Bonn Econ Discussion Papers

Discussion Paper 13/2008

Positive and Negative Team Identity in a Promotion Game

by

Marion Eberlein, Gari Walkowitz

June 2008

Bonn Graduate School of Economics
Department of Economics
University of Bonn
Adenauerallee 24 - 42
D-53113 Bonn

The Bonn Graduate School of Economics is
sponsored by the

Deutsche Post World Net
MAIL EXPRESS LOGISTICS FINANCE

Positive and Negative Team Identity in a Promotion Game

Marion Eberlein Gari Walkowitz*

May 28, 2008

Abstract

In this paper we experimentally investigate whether the so-called in-group/out-group bias leads to a favoring of own team members as candidates in promotion (by voting for them) relative to other teams and their members. In contrast to psychological approaches, monetary incentives for voting choices are implemented and objective performance criteria defined and thus the extent of the in-group/out-group bias is exactly measured. Our data show that face-to-face interaction with team members leads more subjects to favor own team-mates than in anonymous interaction. Moreover, not only the frequency but also the average extent of positive team identity is higher with face-to-face interaction according to objective performance measures. A further finding suggests that only anonymous team interaction often leads to substantial discrimination of own team members (i.e., negative team identity), which also is an interesting new finding and extends previous findings of psychologists on the in-group/out-group bias.

JEL-Classification: M5, L2, C9

Keywords: Team identity, promotion, experiments

*Laboratory for Experimental Economics and Department of Economics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany. Please contact: Marion.Eberlein@uni-bonn.de or Gari.Walkowitz@uni-bonn.de.
We thank Jörg Breitung, Heike Hennig-Schmidt, Alois Kneip, Matthias Kräkel, Robert J. Oxoby, Dirk Sliwka and the participants of the Brown Bag Seminar at the University of Cologne, and Erik Theissen for valuable comments and discussions. We are grateful for the help of Eva Benz, Julia Berndt, Patricia Fridrich, Hong Geng, Oliver Gürtler, Sebastién Kronisch, Mariia Labzunova, Judith Przemeck, Rainer Rilke, Anja Schöttner, Martin Schubert, Aiko Tomaschewski, and Daniel Wiesen for supporting the conduct of the experiments. Financial support from the Deutsche Forschungsgemeinschaft (DFG), grant KR 2077/2-3 and IR 43/1-1, is gratefully acknowledged.

1 Introduction

Many company representatives emphasize the importance of creating team spirit or team identity for a good team performance. Generally, team identity is regarded as a positive value especially in the labor environment. But there is also a dark side of team identity that may appear when a manager has to reach or take part in decisions on a business matter that does not only concern his own team and, in particular, could cause comparative disadvantages for his team. Because of team identity he may in this case use all his influence for a decision in favor of his team, even if this does not lead to an efficient decision as a whole.

An application to a particular case is an enterprise where the responsible heads of department are meeting in order to select one of their staffs as a candidate for a management recruiting program. The complete personal records of each staff are assumed to be known. Their performance is consequently well-known to each decision-maker. Then the question has to be raised whether the heads of department prefer to choose a staff of their own department in spite of obvious criteria in favor of staffs from other departments.

These aspects of team identity - the so called in-group/out-group bias, which leads to a favoring of the own group and their members relative to another group's members - have been investigated in a few experimental economic studies: Charness, Rigotti, and Rustichini (2007) investigate whether group membership and the saliency of the own group influences behavior in the Battle of the Sexes and the Prisoner's Dilemma Game. They find that significantly more subjects choose the strategy which maximizes the payoff of the own group when the group is most salient compared to when it is less salient. In the treatment where groups are most salient, the group of one player (who plays against someone from the out-group) sits as audience in the same room and is informed about the player's payoff, which also influences the group's payoff. Chen and Li (2007) investigate how team identity affects various dimensions of social preferences in the Dictator Game and in games with a response possibility like, e.g., reward or punishment. They show, besides other results, that subjects show more charity and less behindness-aversion to in-group than to out-group members. In addition to that, subjects reward good intentions of in-group members more than those of out-group members and forgive bad intentions of in-group members more than those of out-group members. Furthermore, Eckel and Grossman (2005), Goette, Huffman, and Meier (2006), and McLeish and Oxoby (2007) investigate whether subjects behave more cooperatively to members of the in- group than to members of the other group (in a Public Goods Game respectively in Prisoner's Dilemma Game respectively in a bargaining game). As a main result, each of these studies shows that subjects behave significantly more cooperatively to

their in-group mates than to out-group members[1]. In our experimental study we want to abstract from aspects of two-person-games but consider a promotion decision problem that involves more than one representative of each team.

There are much more experimental studies which focus on the favoring of own group members relative to out-group members in psychology. In this field, the in-group/out-group bias has extensively been investigated since the sixties. There is a considerable research on the assessment of group members' attitudes and characteristics in comparison to other groups (e.g., Sherif, Harvey, White, Hood, and Sherif 1961, Rabbie and Horwitz 1969, Tajfel, Billig, and Bundy 1971, Doise and Sinclair 1973, Tajfel and Turner 1979, Rabbie, Schot, and Visser 1989, Mummendey, Simon, Dietze, Grünert, Haeger et al. 1992, Schaller 1992, Wit and Wielke 1992, Yamagishi, Nobuhito, and Kiyonari 1999, Otten and Wentura 1999, Crisp and Hewstone 2000). As a common and consistent result it is shown that the own group is assessed more positive than the out-group. The subjects' attitude concerning the evaluation of the own group is substantially more affirmative. There are much fewer investigations in the field of psychology with a view to the evaluation of performance (e.g., Sherif et al. 1961, Ferguson and Kelley 1964, Brewer 1979, Kraiger and Ford 1985, Downing and Monaco 1986). As a general result it can also be stated that the performance of the own group is assessed more favorably than the one of other groups.

To the best of our knowledge, none of the studies so far carried out on team identity focuses on the aspect of competition within teams, next to competition between teams. It is not clear if team identity endures when also the own team-mates compete for attaining a specific goal. Will people then favor their team-mates in relation to other groups' team-mates or does competition neutralize team identity or even lead to a discrimination of own team-mates? Imagine a subject who does not win a tournament for promotion or just thinks that his chances are very small. For this subject it may be even worse when one of his team-mates wins the tournament and not a third person from another team, who is perhaps not even personally known to this person[2]. The closer relationship to the team-mate on the one hand and the distance and more abstract perception of the out-group member on the other

[1]There are many laboratory economic experiments focusing on group identity in real groups, which also exist outside the laboratory, as for example nationalities. For example, in the investment game, Buchan, Johnson, and Croson (2006) find that American students exhibit an in-group bias: They are more willing to trust other American students than students of other nationalities. Ruffle and Sosis (2006) observe that kibbutz members cooperate more with members of their own kibbutz than with city residents. Solow and Kirkwood (2002) compare contributions in a Public Goods Game between members of a real group, members of an experimentally created group, and strangers. They only find significantly higher contributions when comparing contributions of members of a real group with contributions of strangers.

[2]A similar aspect is discussed by Grund and Sliwka (2005) in a theoretical tournament model with inequity averse competitors. The authors argue that lateral promotions reduce inequity costs compared to vertical promotions: In vertical promotions the winner of the promotion tournament becomes the superior of

hand, exposes this subject to a higher action level of envy, revenge, or even discrimination against the own team-mate (see, e.g., Suls and Wheeler (2000) for an overview of Social Comparison Theory).

Our design enables us to investigate both positive and negative team identity[3] by incorporating competition - also between team-mates - to a promotion decision problem. From an efficiency point of view, the subjects with the best performances shall be promoted to an elite-team in our experiment. Each subject has to assign promotion points to each of the other subjects. These promotion decisions may be distorted for two different reasons. First, the performance evaluation of certain subjects may be influenced by their membership to a certain group, e.g., a subject may favor his own team-mates for promotion although they are objectively not good enough to join the elite-team. Second, a subject who wants to join the elite-team himself may try to discriminate his team-mates so that his chance to get into the elite-team becomes higher than theirs[4]. Within the experimental design of our work, we can isolate and analyze such distortion effects on performance evaluation of in-group and out-group members.

