matching group  $j$  in period  $t$ , and  $TREATM$  is the dummy to code the two treatments that are included in the regression. The coefficient  $\beta_1$  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. We ran the regressions using general linear latent and mixed models GLLMM (Rabe-Hesketh and Skrondal, 2005). In the regressions, we take into account that subjects and groups are nested in matching groups by including nested random effects, which are assumed to be independently normally distributed (cf.  $\eta_i$  and  $\eta_{ij}$ ). To test for differences in total payoffs and the cooperation index, we used similar regressions. The results are reported in Table 3,<sup>14</sup> where the main comparisons 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 leader quantities between individual- and group-player treatments are significant in the 1-period treatments. However, first-movers in treatment SEQ-IND-15-RM choose significantly higher quantities than first-movers in the corresponding team treatment SEQ-TEAM-15-RM. This contradicts Hypothesis 1a and supports Hypothesis 1b. Note that average first-mover choices in treatment SM-IND-15-RM and the corresponding team treatment SM-TEAM-15-RM do not differ significantly.

##### 4.1.2. Second-mover behavior

Let us first consider second-mover behavior in the 1-period treatments. Fig. 1 shows the average response function observed in these treatments (for the sequential-play treatments in the left panel and for the strategy-method treatments in the right panel). Although there is weak visual evidence indicating that the observed response functions of the teams are closer to the best-response function than those of individual players in the 1-period treatments (which is in line with earlier results in the literature and our Hypothesis 1a), the estimation of simple linear response functions does not deliver

<sup>12</sup> Additional support for this finding is reported in, e.g., Meier and Hinsz (2004).

<sup>13</sup> The cooperation index is defined as  $(\pi^{\text{Observed}} - \pi^{\text{SPNE}}) / (\pi^{\text{Cartel}} - \pi^{\text{SPNE}}) = (\pi^{\text{Observed}} - 108) / 36$ , where  $\pi^{\text{Observed}}$ ,  $\pi^{\text{SPNE}}$ , and  $\pi^{\text{Cartel}}$  are, respectively, the total profits observed, the total profits in the subgame-perfect Nash equilibrium, and the total profits in the cartel solution. The cooperation index is 0 (1) in the subgame-perfect equilibrium (cartel) outcome. The test results reported below do not change if we replace  $\pi^{\text{SPNE}}$  by total profits in the Cournot outcome.

<sup>14</sup> Using the Tobit regression techniques delivers very similar results.

**Table 2**  
Summary of experimental results.

	Prediction		Sequential play				Strategy method			
	Leader	Follower	SEQ-IND		SEQ-TEAM		SM-IND		SM-TEAM	
			Leader	Follower	Leader	Follower	Leader	Follower	Leader	Follower
<b>1-period treatments</b>										
Individual quantities	12	6	9.11 (0.77)	8.11 (0.70)	9.33 (0.99)	7.67 (0.42)	9.67 (1.00)	7.11 (0.45)	10.67 (0.84)	7.00 (0.45)
Total quantities		18	17.22 (0.62)		17.00 (0.73)		16.78 (0.55)		17.67 (0.61)	
Individual payoffs	72	36	59.78 (4.64)	53.89 (5.44)	62.00 (3.79)	54.33 (7.66)	65.44 (4.20)	53.33 (8.06)	65.33 (3.82)	44.67 (6.15)
Total payoffs		108	113.67 (6.06)		116.33 (8.20)		118.78 (4.12)		110 (7.21)	
Cooperation index		0	0.16 (0.17)		0.23 (0.23)		0.19 (0.18)		0.05 (0.12)	
<b>15-period random-matching treatments</b>										
Individual quantities	12	6	10.37 (0.19)	7.77 (0.22)	8.41 (0.13)	7.99 (0.07)	9.33 (0.30)	7.87 (0.22)	8.73 (0.19)	8.16 (0.15)
Total quantities		18	18.14 (0.26)		16.39 (0.16)		17.21 (0.33)		16.89 (0.27)	
Individual payoffs	72	36	57.59 (2.39)	43.31 (1.69)	60.54 (0.87)	59.23 (1.22)	58.55 (2.39)	50.97 (2.76)	59.09 (1.38)	55.97 (1.44)
Total payoffs		108	100.91 (3.49)		119.77 (2.00)		109.52 (4.53)		115.07 (2.54)	
Cooperation index		0	−0.20 (0.10)		0.33 (0.06)		0.04 (0.13)		0.20 (0.07)	
<b>15-period fixed-matching treatments</b>										
Individual quantities	12	6	8.19 (0.44)	7.62 (0.39)	6.71 (0.14)	6.67 (0.19)	7.73 (0.62)	7.28 (0.30)	7.34 (0.51)	7.16 (0.37)
Total quantities		18	15.81 (0.65)		13.38 (0.33)		15.01 (0.66)		14.5 (0.88)	
Individual payoffs	72	36	62.17 (2.98)	58.8 (2.83)	67.86 (0.83)	68.12 (0.33)	63.09 (2.32)	61.33 (3.97)	64.73 (2.40)	64.03 (3.14)
Total payoffs		108	120.97 (4.35)		135.98 (1.14)		124.42 (4.51)		128.77 (5.49)	
Cooperation index		0	0.36 (0.12)		0.78 (0.03)		0.46 (0.13)		0.58 (0.15)	

Notes: The results of the 1-period (15-period) treatments are presented in the upper (lower) part of this table. For the strategy-method treatments, only the relevant quantities of the second-movers are taken into account (i.e., only those quantity choices of second-movers at quantities actually chosen by first-movers). Standard errors of the mean are in parentheses.

any statistically significant differences, neither with respect to the intercept nor to the slope. (For details of this analysis, see Subsection B.2 in the Web Appendix.)

