Universitat Autònoma de Barcelona

Department of Business

Doctorate in Economics, Management, and Organization

DOCTORAL THESIS

THREE ESSAYS ON COMMUNICATION AND ADVICE IN EXPERIMENTAL ECONOMICS

Author:
Christina Rott

Doctoral Thesis Supervisor:
Jordi Brandts Bernad, Ph.D.

Barcelona, June 2014

ABSTRACT

The doctoral thesis is composed by three essays, in each of which we analyze with laboratory experiments how individuals use and react to different communication and advice forms. Individuals make their decisions anonymously and communication and advice, either free form or pre-formulated, is transmitted through the computer.

In the first essay, we study how advice by a more experienced and better-informed person affects an individual's entry into a real-effort tournament and the gender gap. Overall, advice improves the entry decision of subjects and the improvements are mainly driven by increased entry of strong-performing women, who also become more confident, and reduced entry of weak-performing men. We find that the overall gender gap persists even though it disappears among low and strong performers. The persistence is due to an emerging gender gap among intermediate performers.

In the second essay, we investigate how after a history of decay in cooperation, organizations can revive cooperation in a repeated voluntary contribution game in an enduring way. Simply starting the repeated game over - a pure restart - leads to an initial increase of cooperation, but to a subsequent new decay to the previous level. Motivated by cooperation failure in organizations we study the potential of three interventions of triggering higher and sustained cooperation. We find that a detailed explanation of the causes of the decay in cooperation combined with an advice on how to prevent the decay does not have an effect beyond that of just starting over. In contrast, a one-way free form communication message sent by the leader to the followers strongly revives cooperation, independent of being preceded by explanation and advice.

In the third essay, we study how the separation between making and communicating a choice affects fairness and reactions to harsh decisions. A decision-maker allocates a fair or unfair amount of money to herself, two receivers, and a third party. The decision-maker or the third party communicates the allocation chosen to the receivers, who then decide whether to punish or not. With aligned punishment, receivers have to target both decision-maker and third party with the same amount of punishment, whereas with independent punishment they are free to decide whom to punish. Decision-makers choose more often the unfair allocation when punishment is aligned as opposed to independent, but in both cases, decision-makers who choose the unfair allocation are more likely to delegate the communication to the third party. With independent punishment, receivers punish the decision-maker and the third party more when the later communicates the unfair allocation decision. The third party expresses more often remorse than need as opposed to decision-makers, which receivers seem to perceive as an attempt of shifting blame with independent punishment.

ACKNOWLEDGEMENTS

A very special and enriching period of my life is about to finish to give space to new wonderful experiences. Thank you to all those who shared part of this life span with me, who support me or with whom I had a great time. I am very lucky to be surrounded by so many wonderful people. I would like to express my honest acknowledgement to all of you though I may forget to mention some by name given the advanced (or early) time of the day and a closer coming deadline…

I would like to thank the Business Department of Unversitat Autònoma de Barcelona and the colleagues and professors for the opportunity to learn and work with you. I am very grateful for all the honest advice and support for the doctoral thesis and for research and professional life in general from many colleagues and friends. A particular thanks to my coauthors.

Furthermore, my thanks go to many colleagues and friends I met during the journey through the Ph.D., be it during several research stays in New York (a particular thank you to Ernesto and Kika), at conferences or in discussions at different universities or elsewhere. Part of the New York group will meet in the Netherlands again, and another part will simply fly in from Switzerland.

La oficina buena will soon be spread all over the place, but the memories will remain and Europe is not that big. Thanks to all of you for these great years in which we had a great time in our office and at sunnier places (basically all places other than the office)!

Finally, I want to thank my great friends Steffi and Valeska and all my very good friends from Barcelona and back home for the unforgettable moments and laughers. Miguel and my wonderful family, in particular my parents and my sisters Manu, Mimi, and Danny, have constantly supported and encouraged me in this phase and shared all the difficult and happy moments. I am extremely grateful to have you and thank you very much! Without you this would not have been possible.

Last, but not least, a very special thank you to Jordi. I am glad that you decided to supervise my doctoral thesis and appreciate very much your constant support during the Ph.D. in all possible way.

TABLE OF CONTENTS

Chapter IV: Spokesperson – To Be or not to Be?

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Every day, people communicate with one another. Communication can take different forms, which can be roughly split into two groups: Verbal communication and non-verbal communication. Whereas the former is characterized by the expression of words in written or spoken form, the later can take different forms such as gestures, facial expression, or signs. Even though communication is often non-binding, i.e. cheap talk, it has been shown to improve outcomes in different contexts in which individuals interact, for example coordination (Brandts and Cooper 2007) or cooperation (Koukoumelis et al. 2012) situations.

A particular form of verbal communication is advice, in which one or several individual(s) give(s) a recommendation or opinion to another (group) of people as to what action to take in order to reach a desired outcome. Advice occurs on a regular basis in the private life, for example between parents and children or among friends, as well as in the professional life between professors and students or among colleagues. There is wide evidence in the experimental economic and organizational psychology literature that advice improves outcomes in different situations (Bonaccio and Dalal 2006, Brandts, Groenert, and Rott 2014; for a review see Schotter 2003) and thus constitutes an important element of human interaction.

The doctoral thesis is composed by three essays, each of which is presented in a separate chapter. In each essay, we analyze with laboratory experiments how individuals use and react to verbal communication and advice. Individuals make their decisions anonymously and communication and advice, either free form or pre-formulated, is transmitted through the computer. Laboratory experiments bear the advantage of a controlled environment in which the usage of communication and the effect thereof on decisions can be documented and analyzed in detail. Furthermore, they allow analyzing the causal effect of (cheap talk) communication and advice on behavior, which is more difficult with empirical data.