Compared to the psychologists' approach, we apply an experimental procedure that has the following advantages: First, in contrast to psychological analyses, objective performance criteria are defined and thus the extent of the in-group/out-group bias exactly measured. In the psychological studies, the authors only quantify to what degree the own group is evaluated higher than others. But there is no objective evaluation criterion that enables the authors to measure the exact degree of team identity. For example Downing and Monaco (1986) investigate whether subjects evaluate the own team members better than the out-group when they have to show ski exercises. Then the authors compare whether the own group or the other groups receive better evaluations, but an objective criterion which group is actually better is missing. However, in our experiment, subjects solve multiplication tasks and we can exactly rank them according to their performance. Hence, objectively, it is clear who should be elected into the elite-team in our experiment. Second, in contrast to psychological

his former colleague and both are faced with the result of the tournament permanently. Contrarily, in lateral promotions neither the winner nor the loser face their former colleague anymore. Thus, feelings of envy (and compassion) are higher in vertical than in lateral promotions.

[3]With "positive team identity" we refer to the favoring of own team-mates, with "negative team identity" to the discrimination of own team-mates. A more detailed definition is given in section 3.2.

[4]This aspect is related to sabotage in tournaments. In the respective literature competitors cannot only try to win the tournament by exerting much productive effort and therefore reaching high output, but can also try to reduce the competitors' output by sabotaging them. Because only the relative order of the amount of outputs between the rivals determines who wins the tournament, sabotage can be optimal for a competitor. For theoretical literature on sabotage in tournaments see, e.g., Lazear (1989), Kräkel (2005), and Gürtler (2008). For experimental studies in this field, see, for example, Harbring, Irlenbusch, Kräkel, and Selten (2007), and Harbring and Irlenbusch (2008).

studies we provide monetary incentives which counteract in-group favoring. Each subject is paid an extra amount of money which depends on the performance of the elite-team. Hence, an incentive is given to elect subjects with the best performances into the elite-team. If, in spite of this environment, a team favoring is observed, it provides a stronger evidence for team identity than without monetary incentives.

Our data show that face-to-face interaction between team members leads more subjects to favor own team-mates than in anonymous interaction. Moreover, not only the frequency of positive team identity is higher but also the average extent. There are not only more subjects who favor own team-mates with face-to face interaction, but own team-mates are favored to a higher degree according to objective performance measures. This aspect is not investigated by psychological studies on team identity. Another striking result suggests that anonymous group interaction often leads to discrimination of own team members, i.e., negative team identity. This is also an interesting new finding that enriches previous findings of psychologists and enhances the understanding of team interaction.

The remainder of this paper is organized as follows: In the next section we present our experimental design, treatments, and procedure. Afterwards, we present our results. The final section discusses our findings and concludes.

2 Experimental Design, Treatments, and Procedure

2.1 Basic Experimental Design

Our basic experimental design consists of three subsequent stages: i) First real effort performance stage, ii) Voting-for-promotion stage, iii) Second real effort performance stage. At the beginning of the first stage, each subject draws an individual code which is kept secretly and guarantees full anonymity. In the following, we describe the features of each of these stages in detail.

(i) First real effort performance stage (stage 1)
At the first stage, each of 12 invited subjects gets a list with 100 arithmetical calculations to be carried out within a time limit of 15 minutes. The set is made up of 100 simple products of numbers which can be multiplied easily. For each correct solution subjects earn 3 Cents. Auxiliary means except pen and paper are not permitted.

(ii) Voting-for-promotion stage (stage 2)
After the multiplication work, subjects are instructed about the remaining two stages of

the experiment. In particular, they are informed that, at stage 3, equally structured and comparable arithmetical tasks have to be solved, for which again 15 minutes of time are provided and 3 Cents for each correct solution are paid. In contrast to the first stage, there are two different roles for the players at stage 3: role A and role B with respectively 5 and 7 players. The 5 individuals playing in role A receive a bonus of 10 Euro each. Contrarily, those 7 subjects playing in role B do not get such a bonus. Furthermore, each player (independent of role A or B) earns an extra payment, in addition to his individual achievement, that depends on the performance of the players in role A. It amounts to the average number of multiplications correctly calculated by the players in role A at stage 3 multiplied by 3 Cents. By this mechanism an incentive is given to the players to have the best 5 of them in role A. On the other hand, each subject has an incentive to be in role A himself. The subjects do not learn which role they played until the end of the experiment[5].

Before the calculations of the second set of multiplication tasks are actually carried out, the players for role A are elected by all the 12 participants of the experimental session. After the participants have been instructed about this fact and also about stage 3, a piece of paper is handed out to everyone containing a list of the code names of the other 11 subjects together with the number of products they have solved correctly at stage 1. Subjects are not informed about their own performance, which could otherwise influence their voting. In order to decide upon the role of each player, every individual has to propose which subjects shall act in role A. No player can nominate himself. According to Borda's rule, 11 points shall be attached to the code name of that subject who is preferred most for acting in role A, 10 points to the person who is preferred second most, etc. To the code name of that person who is preferred least for role A, 1 point is to be assigned. Subsequently, the subjects are ranked according to the total number of points they received from all their fellows. Those 5 persons with the most points are appointed to role A[6]. The others have to act in role B[7].

Because the subjects get an extra payment which depends upon the average achievement of the members of the elite-team (players in role A), there is an incentive to choose the most qualified persons for the elite-team. But also another, a "strategic" behavior is possible: In order to increase their own chances to get into the elite-team, subjects might try to weaken

[5]This is to prevent a possible influence on a player's effort when he knows that he has already earned the 10 Euro reward as player in role A.

[6]See, e.g., YOUNG, (1974) for an axiomatization of Borda's rule. As Borda's rule violates the principle of "independence of irrelevant alternatives" it is possible to manipulate its outcome by introducing extraneous alternatives (see, e.g., YOUNG 1995). In our experiment, no additional alternatives (i.e., other subjects) are available - the number of teams and team members is constant.

[7]In case of equality of points, which does not allow to clearly define the 5 persons with the most points, a random decision is made. But this case never occurred.

the position of their strongest competitors. In the most extreme case they might assign only 1 point to the player with the best achievement and 11 points to the player with the worst achievement in stage 1 (in the following we will refer to this strategy as the COMPETITIVE one)[8].

As subjects have been informed that the multiplication tasks they have to solve at the next stage are equally structured to those they have already solved, they know that someone who performed very well at stage 1 is likely to perform very well at stage 3, too. Subjects therefore know who should - objectively - join the elite-team. This also means that they can estimate the resulting costs (i.e., the lower extra payment) if they instead favor a team-mate for the elite-team who does not belong to the 5 best ones.

(iii) Second real effort performance stage (stage 3)
At this stage, subjects have to carry out multiplication tasks within a 15 minutes' time limit. The calculation tasks are the same as at stage 1, but in another randomized order. This allows us to exactly compare the performances at stage 1 and 3. For each correct answer 3 Cents are paid.

2.2 Experimental Treatments and Procedure

We apply three different treatment conditions to investigate which environment creates team identity and by which means this is affected, especially with having in view the favoring or discrimination of own team-mates as distinctive essential outcomes. Each of our treatments is equally structured in the three stages described above, but different in team identity enhancing factors.

Anonymous Treatment (ANT)
In a session of the first treatment, the Anonymous Treatment (ANT), 12 subjects are randomly assigned to 4 groups (i.e., teams), each made up of 3 team members. However, neither team members nor members of any other team are able to see each other after team division or to have the opportunity to communicate at any time during the experiment.

The scores of the multiplication tasks of the first stage are summed up for all the members of a team so that a common team score can be determined. These team scores are compared between the 4 teams. Each subject of the team with the highest team score receives a team bonus of 5 Euro. By this a bit of (presumably weak) team identity may be generated because

[8] Although it is no equilibrium that every subject uses the competitive strategy, subjects might use this strategy because they think that it is profitable.

of common economic interest[9]. Which team finally gets the team bonus is announced not before the end of the experiment. In that way a difference in team identity, which may be caused by a different angle of winners and losers, shall be avoided.

But already at this early phase of the experiment, the subjects are informed about how their team results are assessed. In detail, each subject is informed about how many right calculations his team fellows achieved together, but is not told how many of the tasks he himself has solved correctly.

After every subject has received such an intermediate status of specific team performance, they are instructed about the remaining stages of the experiment and get the voting list thereafter. By this list, subjects learn all other subjects' code names and scores and whether they belong to their own team or not. Subjects not adjoined to the own team cannot be connected with a certain other team. Therefore, none of the other team scores is known[10]. Taken together, by the team score competition and through the handling of team membership - using the denotion "group" in the instructions and adjoining subjects to a certain team per se - subjects are given the possibility to develop some degree of team identity. Special emphasis lies hereby on economic team identity, which can arise from the common will to win the team bonus.