Next, we turn to second-mover behavior in the 15-period treatments. The two panels in Fig. 2 show the average response functions in the 15-period truly sequential (left panel) and the 15-period strategy-method treatments (right panel). Inspecting the two panels of Fig. 2, it seems fair to say 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 Fig. 2 as well as simple regressions suggest that team second-movers reward more and punish harder than individual followers.<sup>15</sup> Interestingly, all the observed response functions show a peculiar and somewhat surprising “first slope downward, then upward, then 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 downward-sloping again for higher leader

<sup>15</sup> Recall 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 the Web Appendix, Section B): First, both the intercept and the slope of the response function employed in the individual-player treatment SEQ-IND-15-RM are significantly closer to those of the rational best-response function than the intercept and slope of the response function in the team-player treatment SEQ-TEAM-15-RM. This suggests that individual second-movers behave more selfishly than team second-movers. Second, the reaction function in treatment SEQ-IND-15-RM is downward-sloping, while the reaction function in treatment SEQ-TEAM-15-RM 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.

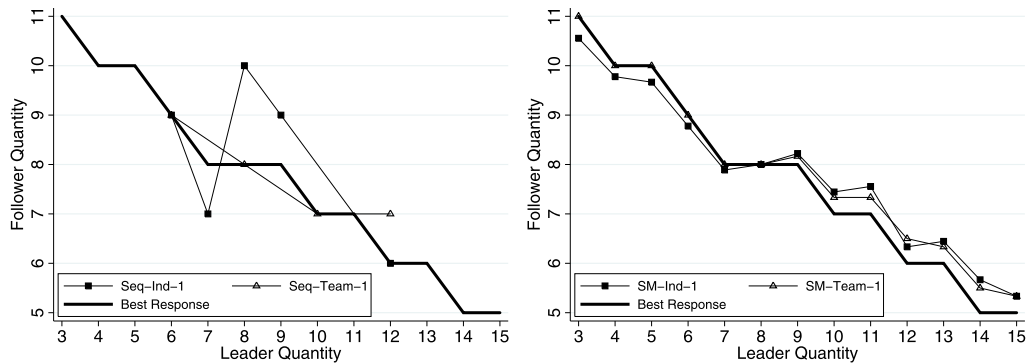


**Table 3**

Results of parametric tests for differences in choices and outcomes. (Estimates for the coefficient  $\beta_1$ .  $H_0: \beta_1 = 0$ .)

	Comparison based on player types		Comparison based on elicitation method	
	SEQ-IND versus SEQ-TEAM	SM-IND versus SM-TEAM	SEQ-IND versus SM-IND	SEQ-TEAM versus SM-TEAM
<b>1-period treatments</b>				
Leader quantity <sup>a</sup>	-0.22 (1.45)	-1.00 (1.31)	-0.56 (1.19)	-1.33* (0.69)
Total quantity	0.22 (0.96)	-0.89 (0.84)	0.67 (0.95)	-0.44 (0.83)
Total profits	-2.67 (9.29)	0.33 (7.91)	-1.13 (8.23)	1.17 (10.99)
Cooperation index	-0.07 (0.26)	0.01 (0.22)	0.17** (0.06)	0.05 (0.31)
<b>15-period random-matching treatments</b>				
Leader quantity <sup>a</sup>	2.62*** (0.21)	0.91 (0.24)	0.94*** (0.21)	-0.37 (0.30)
Total quantity	1.75*** (0.42)	0.39 (0.52)	0.86* (0.5)	-0.49 (0.32)
Total profits	-20.41*** (5.65)	-5.60 (6.87)	-8.65 (7.09)	5.69* (3.48)
Cooperation index	-0.57*** (0.16)	-0.16 (0.19)	-0.24 (0.20)	0.16* (0.10)
<b>15-period fixed-matching treatments</b>				
Leader quantity <sup>a</sup>	1.22*** (0.25)	1.03*** (0.26)	1.00*** (0.23)	-0.77 (0.49)
Total quantity	1.61*** (0.48)	0.05 (0.75)	9.99*** (0.88)	8.68*** (0.73)
Total profits	-15.95*** (5.00)	-7.51 (6.39)	-3.16 (6.05)	5.49 (4.44)
Cooperation index	-0.50*** (0.09)	-0.21 (0.18)	-0.09 (0.17)	0.15 (0.12)

Notes: <sup>a</sup> Estimated equation:  $q_{ijt}^L = \beta_0 + \beta_1 \times TREATM + \eta_i + \eta_{ij} + \varepsilon_{ijt}$ , 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. For the total quantities, total profits and the cooperation index, we used similar regressions. In all the regressions, the dummy variable  $TREATM$  is coded such that it is equal to 1 (0) for the treatment mentioned in the upper (lower) entry in each column of this table. We report as  $p$ -levels  $P > |t|$ . \*\*\*, \*\*, \* indicate significance at the 1%, 5% and 10% levels. For completeness, in columns 3 and 4 of this table we also report test results based on the elicitation method. Standard errors are in parentheses.



**Fig. 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.

choices.<sup>16</sup> Due to the more limited number of different choices for first-movers in the sequential treatments, this pattern is less clear in the left panel of Fig. 2.