Group Treatment (GRT)

By the GRT, we expand the spectrum of possible team identity in the way that we also introduce social factors in addition to the same factors already applied in ANT. For this, we organize a comprehensive communication phase prior to stage 1, where members of each team are given the opportunity to meet with their team fellows and to talk face-to-face in a separate room. They are admitted and advised by an assigned experiment assistant, who is responsible for this specific room during the whole experiment and can give supplementary support. In order to support and enhance social identity the team members in each room are asked to carry out some tasks together. By performing a simple task together and directing effort to a common objective, we want to ensure that the subjects get somewhat acquainted.

First, the 3 members of a team have to determine a name for their team. This name has to consist of 3 words, each beginning with the initials of one of the team members' forenames.

[9]Psychological studies show that a "minimal group condition" can already result in favoring members of the own group (see Bewer (1979) for an overview). Psychologists usually employ as a minimum group condition a reasonable division into groups, for example Doise and Sinclair (1973) tell subjects that they are divided into two groups according to their preference to one of two photographs.

[10]This is to prevent a possible influence on voting, e.g., by disadvantaging the team with the best team score.

After having agreed on a team name, they have to write it on a cardboard and affix it outside to the door. While the team is thinking about an appropriate name, the experiment assistant is leaving the room in order to avoid disturbance of the social team identity finding process. Thereafter, subjects have to read two short newspaper articles about the lawn of the soccer world cup 2006 in Germany and about the first bio-energy village in Germany. For these tasks they get 5 minutes' time. Only two copies of the articles are provided for each 3-person-team to enhance cooperation and coordination among the team members. Then, the team gets 3 minutes to deal with a fragmentary text with statements concerning the text read before. For each right answer completing the lacking parts of the article, 50 Cents are paid and directly donated to the Beethoven Memorial House in Bonn[11]. By donation instead of paying off, we want to foster both subjects' effort and high-minded approach[12].

The described communication phase is used as an instrument for the process of establishing a social identity. It is important to create a relaxed atmosphere prevailing during the whole phase. To support this intention, team rooms allow sitting around a shared table. After completing the communication phase, the economic phase begins, which is identical to ANT. During this stage, the subjects must not talk and have to use their anonymous code names, which the other team members do not know and cannot link to the other two persons of the team[13]. Each team member is now seated at a separate table in the team room.

Writing the instructions for both ANT and GRT we use the term "group" only to describe the formal division into different teams. This is done very carefully to avoid any framing bias or demand effects.

Individual Control Treatment (ICT)

In ICT, we exclude all factors which may cause economic or social team identity. So the ICT

[11] We choose a social objective in form of the donation to the Beethoven Memorial House because it stresses the regional relation and thus may strengthen identity.

[12] We could, of course, pay off the earnings immediately to the subjects but that would correspond more to an economic identity, which we want to attain at a later point of time by paying the team bonus. Moreover, we want to avoid a shifting of marginal incentives for elaborating the multiplication tasks, which may happen if already the communication phase is profitable.

[13] In our experiment we have to distribute papers with certain information to certain subjects when informing them about the scores of their two team-mates and about the contents of their voting list (remember that a subject's score does not appear on his list). At distributing we have to be careful not to hurt anonymity. Therefore, we use the following procedure: The assistant in the room picks up a subject's secret code name, which is written on a piece of paper and placed folded on the subject's desks. The code name cannot be read by the assistant or the other two persons in the room. Afterwards, the assistant takes it to the experimenter waiting outside the room, who then finds out the associated paper, which is then given to the subject by the assistant together with the code picked up before. Again this is done in the same secret way as at the start of the procedure so that the other two fellows and the assistant herself cannot read the information. Also the experimenter is not completely informed because, through this procedure, he cannot see any of the subjects in the room so that he is incapable of associating code names with persons. By this it is ensured that each subject gets the necessary particular information without that anyone gives up anonymity.

includes only the basic design and subjects are not separated into different groups. We use this treatment as a reference for the experimental treatments ANT and GRT. By a comparison, we are particularly able to analyze whether the underlying voting mechanism leading either to an objective or a competitive strategy is influenced by our treatment conditions.

Table 1 shows an overview of our treatments and its features.

	Economic identity Team features				Social identity Communication phase			
Treatment	**Creating "teams"**	**Team bonus**	**Information about team performance**	**Information about team composition**	**See/ talk**	**Joint task**	**Joint feedback**	**Group name**
ICT	−	−	−	−	−	−	−	−
ANT	+	+	+	+	−	−	−	−
GRT	+	+	+	+	+	+	+	+

Table 1: Treatment specifics.

The pen and paper experiment was conducted at the Rheinische Friedrich-Wilhelms-University Bonn. All individuals were students from this university. Subjects were recruited via the internet by using ORSEE software (Greiner, 2003) announcing the possibility to earn an amount of money dependent on their behavior. The treatments lasted about 90 - 120 minutes. On average, subjects earned 15.50 Euro. The Individual Control Treatment (ICT) consisted of two sessions with 12 subjects per session. The other two treatments (ANT and GRT) were made up of three sessions with a total of 36 participants per treatment.

3 Results

In this section we present our results. First, we analyze the strategies that underlie subjects' voting behavior. Next, we explore whether team identity has a substantial impact on subjects' votes, using the underlying strategies as a reference for comparison. We develop quantitative and qualitative measures to investigate team identity. Finally, we enter into a discussion about the efficiency of the composition of the elite-team.

3.1 Reference Strategies

When investigating whether subjects favor or discriminate their team-mates, a reference strategy is needed to which such deviations can be measured. We distinguish two strategies which can serve as a reference: the objective and the competitive one. The individual degree of the observed in-group/out-group bias can differ depending on what reference strategy is pursued. A strict objective strategy is defined as a ranking where the subject with the best performance gets the most (i.e., 11) points, the one with the second best performance the

second most (i.e., 10) points, etc. Hence, the participant with the lowest performance receives the least (i.e., 1) points. Contrarily, a strict competitive strategy is defined conversely: The subject with the highest performance receives only 1 point and the participant with the worst performance gets the most (i.e., 11) points, etc.

When a subject basically wants to send high-performance players to the elite-team and therefore gives most points to them but modifies this strategy a bit in favor of team fellows, team identity should, as a matter of course, be measured in relation to the objective ranking. On the other hand, when a subject decides in favor of a competitive strategy, he can show team identity by changing the team fellows' rank relatively to this strategy. If all the subjects' rankings were, e.g., compared to the objective rankings and team identity was determined on that basis, an underlying strategy would be attributed which does not correctly describe the behavior of certain subjects and thus would yield a wrong conclusion.

To use the appropriate reference strategy, we have to define an exact measure by which we can decide whether a subject basically uses the objective or the competitive strategy. Thus, for each subject, we calculate a measure m_1 to determine the individual reference strategy. In the following we define m_1 formally[14].

With a subject's score we denote a subject's number of correctly solved products at stage 1. With score s_i at position i $(i = 1, ..., 11)$ of a subject's ranking[15] the number K of all elements s_i greater or equal than the preceding element s_{i-1} can be calculated by

$$k_i = \begin{cases} 1 & \text{for} \quad (s_i - s_{i-1}) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \text{ with } i = 2, ..., 11, \sum_{i=2}^{11} k_i = K$$

Conversely, for the L elements in decreasing order

$$l_i = \begin{cases} 1 & \text{for} \quad (s_i - s_{i-1}) < 0 \\ 0 & \text{otherwise} \end{cases}, \text{ with } i = 2, ..., 11, \sum_{i=2}^{11} l_i = L$$

It follows $k_i + l_i = 1$ for every i. Thus, it holds $K + L = 10$. The measure m_1 is defined as $m_1 = K/10$. There is a mainly increasing order if $K > L$. Consequently this yields

[14] Please see Appendix for a calculation example.

[15] With a subject's ranking we refer to his distribution of points toward the other subjects in his voting list. The subject who is preferred most in a subject's ranking is at position $i = 1$ (i.e., gets 11 points).

$K > 10 - K$ and $K/10 > 0.5$. Besides, from $K \leq 10$ follows $K/10 \leq 1$. Hence, a measure is yielded for mainly increasing order of scores (which corresponds to predominantly competitive ranking) $0.5 < K/10 \leq 1$ or $0.5 < m_1 \leq 1$. Conversely, it can easily be shown for $K < L$ that $0 \leq K/10 < 0.5$ respectively $0 \leq m_1 < 0.5$, i.e., for mainly decreasing order of scores respectively predominantly objective ranking. For $K = L$, i.e., $K/10 = 0.5$, the order has no predominant direction. In this case the order is considered as objective[16]. Hence, for predominantly objective ranking it is $0 \leq m_1 \leq 0.5$.

To sum up: When a subject's ranking is predominantly arranged in an increasing order of performance, the value of m_1 takes on values between 0.5 and 1 ($0.5 < m_1 \leq 1$) and the competitive strategy is assigned. In the other case, where $0 \leq m_1 \leq 0.5$, the objective strategy is assigned.

It may be assumed that the choice of an underlying strategy depends on the given set of team identity factors. As the ICT is the treatment where no team identity factors are given, this may yield to more competitive voting behaviors. Contrarily, the objective strategy may occur more often in ANT and GRT. An overview of the applied reference strategies per treatment can be found in Table 2.

As can be seen from Table 2, the objective strategy is applied more often than the competitive one in all three treatments. Moreover, the objective strategy is applied most in GRT and least in ANT. But when testing whether the frequencies of the chosen strategies (objective or competitive) differ between our treatments, no significant difference can be found[17].

Treatment	**ICT**	**ANT**	**GRT**
Objective strategy	79.17	69.44	83.33
Competitive strategy	20.83	30.56	16.67

Table 2: Relative frequency (in %) of strategies dependent on treatment.

This means that the choice of strategy does in most cases not depend on the given team identity factors. Therefore, the application of either strategy does not depend on whether no team identity is provoked, economic team identity is introduced, or a communication phase is established in addition to economic identity to foster team identity.

Because the objective strategy is used most often, most subjects seem to have the election of a high-performance elite-team in mind when voting. But as they also would benefit from

[16] A tie occurred only twice in our whole data set.

[17] Results from two-sided Chi square tests: Pearson $\text{chi}^2 = 0.6960, p = .404$ when comparing ICT and ANT, Pearson $\text{chi}^2 = 1.9251, p = .165$ when comparing GRT and ANT, Pearson $\text{chi}^2 = 0.1670, p = .683$ when comparing ICT and GRT.

being elected to the elite-team, their voting may be based on another kind of reasoning: If subjects want to enhance their own chance to get into the elite-team on an indirect way, their behavior depends on their belief of the other subjects' strategy choices and on subjects' assessment of how many tasks they themselves have solved correctly at stage 1[18]. If a subject believes that he has performed rather bad and believes that many other subjects use the competitive strategy, he may vote objectively to increase his chance to join the elite-team.

Continuing this train of thoughts it can analogously be argued that if a subject believes to have performed really well and believes that many other subjects vote objectively, he may further enhance his own chance of getting into the elite-team by voting competitively. With our data set, we can test by the calculation of a point bi-serial correlation coefficient for ICT, ANT, and GRT whether there is a correlation between the assessed performance and the choice of strategy, but no sufficient correlation can be found[19].

3.2 Quantitative Measurement of Team Identity

Method

Based on the subjects' individual lists of preferences as stated in the ranking, we can investigate whether their voting is biased by team identity, given a certain reference strategy. As a prerequisite, team identity has to be measured in a suitable way. When evaluating voting behavior, each subject has to be considered individually because each subject gets a different voting list, as subjects do not appear on their own list. We define positive (negative) team identity as follows: When a subject assigns his own team-mate to a better (worse) rank - given his reference strategy - positive (negative) team identity is shown.

At the beginning, we investigate quantitatively whether subjects show team identity or not and neglect the extent of team identity for the moment. For each subject we verify typical characteristics in view of team identity in his ranking. For a short characterization, we use the 3-tuple (positive team identity, negative team identity, no team identity) with a dichotomous variable (1 denotes "yes" and 0 "no") for each characteristic. For example, if a subject does not dislocate any team member, his personal feature is (0,0,1)[20]. By these features, we

[18] As already mentioned, no subject knows his own performance. But subjects are asked in the questionnaire how many tasks they think to have correctly solved at the first stage.

[19] Subjects' statements about their self-assessment have to be treated with some caution because they are not paid for a correct self-assessment. In spite of that, their self-assessments are relatively accurate. On average, they underestimate themselves a bit: Average deviations from actual scores are: -1.58 ($SD = 7.27$) in ICT, -3.97 ($SD = 11.56$) in ANT, and -4.44 ($SD = 7.81$) in GRT.

[20] If a subject shows positive team identity to both team-mates his feature is (1,0,0) and if a subject shows negative team identity to both team-mates his feature is (0,1,0). There are subjects who do not treat both team-mates equally. If a subject ranks one team-mate higher than in his reference ranking and the other lower,

investigate the quantity of team identity bias occurring in the treatments and whether the detected frequencies differ across ANT and GRT[21].

Working Hypothesis

Because of the findings from literature presented at the beginning of our paper and since more team identity enhancing factors are given in GRT than in ANT, we state the following hypothesis concerning a comparison between the frequencies of positive and negative team identity in GRT and ANT:

Hypothesis: *(i) In GRT more subjects show positive team identity than in ANT.*
(ii) In GRT less subjects show negative team identity than in ANT.

Frequencies of positive and negative team identity

Table 3 displays the percentages of subjects who show positive, negative, or no team identity in ANT and in GRT[22].

	Positive team identity		**Negative team identity**		**No team identity**	
Treatment	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
ANT	33.33	66.66	30.56	69.44	66.66	33.33
GRT	52.78	47.22	11.11	88.89	58.33	41.67

Table 3: Relative frequencies (in %) of positive, negative, and no team identity.

Interestingly, in both treatments positive and negative team identity can be detected. Considering ANT, 33.33% of all subjects participating in ANT show positive team identity to at least one team-mate and 30.56% show negative team identity to at least one team-mate. In GRT, 52.78% show positive team identity to at least one team-mate and only 11.11% show negative team identity to at least one team-mate.

To examine the first part (i) of our working hypothesis, we test "positive team identity" and "no positive team identity" across ANT and GRT. As a result, we find that the frequencies of these two characteristics differ significantly between the treatments (Chi square test, Pearson $\text{chi}^2 = 2.7758, p = .048$, one-sided), with positive team identity occurring more often in GRT. Thus, exploring whether subjects show positive team identity or not and neglecting the extent of this identity indicates that significantly more subjects in GRT show positive team identity than in ANT. Consequently, our working hypothesis (i) cannot be rejected. Our data show

the feature is (1,1,0). If one team-mate is not dislocated but the other is dislocated, his feature is (1,0,1) when favoring this team-mate, or (0,1,1) when discriminating this team-mate.

[21] This analysis is not undertaken for ICT since no teams are formed.

[22] In Table 3, the percentages of subjects who are written under "Yes" and "No" for a certain category for a certain treatment add to 100%. Because each subject can treat his team-mates differently, percentages across categories for a certain treatment do not necessarily add to 100%.

that giving subjects the possibility to talk to each other face-to-face, so that they get to know each other, and to solve some tasks together, with a high-minded aim in view, results in more positive deviations from the reference ranking concerning the own team-mates than without these additional social group identity factors.

To examine the second part of our working hypothesis, we test the two categories "negative team identity" and "no negative team identity" between ANT and GRT. We find that the frequencies of these characteristics also differ significantly between the treatments (Fisher exact test, $p = .040$, one-sided[23]), with negative team identity occurring more often in ANT. Thus, omitting the communication phase leads more subjects to discriminate their own team-mates by shifting them down in their ranking. Getting acquainted with one's team-mates and fulfilling a task together seem to work against this discrimination and neutralize negative team identity or even change it into more positive attitudes toward team members. Hence, our working hypothesis (ii) can also not be rejected.

This is a striking result: In GRT, goodwill toward own team-mates is not only expressed by favoring them more often than in ANT (i.e., placing them higher in the individual ranking) but also by not putting them into a worse position compared to the reference ranking. Thus, in GRT team identity appears twice - in positive and in not negative actions undertaken by own team-mates.

In the next step we test if the number of subjects who show "no team identity" to at least one team-mate and the number who shows "any kind of team identity" differ between ANT and GRT. We find that the quantities do not differ significantly across treatments (Chi square test, Pearson chi$^2 = 3.1280, p = .209$, two-sided).

The so far accomplished statistical tests are not totally independent of each other[24]. To overcome this methodological issue we additionally investigate combinations of the three investigated characteristics. In view of our research topic, the *complete* characterization tuples (1,0,0) and (0,1,0) - subjects who show *only positive* or *only negative* identity to both team-mates - are of particular interest. So we test whether the frequencies of the complete

[23] A Fisher exact test is applied here because the required minimum number of observations per category for the conduct of a Chi square test (which is 5) is not satisfied.