4.1.3. A closer look at follower data: structural estimations

While the estimation of linear and monotonic response functions may serve as a quick diagnostic tool (see footnote 15), from the preceding discussion we conclude that simple linear estimations are inappropriate and incapable of accounting

<sup>16</sup> The basic pattern and the relationship of the two response functions to each other are already clearly visible in the first period of the strategy-method treatments. More details on this and the evolution of first- and second-movers' behavior over time can be found in the Web Appendix, Section C.

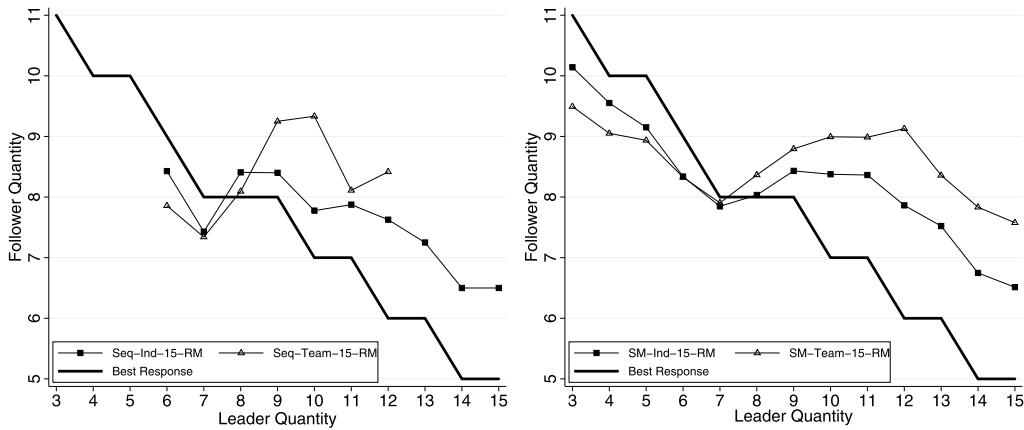


Fig. 2. Average response functions observed in the 15-period sequential treatments (left) and the 15-period strategy-method treatments (right).

for the patterns observed in the average and individual response functions. Furthermore, although basic patterns are easily identifiable at the individual and team levels in the repeated strategy-method treatments, this is not easy in the repeated sequential treatments since in these 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. It turns out that the patterns observed in Fig. 2 (and at the individual and group level) are consistent with the predictions of models of other-regarding preferences, especially the model suggested by Fehr and Schmidt (1999). Therefore, in this section we will account for the 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 the Cox et al. (2007) model of emotion-driven reciprocity. The important result of these estimations is that, irrespective of the model we estimate, individuals appear to be more selfish than teams. We are able to make this assertion as the standard selfish best-response function is nested in both of the social-preference models that we estimate. Therefore, we have a clear and unambiguous method to formally test which of the two observed average response functions is closer to the prediction of subgame perfectness: it is the function whose estimated preference parameters are closer to those representing the standard selfish best-response function. We must stress that, given the specific non-monotonic shapes of the observed response functions of groups and individuals (which are in line with the predictions of, e.g. the Fehr and Schmidt (1999) model), we employ the structural estimation of these preference models as an econometric technique in order to adequately estimate and compare the response functions.<sup>17</sup>

**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 by 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 non-standard types. It turns out that this response function accurately predicts the shape of the average response function that we observe in our 15-period sessions (see Fig. 2). Lau and Leung then develop a maximum-likelihood model in which a share  $\phi_{ns}$  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), Lau and Leung 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 those of both the standard or non-standard types provides a rationale to apply Lau and Leung’s model to our data so as to account for follower behavior. In the following, we will briefly introduce the model put forward by Lau and Leung, while closely following their exposition. We will then estimate it for the four 15-period random-matching treatments.

Denote player  $i$  and  $j$ ’s payoffs by  $\pi_i$  and  $\pi_j$ , respectively. Then, Fehr and Schmidt preferences are given by

$$u_i = \pi_i - \alpha_i \max\{\pi_j - \pi_i, 0\} - \beta_i \max\{\pi_i - \pi_j, 0\}, \tag{1}$$

<sup>17</sup> Clearly, 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 for the group treatments as well, 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).

where  $0 \leq \beta_i < 1$ ,  $\beta_i \leq \alpha_i$ ,  $i, j = L, F$  with  $i \neq j$ . The parameter  $\alpha_i$  measures player  $i$ 's aversion towards disadvantageous inequality, whereas the parameter  $\beta_i$  measures player  $i$ 's aversion towards advantageous inequality. For estimation purposes, Lau and Leung make two assumptions. First, 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 second-movers 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  $\phi_{ns}$  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  $A = [3, 12(\frac{1-b}{3-2b})]$ ,  $B = [12(\frac{1-b}{3-2b}), 12(\frac{1+a}{3+2a})]$ , and  $C = [12(\frac{1+a}{3+2a}), 15]$ . 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 first-mover quantities, slopes upward for intermediate ones, and slopes downward again for high first-mover quantities. Hence, it predicts the pattern observed in Fig. 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 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  $\varepsilon_i$  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  $\phi_{ns}$  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 (see Lau and Leung, 2010, p. 678). The log likelihood function of observing the sample  $(x_i, y_i)_{i=1}^{N_{\text{Treatm}}}$  of leader and follower choices is then given by

$$\ln L(a, b, \phi_{ns}, \sigma; (x_i, y_i)_{i=1}^{N_{\text{Treatm}}}) = \sum_{i=1}^{N_{\text{Treatm}}} \ln \{ (1 - \phi_{ns}) f_S(y_i) + \phi_{ns} f_{NS}(y_i|x_i; a, b, \sigma) \},$$

where  $N_{\text{Treatm}}$  is the number of observations in the treatment under consideration. To control for the non-independence of observations, we cluster standard errors on the matching group level.