[24] The specifications 0 or 1 in the categories "positive team identity", "negative team identity", and "no team identity" are not totally independent of each other: If a subject neither shows positive nor negative team identity to any team- mate, it necessarily follows that he shows no team identity to any team-mate. This is why also the tests of the categories are not totally independent of each other: As we have shown, the test concerning "no team identity" yields no significant difference between ANT and GRT and the test concerning "positive team identity" shows that there is more positive team identity in GRT than in ANT. Then it is logical that there are more subjects in ANT showing negative team identity than in GRT. But, in any case, the results of the specific tests are meaningful, because within these tests the data are independent.

characterization tuples [(1,0,0) or not (1,0,0)] and [(0,1,0) or not (0,1,0)] substantially differ across ANT and GRT. In addition to that, we test the frequencies of the complete tuples [(0,0,1) or not (0,0,1)] in ANT and GRT. Table 4 gives an overview of the distribution of these specific cases.

	Only positive team identity (1,0,0)		**Only negative team identity (0,1,0)**		**Only no team identity (0,0,1)**	
Treatment	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
ANT	8.33	91.67	13.89	86.11	47.22	52.78
GRT	33.33	66.66	0	100	44.44	55.56

Table 4: Relative frequencies (in %) of only positive, only negative, or no team identity.

As can be seen from Table 4, 8.33% of all subjects participating in ANT show "only positive team identity", i.e., favor both team-mates[25]. A higher percentage of subjects (13.89%) shows "only negative team identity", i.e., discriminate both team mates. This is quite different when considering GRT: 33.33% of all subjects participating in GRT favor both team-mates while there is no subject in this treatment who discriminates both team-mates.

When testing if the amount of subjects who show "only positive identity" or "not only positive identity" differ between ANT and GRT, we find a significant difference (Fisher exact test, $p = .009$, one-sided). There are significantly more subjects in GRT than in ANT who show positive team identity to both team-mates. When testing if the number of subjects who show "only negative identity" and of subjects who do not show "only negative identity" differ between ANT and GRT, there is also a significant difference found (Fisher exact test, $p = .027$, one-sided). Here, significantly more subjects discriminate both team members in ANT than do subjects in GRT. Considering "only no team identity", we find no significant difference among the two treatments (Chi square test, Pearson chi$^2 = 0.0559, p = .813$, two-sided)[26].

Finally, we investigate if there is a correlation between a subject's underlying strategy and the occurrence of positive or negative or no team identity. It seems plausible that someone who uses the competitive strategy is mainly focused on joining the elite-team. Therefore, he may regard his team-mates as (closer) competitors and is not willing to give them a higher chance to join the elite-team by favoring them because of envy. Perhaps he even wants to discriminate them in order to reduce their chances to win the elite bonus - independent of their performance at stage 1.

[25]Also in Table 4, the percentages of subjects who are stated under "Yes" or "No" of a certain category for a certain treatment add to 100%. Again, across these categories, the percentages do not necessarily add up to 100% for a certain treatment.

[26]These three tests of tuples are totally independent of each other because a subject who is not characterized by tuple (1,0,0) is not necessarily characterized by tuple (0,1,0) or by (0,0,1).

Concerning ANT we find a negative correlation between using the competitive strategy and showing positive team identity ($\phi = -0.3$, phi-correlation coefficient for two nominal and dichotomous variables). This means that a subject who uses the competitive strategy, less frequently shows positive identity than when using the objective strategy. There is no correlation between the chosen strategy and negative team identity in ANT. However, we find a positive correlation between using the competitive strategy and showing no team identity ($\phi = 0.4$). This is in line with competitive behavior that includes neither positive nor negative team identity. In GRT, there is a positive correlation between the competitive strategy and negative team identity ($\phi = 0.3$). When a subject uses the competitive strategy, he more frequently shows negative team identity than when using the objective strategy. Again this finding can be interpreted in such a way that a subject who is competitive preferably does not want his team-mates to join the elite-team and therefore discriminates them. For the other cases no correlation can be found.

The competitive strategy seems to counteract favoring of own team-mates in both treatments, but through different ways: While it leads to less occurrences of positive team identity in ANT it provokes more occurrences of negative team identity in GRT.

In order to get a better insight into the driving forces behind team identity we conduct an order probit regression analysis. Table 5 displays the estimated coefficients and their effect on a subject's probability to show positive, negative, or no team identity to a certain team-mate.

In our regression we model "team identity" toward a certain team-mate as the dependent variable which can either be negative (-1) or positive (1), or zero (0) for the case that neither type of team identity occurs. According to our research agenda, we investigate the impact of the independent variable TREATMENT (ANT with anonymous interaction of team-mates and GRT with face-to-face interaction). To find out whether a subject's team identity is influenced by the team-mates' absolute and relative performances we include the independent variables ASSOCIATED SUBJECT'S SCORE (a team-mate's score at stage 1) and ASSOCIATED SUBJECT'S RANK (a team-mate's rank in the reference ranking) in our regression.

Independent Variable	Coefficient (Standard error)
TREATMENT	.63558***
	(.21022)
ASSOCIATED SUBJECT'S SCORE	.01466*
	(.00790)
ASSOCIATED SUBJECT'S RANK	.06407*
	(.03828)
REFERENCE STRATEGY	.27737
	(.24168)
TEAM BONUS EXPECTED	.20902
	(.21259)
OWN ACTUAL SCORE	.00970
	(.01111)
SELF-ASSESSMENT OF OWN ACTUAL SCORE	−.00678
	(.01103)
SEX	.17260
	(.24418)
Cut1	1.13856
	(0.89671)
Cut2	2.83161
	(0.91487)
N	140
(Pseudo) R^2	0.0761
LR-χ^2	20.69
Prob $> chi^2$	.0080

Table 5: Ordered probit estimates for dependent variable TEAM IDENTITY. 1 [-1] represents positive [negative] team identity and 0 stands for no team identity. We apply 0 [1] for ANT [GRT], objective [competitive] reference strategy, no expectation of group bonus [expectation of group bonus], female [male] as dummy variables. Asterisks indicate variables as being significant at $1\%^{***}, 5\%^{**}$**, and** $10\%^{*}$**.**

A high absolute and relative performance of a team-mate are expected to positively influence a favoring of him. Moreover we include the variable REFERENCE STRATEGY with which we now examine the general effect of the chosen reference strategy on team identity. Furthermore, our regression contains the variables TEAM BONUS EXPECTED, OWN ACTUAL SCORE, SELF-ASSESSMENT OF OWN ACTUAL SCORE, and SEX[27]. Whether a subject expects his team to win the team bonus or not may also guide his behavior toward team-mates, moderated by his own actual performance at stage 1 (which is unknown to the subject) and, more important, the assessment of his actual performance at stage 1. The expectation of the team bonus - dependent on team members' scores -, together with a positive self-assessment, might lead to a more positive, generous, and favoring attitude toward team-mates. When the team bonus is not expected this may lead to negative feelings (e.g., anger, disappointment, frustration) inducing implicit negative reciprocity toward team-mates. A voter might blame his team members for not winning the team bonus resulting in punishment by discriminatory

[27]Of course we observe a subject's actual score in the multiplication task. Furthermore, we ask subjects whether they expect to have won the team bonus, how they assess their performance and whether they are male or female in our questionnaire at the end of our experiment.

voting behavior. However, it is also plausible to expect poor players to form a coalition of "losers".

Table 5 shows that TREATMENT highly significantly influences subjects' voting attitudes toward team-mates. Thus, in GRT, positive team identity is significantly more likely to occur, and at the same time, negative or no team identity are less likely to be exhibited. As expected, a higher ASSOCIATED SUBJECT'S SCORE and a better ASSOCIATED SUBJECT'S RANK do also have a positive influence on voting behavior that favors associated team-mates. The underlying REFERENCE STRATEGY - in general - does not seem to sufficiently guide subjects in their voting behavior. Also TEAM BONUS EXPECTED, OWN ACTUAL SCORE, SELF-ASSESSMENT OF OWN ACTUAL SCORE, and SEX do not significantly influence the probability of favoring or discriminating a team-mate or to show no team identity toward him.

3.3 The Extent of Team Identity

To test the extent of team identity, we develop two measures, m_2 and m_3. As already explained, we determine measure m_1 for each subject first in order to find out whether we have to use the objective or the competitive strategy as reference for team identity measurement.