In an effort to first estimate the average response functions as shown in Fig. 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.<sup>18</sup>

We note that the parameter estimates of the inequality-aversion parameters  $a$  and  $b$  are significantly different from 0 in all the 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 model. Most importantly for the purpose of deciding which observed average response function is closer to the 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-RM the parameter  $a$  is estimated as 0.629, it is only 0.303 in treatment SEQ-IND-15-RM. This is in support of Hypothesis 1b and in contrast to Hypothesis 1a, according to which the observed response function of teams should be closer to the 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 the two relevant treatments comparisons, the parameters  $a$  and  $b$  are the same.<sup>19</sup>

<sup>18</sup> 10 out of 4998 choice pairs result in negative payoffs for 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 quantities 3 and 15, causing difficulties in achieving convergence.

<sup>19</sup> We apply the Wald test for testing parameter significance. First, we combine data from different treatments into a large, unrestricted model. Next, we put restrictions on the coefficients to see whether they are equal to zero.

**Table 4**  
 Estimation results for Lau and Leung’s (2010) implementation of the Fehr and Schmidt model (data from random-matching treatments).

	Truly sequential play		Strategy method			
	SEQ-IND-15	SEQ-TEAM-15	All data		Relevant data	
			SM-IND-15	SM-TEAM-15	SM-IND	SM-TEAM-15
$\phi_{ns}$	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*** (0.001)
$b$	0.216*** (0.029)	0.252*** (0.017)	0.192*** (0.062)	0.215** (0.090)	0.215*** (0.035)	0.400*** (0.001)
$\sigma$	1.506*** (0.164)	0.862*** (0.190)	1.485*** (0.200)	1.740*** (0.423)	1.658*** (0.263)	1.322*** (0.216)
LL	−427.864	−198.222	−4599.221	−3334.489	−436.902	−262.242
N	234	156	2535	1690	173	156
Hypothesis Testing	$a^{SEQ-IND-15} = a^{SEQ-TEAM-15}$ & $b^{SEQ-IND-15} = b^{SEQ-TEAM-15}$ $p = 0.075$ ( $\chi^2_{(2)} = 5.17$ )		$a^{SM-IND-15} = a^{SM-TEAM-15}$ & $b^{SM-IND-15} = b^{SM-TEAM-15}$ $p = 0.077$ ( $\chi^2_{(2)} = 5.14$ )		$a^{SM-IND-15} = a^{SM-TEAM-15}$ & $b^{SM-IND-15} = b^{SM-TEAM-15}$ $p < 0.001$ ( $\chi^2_{(2)} = 28.72$ )	

Note: Estimations for the case  $\phi_{ns} = 1$ .

**Table 5**  
 Estimation results for Lau–Leung’s implementation of the Fehr and Schmidt model (data from random-matching treatments).

	Truly sequential play		Strategy method			
	SEQ-IND-15	SEQ-TEAM-15	All data		Relevant data	
			SM-IND-15	SM-TEAM-15	SM-IND-15	SM-TEAM-15
$\phi_{ns}$	0.277* (0.162)	0.773*** (0.073)	0.276*** (0.081)	0.418*** (0.130)	0.279*** (0.087)	0.543*** (0.137)
$a$	1.713 (1.970)	0.949*** (0.181)	1.035*** (0.203)	1.479*** (0.463)	15.584 (13.648)	3.327*** (0.903)
$b$	0.383 (0.513)	0.470*** (0.041)	0.823*** (0.001)	0.828*** (0.001)	0.605*** (0.02)	0.476*** (0.013)
$\sigma$	1.094*** (0.140)	0.717*** (0.257)	0.818*** (0.123)	1.034*** (0.181)	0.968*** (0.122)	0.931*** (0.179)
LL	−426.523	−196.848	−4149.451	−3232.943	−377.509	−241.571
N	234	156	2325	1690	227	156
Hypothesis Testing	$\phi_{ns}^{SEQ-IND-15} = \phi_{ns}^{SEQ-TEAM-15}$ $p = 0.034$ ( $\chi^2_{(1)} = 8.55$ )		$\phi_{ns}^{SM-IND-15} = \phi_{ns}^{SM-TEAM-15}$ $p = 0.09$ ( $\chi^2_{(1)} = 2.68$ )		$\phi_{ns}^{SM-IND-15} = \phi_{ns}^{SM-TEAM-15}$ $p = 0.079$ ( $\chi^2_{(1)} = 3.08$ )	

Note: Estimations for the unrestricted model.

We next estimate the full model, dropping the restriction  $\phi_{ns} = 1$ , and concentrating on the estimated share of standard- and non-standard-type decisions in the two related treatments. The results are shown in Table 5. With the exception of treatment SEQ-IND-15-RM, the share  $\phi_{ns}$  of non-standard type decisions is estimated to be significantly larger than 0 in all treatments and ranges from about 0.27 in the individual treatments to 0.773 in treatment SEQ-TEAM-15-RM. More importantly for our purposes, the share of non-standard type decisions is estimated to be consistently higher in the team 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. This again is strong evidence against Hypothesis 1a (according to which groups are expected to be more in line with the predictions of subgame perfectness) but in support of Hypothesis 1b.

**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 et al., 2007). Clearly, reciprocity is also 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 Fig. 2. Therefore, as a robustness check of our finding that team second-movers are less myopic than individual second-movers in the 15-period treatments, we also estimated the model put forward by Cox et al. (2007), see Section D of the Web Appendix for details. 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 the group-player 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 Hypothesis 1a, but supports Hypothesis 1b.

## 5. Discussion: Potential explanations 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 random-matching 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 what could explain the different results in the one-shot (both in our own and in earlier experiments) and the repeated treatments. Under the headlines “heterogeneity in subjects’ types”, “beliefs”, and “repeated-game strategies”, we next discuss three issues that provide possible explanations of our findings.

*Heterogeneity in subjects’ types:* A first possible explanation rests on the observation that there is heterogeneity in the subjects’ types of behavior and that, in comparison to the interaction of individual players and depending on the time horizon of the interaction, the exchange of arguments via discussions is likely to lead groups to more selfish behavior in one-shot interactions and to more cooperative behavior in repeated interactions. This is what we explain in what follows.

Regarding the heterogeneity of the subjects’ types, 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, independently of the time horizon of interaction. Strat-R&Ps reward “nice” low leader quantities and punish “greedy” high leader quantities during every period but the final period, where they revert to 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 (2012) for the existence of Strat-R&Ps and Pref-R&Ps).<sup>20</sup>

Let us now 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 interindividual, one-shot treatments is likely to be a mixture of payoff-maximizing 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 intergroup 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 interindividual treatments, average behavior will be a mixture of other-regarding behavior (displayed by both Pref-R&Ps and Strat-R&Ps) and PMs. However, in the multiple-period team treatments, it is conceivable that Strat-R&P now side with Pref-R&Ps in an effort to convince the potentially present PMs 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 intergroup, multiple-round treatments is likely to be more homogeneous 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.<sup>21,22</sup>

We next provide evidence for the existence of the different types of subjects mentioned above. 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 probability that second-movers on average use standard best response. As this share is estimated to be

<sup>20</sup> Reuben and Suetens’s (2012) results are partly based on the work of Kreps et al. (1982), who show that players in a *finitely* repeated prisoners’ dilemma might cooperate strategically if they entertain the belief that rival players reciprocate cooperation for non-strategic reasons. As suggested by a referee, in our treatments with random-matching, the Kreps et al. (1982) argument might still hold because our matching groups are relatively small, such that there is a relatively high chance of meeting the same opponent in successive periods.

<sup>21</sup> Note 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.

<sup>22</sup> We believe, furthermore, that the mechanisms described here are also 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 or trust games).



significantly larger than 0, no matter which of the individual-treatment data sets we use, this provides (indirect) evidence for the existence of myopic profit maximizers.

The second, more direct evidence, is delivered by means of a cluster analysis of the individual response functions of second-movers in rounds 14 and 15, respectively, in treatment SM-IND-15-RM. Roughly, we find that seven out of 18 subjects can be classified as *PMs*, five out of 18 subjects as *Strat-R&P*, and two out of 18 subjects as *Pref-R&P*.<sup>23</sup> This provides further evidence for the existence of the types mentioned above.<sup>24</sup>

The third kind of evidence is provided by the analysis of chat protocols, in which we also find ample evidence for the kinds of subject types introduced above, and that a large part of the group discussions can be characterized as a conflict between these subject types (for details, see the Web Appendix, Section J).

*Beliefs:* Could beliefs play a role in explaining our results? More precisely, are groups better at predicting the behavior of rivals, enabling them to make more “appropriate” decisions? Clearly, in our Stackelberg game, second-movers observe the first-movers’ choices before they make their decisions, so that beliefs on the part of second-movers are irrelevant. However, first-movers need to predict the likely reaction of second-movers when making their quantity decisions. So, differences in the belief formation of individuals and groups acting in the role of the leader could, in principle, help explain our results. Indeed, in one-shot interactions it is more likely that at least one member of a leader group suggests that the second-mover group might play a best response, possibly persuading the leader group to choose a quantity that is on average higher than individual leaders would choose. In repeated interactions, it might be that groups are simply better at predicting the followers’ reactions (which in our data means harsher punishment for higher leader quantities and more rewarding for lower leader quantities), which might prompt leader groups to choose, on average, lower quantities than individuals. In Section E of the Web Appendix, we argue that leaders, to a large extent, appear to adjust to followers’ reactions. But this still leaves room for differences in belief formation, which might partly drive the differences in leader behavior that we observe, especially in our repeated SEQ treatments.

Unfortunately, we did not elicit beliefs in our experiment. However, the chats among the members of leader groups could provide hints as to what extent predicting followers’ reactions plays a role in their decision-making. Analyzing leader group chats, we find that groups do discuss followers’ likely reactions to certain quantity choices. In particular, in the early periods of the experiment, leader groups most commonly discussed the possibility of followers best responding. This, however, clearly becomes less frequent in leader group discussions over the course of the experiment. On the other hand, discussions about followers rewarding, exploiting or punishing possible leader quantity choices can be observed throughout the experiment. Moreover, there is clear evidence that “belief” discussions sharply decline after the first five periods, which is consistent with the idea that beliefs are formed early on in the experiment. (We provide details of this analysis in Section K of the Web Appendix.)

Not eliciting beliefs in our experiment and not having direct evidence on the thought processes of individuals prevents us from making definitive statements about possible differences in belief formation by individuals and groups. We think this is a worthwhile task for future research.