Measure m_2

After the determination of the reference strategy, measure m_2 is applied. It is used for investigating *where* members of the own team are placed in the ranking. The more positions they are shifted, the higher is the degree of team identity. Measure m_2 is constructed by the actual number of shifted positions - in relation to the reference sequence - and the maximum number of ranks which are possible to leap over (in the same direction)[28]. When a team fellow is shifted from position i to position J, i.e., from rank r_i to rank r_J, *relative to the reference ranking*, the degree of shifting is measured by the difference in ranks normalized by division through the maximum shift possible in the same direction. Formally this yields for either team fellow:

$$\hat{m}_2 = \text{sign} \cdot (r_i - r_J)/(r_i - R), \text{ with } r_i = i, i = 1, ..., 11^{29} \text{ and}$$

$$\text{sign} = \begin{cases} +1 & \text{for upward-shifting} \\ -1 & \text{for downward-shifting} \end{cases}$$

[28] Please see Appendix for a calculation example.

[29] Again, $i = 1$ refers to that subject who gets assigned 11 points in a subject's ranking.

$$\mathrm{R} = \begin{cases} r_1 & \text{for upward-shifting} \\ r_{11} & \text{for downward-shifting} \end{cases}$$

We distinguish whether an individual is shifted to a higher or to a lower position. In the first case, i.e., positive team identity, the measure is taken as positive. In the latter case, i.e., negative team identity, it is set negative. It is also calculated by the number of positions the considered person is shifted - but downwards - divided by the possible maximum number the person can be shifted - but of course also counted downwards - and affixed with a negative sign.

If shifting is impossible because $r_i = R$ already holds, then $\hat{m}_2 = 0/0$ is formally yielded, which is set to 0 by definition. Hence $-1 \leq \hat{m}_2 \leq 1$. Finally, team identity of the voting subject is measured by the average degree of shifting he carried out for his two team fellows: $m_2 = 0.5 \cdot (\hat{m}_2(\text{fellow1}) + \hat{m}_2(\text{fellow2}))$[30]. Because it is impossible that $\hat{m}_2(\text{fellow1}) = \hat{m}_2(\text{fellow2}) = -1$ or $+1$[31], it is $-1 < m_2 < 1$.

In the case that one fellow cannot be shifted in one direction, because $r_i = R$ already holds and it can therefore not be decided whether the subject is team-minded or not, an average is not computed and only the one team-mate that can be shifted is taken into consideration.

Measure m_3

In order to measure not only the subjects' team identity by the number of shifted ranks, a further measure m_3 is developed, which additionally takes the amount of overleaped scores into account[32]. Hereby the measurement of team identity may be enhanced because there is an essential difference whether an in-group member who is ranked up, for instance to the preceding position, replaces a subject who solved 20 multiplications more or only 1. Indeed, one of the underlying questions for a subject confronted with the voting task is: How much is it worth for me to get my team fellow favored?

Formally, m_3 is constructed quite similar to m_2. We take the difference between the score of the shifted subject and the score of the subject whose rank is replaced. Then this quantity

[30] If a subject favors one team-mate and discriminates the other, calculating the average results in a neutralization of the two effects. At the most extreme case, m_2 can add up to zero. Thus, a m_2 of zero does not necessarily mean that a subject does not show team identity. However, by taking the average we want to account for the fact that a subject might intentionally discriminate one team-mate and might offset this by favoring the other.

[31] Since it would mean both team-mates are shifted to the first respectively last rank.

[32] Please see Appendix for a calculation example.

of scores is normalized by division through the maximum difference of scores that is feasible by a shifting into the same direction.

We denote for a specific subject the sequence of the other 11 subjects' achieved scores *in the reference ranking* as t_i with $i = 1, ..., 11$, and $0 \leq t_i \leq 100, t_i \geq t_{i+1}$ for the objective ranking and $t_i \leq t_{i+1}$ for the competitive ranking[33]. When a team fellow is shifted from rank i to rank J, we define for either team fellow $\hat{m}_3 = \text{sign} \cdot (t_i - t_J)/(t_i - T)$, with

$$\text{sign} = \begin{cases} +1 & \text{for upward-shifting} \\ -1 & \text{for downward-shifting} \end{cases}$$

$$\text{T} = \begin{cases} t_1 & \text{for upward-shifting} \\ t_{11} & \text{for downward-shifting} \end{cases}$$

and $m_3 = 0.5 \cdot (\hat{m}_3(\text{fellow1}) + \hat{m}_3(\text{fellow2}))$[34]. It is $-1 < m_3 < 1$.

Again, if formally $\hat{m}_3 = 0/0$, it is set to 0 by definition. The average is only computed when team identity is feasible in both directions for both team-mates. Otherwise only $\hat{m}_3$ is used. If the considered subject is shifted to a lower rank, the last rank is taken as the point of reference, i.e., in case of objective ranking the one with the lowest score and in case of competitive ranking the one with the highest score. When shifting a subject to a lower position, a negative number is assigned, so that measure $\hat{m}_3$ for an individual can vary in the range $-1 \leq \hat{m}_3 \leq 1$.

Values of measures m_2 and m_3

In the first part of this subsection we consider the averages of m_2 and m_3 which can be found in Figure 1.

As can be seen, team identity is, on average, shown in both treatments. The averages of m_2 and m_3 are both higher in GRT (0.29 and 0.22) compared to ANT (0.02 and 0.04). Thus, in GRT, the extent of positive team identity represents on average 29% of the maximal feasible upward rank shift and 22% of the maximal feasible upward score shift. Contrarily, in ANT, only 2% (4%) of the disposable favoring scale is applied. Moreover, m_2 is higher (smaller) than m_3 in GRT (ANT). Thus, accounting for the actual scores and not only for the ranks

[33] Again, $i = 1$ refers to the subject who is most preferred in a subject's ranking (i.e., gets 11 points).

[34] Again, we are aware that conducting the average of the team identity towards the two team-mates can cancel each other out if one team-mate is favored and the other is discriminated (see footnote 30 for a brief discussion of this issue).

leads on average to a comparatively smaller extent of team identity in GRT but to a higher extent of team identity in ANT[35].

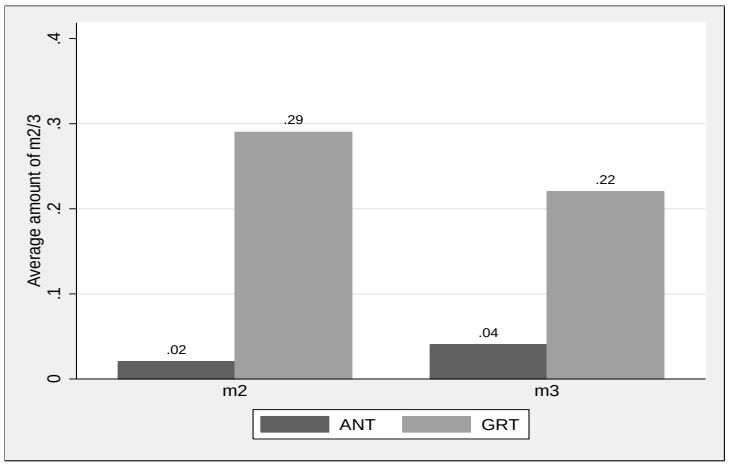

Figure 1: Average amounts for measures m_2 and m_3 dependent on treatment.

We can now test whether the average extent of team identity differs between ANT and GRT. Starting with m_2, a significant difference is found (Mann-Whitney-U test, $p = .004$, two-sided), which shows that the values of m_2 are substantially higher in GRT. A similar result can be found for m_3: When testing the values of m_3 in both treatments against each other, a significant difference is found (Mann-Whitney-U test, $p = .009$, two-sided). Again, m_3 is higher in GRT on average. These results show that the extent of team identity is higher in GRT than in ANT on average[36].

[35]This is due to the fact that in GRT team-mates whose score equals a non team member's score are often favored relative to this non team-mate.

[36]From a statistical point of view, it might be a little critical to test the values of m_2 and m_3 because each subject gets different information from the voting list (because a subject himself is not included). Therefore, the team identity of a subject depends on slightly different information. Consequently, we have to be careful with interpreting the results when testing the values of m_2 and of m_3. However, both measures are normalized and thereby comparable across subjects and deliver useful additional information.

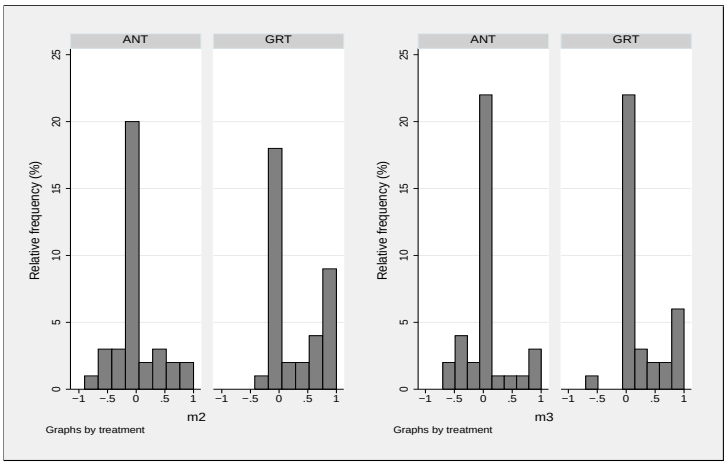

Figure 2: Distribution of m_2 and m_3 in both treatments.

As can be seen from Figure 2 and Table 6, m_2 and m_3 are zero for most subjects in both treatments. Thus, according to these two measures, many subjects do neither favor nor discriminate their team-mates[37]. In addition to that, in both treatments there are subjects who show positive values of m_2 and m_3 (positive team identity) and subjects who show negative values of m_2 and m_3 (negative team identity). As can be seen, the relative frequencies of positive m_2 and m_3 are higher in GRT than in ANT. Moreover, negative amounts of both measures are more often found in ANT than in GRT. This confirms the results of the last section where it is shown that positive (negative) team identity is significantly more often found in GRT (ANT) than in ANT (GRT).

	Positive values of				**Negative values of**				**Zero values of**			
	m_2		m_3		m_2		m_3		m_2		m_3	
Treatment	**RF**	**AV**	**RF**	**AV**	**RF**	**AV**	**RF**	**AV**	**RF**	**AV**	**RF**	**AV**
ANT	25.0	0.49	22.2	0.52	27.78	-0.35	27.78	-0.29	52.78	0	50.0	0
GRT	47.2	0.67	38.9	0.61	5.56	-0.46	2.78	-0.49	52.78	0	41.67	0

Table 6: Relative frequencies (RF, in %) and averages (AV) of m_2 and m_3 when positive, negative, or zero values of m_2 and m_3 are considered.

Moreover, as can also be seen in Table 6, the average extent of positive team identity is higher in GRT than in ANT. Thus, not only the frequency but also the extent of positive team identity is higher with face-to-face interaction than without. Concerning the extent of negative team identity, the average is also higher in GRT than in ANT[38]. However, the average extent of negative team identity is calculated for only 2 (1) persons in GRT concerning m_2 (m_3).

[37] Basically, a m_2 or m_3 which is equal to zero can also result if $\hat{m}_2(\text{fellow1}) = -\hat{m}_2(\text{fellow2})$ and accordingly for $\hat{m}_3$. But this case occurs only once in our whole data set.

[38] When testing the differences of the averages across treatments we do not find consistent significances.

3.4 Composition of the Elite-Team

An interesting question is whether the composition of the elite-team differs dependent on the treatment condition. All subjects profit from a high average performance of the elite-team. Consequently, the composition of the elite-team can be called efficient from an ex post point of view, when it is formed of the five subjects with the best performance at stage 3.

Taking a look at the distribution of performances at the last stage, our data show that subjects elect very good elite-teams, but not the efficient ones. To investigate the extent to which subjects in a certain session fail to compose the efficient elite-team, we consider the average of the actual elite-team's result in a session and compare it with the average result of those five persons with the highest scores at stage 3 in this session. To investigate inefficiencies on the treatment level, we calculate the averages per treatment which can be found in Table 7. The average difference per treatment is lowest in ICT (1.3), higher in GRT (2.3), and highest in ANT (4.9). Thus, subjects in ANT vote most inefficiently. But also in GRT and even in ICT subjects fail to elect the (complete) efficient elite-team. On average, the election of those subjects who should join the elite-team, because of their performance at stage 3, is quite similar across treatments.

	ICT	**ANT**	**GRT**
Average result of elite-team	94.7	89.9	94.9
Average result of efficient elite-team	96	94.8	97.2
Difference	1.3	4.9	2.3
Average number of efficiently elected subjects	3.5	3.3	3.7

Table 7: Comparison of actual and efficient elite-team per treatment (ex post point of view).

The previous analysis is carried out from an ex post point of view because the performances in the last stage are considered. But subjects in our experiment only know the performances of the first stage at the time of voting. In the light of this information subjects act efficiently when the 5 subjects who performed best in the first stage are elected to the elite-team[39].

As can be seen in Table 8, a different picture occurs when regarding the composition of the elite-team from an ex ante point of view. GRT is the only treatment where the best 5 subjects of stage 1 are elected to the elite-team. This means, in this treatment subjects elect an efficient elite-team, given their information at this stage.

[39] Although we use the same multiplication tasks in stage 1 and stage 3 (in just another order), subjects' performances across stage 1 and 3 change a bit.

	ICT	ANT	GRT
Average result of subjects at stage 1 who are elected to the elite-team	90.6	85.1	91.1
Average result of the best five subjects at stage 1	91.9	87.9	91.1
Difference	1.3	2.8	0
Average number of efficiently elected subjects	4	4	5

Table 8: Comparison of actual and efficient elite-team per treatment (ex ante point of view).

To investigate how efficiently subjects vote in our treatments from an ex ante point of view, we first calculate for each session the difference between the average performances at stage 1 of those subjects who are elected to the elite-team and of those who are the best 5 at stage 1 and, second, the average per treatment. We find that the composition of the elite-team is efficient from an ex ante point of view in GRT. The average difference is 1.3 in ICT and 2.8 in ANT.

It seems that the election of the elite-team is "distorted" because - besides team identity - several subjects use the competitive ranking. In GRT, in contrast to ANT, the influence of competitive ranking, of positive, and of negative team identity seem to cancel each other out resulting in efficient outcome from an ex ante point of view. Contrarily, the elite-teams in ANT are worst - both from an ex ante and an ex post point of view.

4 Summary and Discussion

In this paper we systematically investigate whether there exists an in-group/out-group bias that leads to a favoring of own team members as candidates in promotion relative to other teams and their members.

In contrast to psychological approaches, monetary incentives for voting choices are implemented. Using an experimental approach, we are able to isolate and analyze distortion effects on performance evaluation of in-group and out-group members. Moreover, objective performance criteria are defined and the extent of the in-group bias exactly measured.

Our results show that face-to-face interaction and communication and a joint team task can lead more subjects to favor own team-mates than in an anonymous team environment. In-group members are more willing to deliver a voluntary voting support to increase team mates' probability of promotion. This finding clearly confirms the first part of our hypothesis. Moreover, not only the frequency of positive team identity is higher but also the average extent. There are not only more subjects who favor own team-mates, but own team-mates are favored to a higher degree according to objectively measurable performance.

Our second main finding, which also supports the second part of our working hypothesis, is that mere anonymous team interaction can lead to a substantial disadvantage for own team members. This very interesting result extends and even contradicts previous findings of psychologists. When giving subjects the possibility to discriminate their team-mates in voting for promotion, they make use of negative shifting in their promotion ranking. Consequently the introduction of teams (groups) does not only lead to a favoring of own team-mates, as extensively investigated by psychologists, but also negative team identity can be detected.

How can the negative bias in performance evaluation and voting for promotion be explained? One reasonable explanation is that own team members represent a less abstract reference group for the decision making team member than the members of the other teams. In this case the initial (positive) fact, that a team is created, results in a negative outcome for the other team members. This can be due to the fact that the other team members are closer to the decision maker and therefore more exposed to competition. A second assertion that points into the same direction and that also gets support by participants' written comments is that under anonymous conditions a subject attempts to be the "winner" of the team, who is elected and promoted to the elite-team. To reduce their team-mates' chances they are discriminated by assigning less votes to them. In the case when a subject thinks that his chances are very small for getting into the elite-team, he wants to prevent other team members to join it because of envy. Instead, the advantage is left to a third person, who is more distant.

A third contribution of this paper is that we show that a totally anonymous promotion process does not lead to an efficient composition of the elite-team because several subjects behave competitively and disadvantage own team members in the voting stage. Contrarily, in a situation where subjects can easily communicate and have joint group tasks, the negative influence of competitive behavior seems to be cancelled out by positive team identity resulting in an efficient elite-team from an ex ante point of view. From this perspective, GRT is the treatment which would be favored by the management of an organization. Hence, company leaders should care for intensive face-to-face communication and identification creation among team members. The sole creation of virtual or rapidly changing teams, where members have no constant contact, can lead to inferior outcome when subjects shall be determined for promotion. Creating a team context with individuals pursuing their own interest may even negatively affect the efficiency of selected teams. To counteract this, employers should foster social identity between team-mates and not exclusively focus on economic goals.