*Repeated-game strategies:* Recall that in our 15-period treatments there are matching groups consisting of three individual or group leaders and three individual or group followers who were randomly re-matched across periods. This implies that each individual or group plays any other rival individual or group, on average, five times. This might make some sort of repeated-game strategies promoting cooperation attractive, despite the fact that with a fixed and commonly known number of rounds, the unique subgame-perfect equilibrium is to play the unique subgame-perfect equilibrium of the stage game in each period of interaction. Still, in all of our repeated treatments, we see outcomes that are more cooperative than those predicted by subgame perfection (which is in line with, e.g., Selten and Stoecker, 1986, or Andreoni and Miller, 1993). Moreover, this, importantly, is true to a greater extent in the team treatments than in the corresponding individual treatments. This suggests that, perhaps, groups are better at understanding the advantages of cooperation-enhancing repeated-game strategies and are better at employing them.<sup>25</sup> However, would the greater extent of cooperation in team treatments when compared to individual treatments still prevail with fixed instead of random-matching? Although the literature – so far

<sup>23</sup> Note that in treatment SM-TEAM-15-RM there are only two out of 12 teams who can be classified as pure profit maximizers (teams 2 and 9), as shown in Figs. 9 and 10 in Section F of the Web Appendix. 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.

<sup>24</sup> The individual response functions in rounds 14 and 15 in treatment SM-IND-15-RM are shown in Figs. 7 and 8 in the Web Appendix, Section F. (Note that, as outlined in Section C of the Web Appendix, we consider behavior in period 14 as a fair representation of the experienced behavior of followers.) The details of the cluster analysis are provided in the Web Appendix, Section G. Furthermore, Figs. 5 and 6 in Section F of the Web Appendix show the complete-response functions in the one-shot strategy-method treatments. Most of the response functions we observe in these figures show the 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 reward and punishment, we also have evidence for *Pref-R&Ps* in these treatments.

<sup>25</sup> In a strict sense, speaking of repeated-game strategies in the context of 15 periods is something of a stretch. Still, earlier experimental results (see references above) report more cooperation in repeated settings, even with random-matching across periods, just as we observe in our own experiments. Note also that a complete formal derivation of repeated-game strategies assuming an infinite time horizon for the Stackelberg game at hand is beyond the scope of this paper. For the intricacies involved, see, e.g., Mailath and Samuelson (2006).

– is silent as to the effect of cooperation rates between teams when moving from random- to fixed-matching, there is a lot of evidence in this respect for play between individuals. For example, and closest to our context, [Huck et al. \(2001\)](#) report higher cooperation rates in individual fixed-matching Stackelberg markets than in those with random-matching. So, again, would groups still “beat” individuals in terms of higher cooperation rates and higher profits once fixed-matching is employed, given that groups already show higher cooperation rates in our random-matching treatments? If we are justified in promoting an interaction effect between the “size” of players (play among individuals versus play among teams) and the time horizon of play (one-shot versus multiple periods), also with fixed-matching across periods we should observe the behavior of groups to be further away from the subgame-perfect equilibrium of the stage game than that of individuals. Or, in other words, also with fixed-matching, groups should be more cooperative than individuals.

To find out, we also ran four 15-period, fixed-matching treatments, which we refer to as SEQ-IND-15-FM, SEQ-TEAM-15-FM, SM-IND-15-FM and SM-TEAM-15-FM, respectively. The design of these additional treatments is almost exactly the same as with their random-matching counterparts, with the sole difference being the matching protocol. [Table 1](#) provides details of the design as well as the numbers of independent observations and the numbers of new subjects participating. The lower third of [Table 2](#) shows summary statistics, [Table 3](#) provides test results, and Fig. 11 in Section H of the Web Appendix shows second-movers’ average response functions in the additional fixed-matching treatments.

Comparing the repeated fixed-matching with the repeated random-matching treatments, we find that all the summary statistics in [Table 2](#) indicate more collusive outcomes in the fixed-matching treatments when compared to the corresponding random-matching treatments. That is, in the fixed-matching treatments, all the quantities are lower and all the profits are higher than in the corresponding random-matching treatments.

More important for our purposes is the comparison of the IND and the TEAM fixed-matching treatment for each of the elicitation methods (SEQ and SM). First, we find that leaders in both IND treatments choose significantly higher quantities than leaders in the corresponding TEAM treatments. Second, the average total quantities in both IND treatments are higher than in the corresponding TEAM treatments. This translates into average total profits and average cooperation indices being higher in the TEAM than in the corresponding IND treatments. However, whereas these differences are statistically significant in the SEQ treatments, they are not in the SM treatments. Third, the inspection of followers’ average response functions shows that, by and large, team followers reward and punish more than individual followers (see the Web Appendix, Section H). Finally, we also estimated the two versions of the [Lau and Leung \(2010\)](#) model for the fixed-matching follower data. The results, again, provide evidence against [Hypothesis 1a](#) and support for [Hypothesis 1b](#) (see the Web Appendix, Section I).

In all, also in the fixed-matching treatments we find evidence for the behavior of groups being further away from the subgame-perfect equilibrium of the stage game than that of individuals (or that groups are more cooperative than individuals), which is especially true in the sequential treatments.

## 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 treatments. Our main finding is a differential effect that the time horizon of the 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 (i) the mode in which we elicit choices, (ii) the matching protocol and (iii) the model employed to account for second-mover behavior. Importantly, our main finding is in (stark) contrast to the results of earlier studies reporting that groups appear to be more selfish than individuals.

One possible explanation for the different results in our own and earlier studies is that there is heterogeneity in the 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 more selfish behavior in one-shot interactions and to more cooperative behavior in repeated interactions in comparison to individuals. Moreover, there might be differences in belief formation between individuals and groups (in our context first-movers) concerning the likely play of rivals (in our context second-movers). In particular, groups might simply be better at predicting followers’ reactions, leading them to better adjust to their rivals’ play. Finally, we demonstrate that groups appear to be better at understanding the advantages of repeated-game strategies (consisting of punishments and rewards), and also appear to be better at employing them.

In light of our results, and to the extent that the explanations of our results are convincing, it might be worthwhile to revisit other simple sequential-move games (such as the ultimatum, trust, centipede or gift-exchange games) to check for a possible differential effect of the time horizon of interaction. Whereas we concentrate on the length 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 employed (e.g., majority or unanimity voting).<sup>26</sup>

<sup>26</sup> Some 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.

The Stackelberg market game is, arguably, not of the “Eureka” type, where a solution once found is recognized as such by the players. Therefore, the results of our repeated markets are not necessarily in contrast to the findings referred to in the Sutter et al. (2009) quote in footnote 1, which summarizes results from repeated interactions in games with a strong “Eureka” component. In these games, the 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 thing to check who is closer to game-theoretic predictions in interindividual and intergroup comparisons; it is another to check who earns higher profits. Perhaps not surprisingly, there does not seem to be a simple relationship between higher conformity with game-theoretic predictions and higher profits. On the one hand, 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 the victim of the winner’s curse than individuals, and therefore groups make lower profits than individuals. In our repeated Stackelberg markets employing truly sequential play, however, we find that groups earn significantly higher total profits than individuals (see Table 3), 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 as to 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” (Wildschut and Insko, 2007). Clearly, the results of our one-shot and repeated treatments show that, indeed, there is a clear difference – or discontinuity – between interindividual and intergroup interaction. However, our results 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.<sup>27</sup>

Finally, 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 (with random- as well as with fixed-matching) in a unified framework. 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 make in the two environments. This also enables us to uncover the complete shape of the response functions used by experienced Stackelberg followers. 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 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.

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.geb.2013.09.007>.

## References

- Andreoni, J., Miller, J.H., 1993. Rational cooperation in the finitely repeated prisoner’s dilemma: Experimental evidence. *Econ. J.* 103, 570–585.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10 (1), 122–142.
- Bornstein, G., 2008. A classification of games by player type. In: Biel, A., Eek, D., Gärling, T., Gustafsson, M. (Eds.), *New Issues and Paradigms in Research on Social Dilemmas*. Springer, USA, pp. 27–42.
- Bornstein, G., Gneezy, U., 2002. Price competition between teams. *Exper. Econ.* 5, 29–38.
- Bornstein, G., Yaniv, I., 1998. Individual and group behavior in the ultimatum game: Are groups more “rational” players. *Exper. Econ.* 1, 101–108.
- Bornstein, G., Budescu, D., Kugler, T., Selten, R., 2008. Repeated price competition between individuals and between teams. *J. Econ. Behav. Organ.* 66, 808–821.
- Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group decisions in the centipede game: Are groups more “rational” players. *J. Exp. Soc. Psychol.* 40, 599–605.
- Bosman, R., Hennig-Schmidt, H., van Winden, F., 2006. Exploring group decision making in a power-to-take experiment. *Exper. Econ.* 9 (1), 35–51.
- Cardella, E., Chiu, R., 2012. Stackelberg in the lab: The effect of group decision making and deliberation periods. *J. Econ. Psych.* 33, 1070–1083.
- Cason, T., Mui, V., 1997. A laboratory study of group polarization in the team dictator game. *Econ. J.* 107, 1465–1483.
- Charness, G., Sutter, M., 2012. Groups make better self-interested decisions. *J. Econ. Perspect.* 26 (3), 157–176.
- Cooper, D., Kagel, J., 2005. Are two heads better than one? Team versus individual play in signaling games. *Amer. Econ. Rev.* 95, 477–509.
- Cox, J., 2002. Trust, reciprocity and other-regarding preferences: Group vs. individuals and males vs. females. In: Zwick, R., Rapoport, A. (Eds.), *Advances in Experimental Business Research*. Kluwer Academic Publishers, Dordrecht, pp. 331–350.
- Cox, J., Hayne, S., 2006. Barking up the right tree: Are small groups rational agents? *Exper. Econ.* 9 (3), 209–222.

<sup>27</sup> Certainly, 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.