Another finding is that the mere introduction of teams, a joint team task, an implemented

team bonus, and face-to-face communication do not substantially influence subjects' choice of the reference strategy for voting. This may be due to the fact that a certain distribution of "competition-types" and associated beliefs inherently exists, which remains unaffected by experimental treatment conditions. However, on average there is a slightly decreasing tendency that these types are detected in GRT.

Future work can focus on more diverse effort tasks, which are more sensitive to environmental and organizational changes, to investigate how performance and promotion are connected. The real-effort task in our experiment is one-dimensional and just requires simple mathematical skills. Other characteristics or abilities (like e.g., social competence) were not needed to solve the given tasks. By implementing different types of tasks that require complementary skills it can be investigated how the change of multi-level performances interacts with occurrences and extends of positive and negative team identity.

Furthermore, the composition of teams can be varied. For instance, by the introduction of culturally mixed teams it could be studied how these factors influence team identity and promotion processes. Also interesting is the influence of crossed categorization on the in-group/out-group bias: Consider for example an extension of our experiment, where each team consists of two Germans and one Chinese. Would the Chinese then favor his own team members or the Chinese students of the other teams?

Finally, another interesting aspect is the formation of social networks across borders by intentionally fostering or disadvantaging colleagues. This fact seems to be important nowadays with working places being frequently changed and abandoned and former team-mates being spread all over the world. Accounting for these aspects will lead to a better understanding of team interaction - within and between heterogeneously composed teams - in globalized business environments.

References

Brewer, B. (1979): "In-group Bias in the Minimal Intergroup Situations: A Cognitive-Motivational Analysis", *Psychological Bulletin*, 86(2), 307-324.

Buchan, N., E. Johnson, and R. Croson (2006): "Let's get personal: an international examination of the influence of communication, culture and social distance on other regarding preferences", *Journal of Economic Behavior and Organization*, 60, 373-398.

Charness, G., L. Rigotti, and A. Rustichini (2007): "Individual Behavior and Group Membership", American Economic Review, 97(4), 1340-1352.

Chen, Y. and X. Li (2007): "Group Identity and Social Preferences", *Working Paper*.

Crisp, R.J. and M. Hewstone (2000): "Crossed Categorization and Intergroup Bias: The Moderating Roles of Intergroup and Affective Context", *Journal of Experimental Social Psychology*, 36, 357-383.

Doise, W. and A. Sinclair (1973): "The categorization process in intergroup relations", *European Journal of Social Psychology*, 3, 145-157.

Downing, L.L. and N.R. Monaco (1986): "In-Group/Out-Group Bias as a Function of Differential Contact and Authoritarian Personality", *The Journal of Social Psychology*, 126(4), 445-452.

Eckel, C. and P. Grossman (2005): "Managing Diversity by Creating Team Identity", *Journal of Economic Behavior and Organization*, 58, 371-392.

Ferguson, C.K. and H.H. Kelley (1964): "Significant factors in overevaluation of own-group's product", *Journal of Abnormal and Social Psychology*, 69, 223-228.

Goette, L., D. Huffman, and S. Meier (2006): "The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups", *IZA Discussion Papers No. 06-7.*

Greiner, B. (2003): "The Online Recruitment System ORSEE - A Guide for the Organization of Experiments in Economics", Papers on Strategic Interaction, Max Planck Institute for Research into Economic Systems, *Discussion Paper 10-2003.*

Grund, C. and D. Sliwka (2005): "Envy and Compassion in Tournaments", *Journal of Economics and Management Strategy*, 14, 187-207.

Gürtler, O. (2008): "On Sabotage in Collective Tournaments", *Journal of Mathematical Economics*, 44, 383-393.

Harbring, C. and B. Irlenbusch (2008): "How many Winners are Good to Have? On Tournaments with Sabotage", *Journal of Economic Behavior and Organization*, 65, 682-702.

Harbring, C., B. Irlenbusch, M. Kräkel, and R. Selten (2007): "Sabotage in Corporate Contests - An Experimental Analysis", *International Journal of the Economics Business*, 14, 367-392.

Kraiger, K. and J.K. Ford (1985): "A meta-analysis of ratee race effects in performance ratings", *Journal of Applied Psychology*, 70, 56-65.

Kräkel, M. (2005): "Helping and Sabotaging in Tournaments", *International Game Theory Review*, 7, 211-228.

Lazear, E.P. (1989): "Pay Equality and Industrial Politics", *Journal of Political Economy*, 97, 561-580.

McLeish, K.N. and R.J. Oxoby (2007): "Identity, Cooperation, and Punishment", *IZA Discussion Paper No. 2572.*

Mummendey, A., B. Simon, C. Dietze, M. Grünert, G. Haeger et al. (1992): "Categorization is not enough: intergroup discrimination in negative outcome allocations", *Journal of Experimental Social Psychology*, 28, 125-144.

Otten, S. and D. Wentura (1999): "About the impact of automaticity in the minimal group paradigm: evidence of the affective priming task", *European Journal of Social Psychology*, 29, 1049-1071.

Rabbie, J.M., J.C. Schot, and L. Visser (1989): "Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model", *European Journal of Social Psychology*, 19, 171-202.

Rabbie, J.M. and M. Horwitz (1969): "Arousal of Ingroup-Outgroup Bias by a Chance Win or Loss", *Journal of Personality and Social Psychology*, 13(3), 269-277.

Ruffle, B. and R. Sosis (2006): "Cooperation and the in-group-out-group bias: A field test on Israeli kibbutz members and city residents", *Journal of Economic Behavior and Organization*, 60, 147-63.

Schaller, M. (1992): "In-Group Favoritism and Statistical Reasoning in Social Inference: Implications for Formation and Maintenance of Group Stereotypes", *Journal of Personality and Social Psychology*, 60(1), 61-74.

Sherif, M., O.J. Harvey, B.J. White, W.R. Hood, and C.W. Sherif (1961): "Intergroup conflict and cooperation: The Robbers Cave experiment", Norman, Oklahoma: University Book Exchange.

Solow, J.L. and N. Kirkwood (2002): "Group identity and gender in Public goods experiments", *Journal of Economic Behavior and Organisation*, 48, 403-412.

Suls, J. and L. Wheeler (2000): *Handbook of Social Comparison. Theory and Research*, Editors: J. Suls and L. Wheeler, Kluwer Academic/Plenum Publishers, New York. Series editor: C.R. Snyder. The Plenum Series in Social/Clinical Psychology.

Tajfel, H., M.G. Billig, and R.P. Bundy (1971): "Social categorisation and intergroup behaviour", *European Journal of Social Psychology*, 1(2), 149-178.

Tajfel, H. and J.C. Turner (1979): "An integrative theory of intergroup conflict", in W.G. Austin and S. Worchel (Eds.), *The social psychology of intergroup relations*, 33-47. Monterey, CA: Brooks/Cole.

Wit, A.P. and H.A.M. Wilke (1992): "The effect of social categorization on cooperation in three types of social dilemmas", *Journal of Economic Psychology*, 13, 135-151.

Yamagishi, T., J. Nobuhito, and T. Kiyonari (1999): "Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism", in: *Advances in Group Processes*, ed. S.R. Thye, E.J. Lawler, M.W. Macy, and H.A. Walker, pp. 161-97. Stamford, CT: JAI.

YOUNG, H.P. (1974): "An Axiomazation of Borda's Rule", *Journal of Economic Theory*, 9, 43-52.

YOUNG, P. (1995): "Optimal Voting Rules", *Journal of Economic Perspectives*, 9, 51-64.

Appendix

A calculation example

In the first session of GRT, subject c with team-mates' scores 82 resp. 71 states the following individual ranking which is compared to the objective ranking:

Individual ranking	**Objective ranking**
82	99
99	98
98	82
78	78
71	75
75	75
75	71
69	69
66	66
52	52
47	47

The objective ranking is taken as reference point for subject c because the performances in his ranking are predominantly decreasing: $m_1 = 3/10 = 0.3$.

Measure m_2:

$$\hat{m}_2(82) = 2/2 = 1$$
$$\hat{m}_2(71) = 2/6 = 0.33$$
$$m_2 = (1 + 0.33)/2 = 0.67$$

Measure m_3:

$$\hat{m}_3(82) = 17/17 = 1$$
$$\hat{m}_3(71) = 4/28 = 0.14$$
$$m_3 = (1 + 0.14)/2 = 0.57$$