- Cox, J., Friedman, D., Gjerstad, S., 2007. A tractable model of reciprocity and fairness. *Games Econ. Behav.* 59 (1), 17–45.
- Cox, J., Friedman, D., Sadiraj, V., 2008. Revealed altruism. *Econometrica* 76 (1), 31–69.
- Elbittar, A., Gomberg, A., Sour, L., 2004. Group decision-making in ultimatum bargaining: An experimental study. Working Papers 0407. Centro de Investigación Económica, ITAM.
- Engel, C., 2010. The behaviour of corporate actors: How much can we learn from the experimental literature? *J. Inst. Econ.* 6, 445–475.
- Fehr, E., Schmidt, K.M., 1999. A theory of fairness, competition, and cooperation. *Quart. J. Econ.* 117 (3), 817–868.
- Feri, F., Irlenbusch, B., Sutter, M., 2010. Efficiency gains from team-based coordination – Large-scale experimental evidence. *Amer. Econ. Rev.* 100, 1892–1912.
- Fischbacher, U., 2007. Z-tree – Zurich toolbox for readymade economic experiments. *Exper. Econ.* 10 (2), 171–178.
- Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Letters* 71 (3), 397–404.
- Gillet, J., Shram, A., Sonnemans, J., 2009. The tragedy of the commons revisited: The importance of group decision-making. *J. Public Econ.* 93, 785–797.
- Gillet, J., Shram, A., Sonnemans, J., 2011. Cartel formation and pricing: The effect of managerial decision making rules. *Int. J. Ind. Organ.* 29, 126–133.
- Holt, C., 1985. An experimental test of the consistent-conjectures hypothesis. *Amer. Econ. Rev.* 75 (3), 314–325.
- Huck, S., Wallace, B., 2002. Reciprocal strategies and aspiration levels in a Cournot–Stackelberg experiment. *Econ. Bull.* 3 (3), 1–7.
- Huck, S., Müller, W., Normann, H.T., 2001. Stackelberg beats Cournot: On collusion and efficiency in experimental markets. *Econ. J.* 111, 749–765.
- Huck, S., Normann, H.T., Oechssler, J., 2004. Two are few and four are many: number effects in experimental oligopoly. *J. Econ. Behav. Organ.* 53, 435–446.
- Kocher, M., Sutter, M., 2005. The decision maker matters. Individual versus team behavior in experimental beauty-contest games. *Econ. J.* 115, 200–223.
- Kocher, M., Sutter, M., 2007. Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica* 34, 63–88.
- Kreps, D.M., Milgrom, P., Roberts, J., Wilson, R., 1982. Rational cooperation in the finitely repeated prisoners' dilemma. *J. Econ. Theory* 27, 245–252.
- Kugler, T., Bornstein, G., Kocher, M., Sutter, M., 2007. Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *J. Econ. Psych.* 28 (6), 646–657.
- Kugler, T., Kausel, E., Kocher, M., 2012. Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley Interdiscip. Rev.: Cogn. Sci.* 3, 471–482.
- Lau, S.-H.P., Leung, F., 2010. Estimating a parsimonious model of inequality aversion in Stackelberg duopoly experiments. *Oxford Bull. Econ. Statist.* 72 (5), 669–686.
- Levine, J., Moreland, R.L., 1998. Small groups. In: Fisher, G., Lindzey, G. (Eds.), *The Handbook of Social Psychology*, vol. II. 4th edition. Oxford University Press, Oxford, pp. 415–469.
- Lodewijckx, H., Rabbie, J., Visser, L., 2006. “Better to be safe than to be sorry”: Extinguishing the individual – group discontinuity effect in competition by cautious reciprocation. *Eur. Rev. Soc. Psychol.* 17 (1), 185–232.
- Luhan, W., Kocher, M., Sutter, M., 2009. Group polarization in the team dictator game reconsidered. *Exper. Econ.* 12 (1), 26–41.
- Mailath, G., Samuelson, L., 2006. *Repeated Games and Reputations: Long-Term Relationships*. Oxford University Press.
- Meier, P., Hinsz, V., 2004. A comparison of human aggression committed by groups and individuals: An inter-individual–intergroup discontinuity. *J. Exp. Soc. Psychol.* 40, 551–559.
- Milgrom, P., Roberts, J., 1982. Limit pricing and entry under incomplete information: An equilibrium analysis. *Econometrica* 50, 443–459.
- Raab, P., Schipper, B., 2009. Cournot competition between teams: An experimental study. *J. Econ. Behav. Organ.* 72 (2), 691–702.
- Rabbie, J.M., 1998. Is there a discontinuity or a reciprocity effect in cooperation or competition between individuals and groups? *Eur. J. Soc. Psychol.* 28, 483–507.
- Rabe-Hesketh, S., Skrondal, A., 2005. *Multilevel and Longitudinal Modeling Using STATA*. STATA Press, College Station, Texas.
- Reuben, E., Suetens, S., 2012. Revisiting strategic versus non-strategic cooperation. *Exper. Econ.* 15, 24–43.
- Robert, C., Carnevale, P., 1997. Group choice in ultimatum bargaining. *Org. Behav. Hum. Decis. Process.* 72 (2), 256–279.
- Schopler, J., Insko, C.A., Graetz, K.A., Drigotas, S.M., Smith, V.A., 1991. The generality of the individual-group discontinuity effect. Variations in positivity–negativity of outcomes, players' relative power and magnitude of outcomes. *Pers. Soc. Psychol. Bull.* 17, 612–624.
- Selten, R., 1967. Die Strategiemethode zur Erforschung des Eingeschränkt Rationalen Verhaltens im Rahmen eines Oligopol-experiments. In: Sauermann, Heinz (Ed.), *Beiträge zur experimentellen Wirtschaftsforschung*. Mohr, Tübingen, Germany, pp. 136–168.
- Selten, R., Stoeker, R., 1986. End behavior in finite prisoner's dilemma supergames. *J. Econ. Behav. Organ.* 1, 47–70.
- Song, F., 2006. Trust and reciprocity in inter-individual versus inter-group interactions: The effects of social influence, group dynamics, and perspective biases. *Exper. Econ.* 9, 179–180.
- von Stackelberg, H., 1934. *Marktform und Gleichgewicht*. Julius Springer, Vienna.
- Stoner, J., 1961. A comparison of individual and group decisions involving risk. Unpublished master's thesis. Massachusetts Institute of Technology, Boston.
- Sutter, M., Kocher, M., Strauss, S., 2009. Individuals and teams in auctions. *Oxford Econ. Pap.* 61, 380–394.
- Teger, A., Pruitt, D., 1967. Components of group risk taking. *J. Exp. Soc. Psychol.* 3, 189–205.
- Wildschut, T., Insko, C., 2007. Explanations of interindividual – intergroup discontinuity: A review of the evidence. *Eur. Rev. Soc. Psychol.* 18 (1), 175–211.