In the second chapter of the doctoral thesis, we study how advice by a more experienced and better-informed person affects an individual's entry into a real-effort tournament and the gender gap. Our experiment is motivated by the concerns raised by approaching the gender gap through affirmative action policies. Overall, advice improves the entry decision of subjects, in that forgone earnings due to wrong entry decisions go significantly down. The improvements are mainly driven by increased entry of strong-performing women, who also become more confident, and reduced entry of weak-performing men. We find that the overall gender gap persists even though it disappears among low and strong performers. The persistence is due to an emerging gender gap among intermediate performers driven by women (men) following more the advice to stay out of (enter) the tournament in this performance group.

In the third chapter, we analyze how after a history of decay in cooperation, organizations can revive cooperation in a repeated voluntary contribution game in an enduring way. Simply starting the repeated game over - a pure restart - leads to an initial increase of cooperation, but to a subsequent new decay to the previous level. Motivated by cooperation failure in organizations we study the potential of three interventions of triggering higher and sustained cooperation. We find that a detailed explanation of the causes of the decay in cooperation combined with a concrete advice do not have an effect beyond that of just starting over. In contrast, a one-way free form communication message sent by the leader to the followers strongly revives cooperation. Repeated free form communication by the leader further strengthens the reviving effect on cooperation by eliminating the decline in contributions over time almost completely. Letting the leader's communication be preceded by a detailed explanation of the causes of the decay in cooperation combined with a concrete advice does not outperform the pure effect of communication.

In the fourth chapter, we evaluate how the separation between making and communicating a choice affects fairness and reactions to harsh decisions. In organizations and institutions, decision-makers frequently let a spokesperson communicate their choices to those affected. We conduct a laboratory experiment in which a decision-maker allocates a fair or unfair amount of money to herself, two receivers, and a third party. Either the decision-maker or the third party, i.e. the spokesperson, communicates the allocation chosen to the receivers, who then decide whether to punish or not. Receivers can punish in two different ways. In the aligned punishment, receivers are forced to target both decision-maker and spokesperson with the same amount of punishment, whereas in the independent punishment they are free to decide whom to punish. The perception of the delegation decision and of the communication strategy seems to differ strongly depending on the punishment form imposed. Decision-makers choose more often the unfair allocation when punishment is aligned as opposed to independent, but in both cases, decision-makers who choose the unfair allocation are more likely to delegate the communication to the third party. With independent punishment, receivers punish the decision-maker and the third party more when the later communicates the unfair allocation decision. The third party expresses more often remorse than need as opposed to decision-makers, which receivers seem to perceive as an attempt of shifting blame with independent punishment.

REFERENCES

Brandts J, Cooper D (2007) It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. *Journal of the European Economic Association* 5(6): 1223–1268.

Brandts J, Groenert V and Rott C (2014) The impact of advice on women's and men's selection into competition. *Management Science*, forthcoming.

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101(2): 127–151.

Koukoumelis, A., M. V. Levati, and J. Weisser (2012), Leading by words: A voluntary contribution experiment with one-way communication, *Journal of Economic Behavior and Organization*, 81(2), 379-390.

Schotter A (2003) Decision making with naive advice. *American Economic Review* 93(2): 196–201.

CHAPTER II: THE IMPACT OF ADVICE ON WOMEN'S AND MEN'S SELECTION INTO COMPETITION

1. Introduction

The underrepresentation of strong-performing women in top-level jobs continues to be a major concern of managers, politicians, and society in general and has received much attention by researchers in economics and other fields. For some time research on this topic focused on explanations related to differences in human capital, discrimination, and child rearing. More recently, the literature has explored an additional explanation: gender differences in entry into competition as studied in the seminal paper by Niederle and Vesterlund (2007). Participants in their experiment perform a simple addition task, for which they are given the choice between a competitive compensation scheme (in which payoffs depend on relative performance) and a noncompetitive one (in which payoffs depend only on own performance). They show that women opt for the competitive scheme far less than men do, though there are no significant gender differences in performance. In other words, there exists a gender gap in entry into competition. This gender gap is mainly driven by strong-performing women entering too little and weak-performing men entering too often.

Subsequently, a number of papers have explored mechanisms to reduce this gender gap. One set of papers considers the effects of affirmative action, such as quotas or other forms of positive discrimination (Niederle et al. 2013, Balafoutas and Sutter 2012). It is argued that these measures are effective and do not substantially harm efficiency. However, affirmative action policies are highly controversial, and it remains difficult to find majorities for them. Despite the results in the aforementioned studies, there remains a concern that they do not lead to efficient allocations, possibly promoting weak-performing women at the cost of strong-performing men. Another concern is that they actually harm women because women who obtain a job under an affirmative action scheme may be stigmatized of being selected only because of this action. Another set of papers shows that the provision of relative performance feedback leads to a significant improvement of the competition entry behavior and a reduction of the gender gap (Wozniak et al. 2014).

The idea behind our experiment is inspired by the literature on naive advice and the fact that insiders or seniors in firms and institutions possess information and experience about the qualifications necessary for a post. The advice literature consistently shows that advice improves decisions in games such as the ultimatum game or the battle of the sexes game (for a survey, see Schotter 2003). A key difference in our experimental design is that beliefs about own relative performance play an important role for the decision, which is, to our knowledge, not the case in previous studies on the effect of advice. From the results in previous studies, it is not clear whether advice can improve decisions when confidence about own relative performance matters. Our aim is to study whether receiving advice from a person who has experienced the same particular competitive situation and has some informational advantage can improve competitive choices and therefore constitute an alternative to affirmative action. Advice can occur naturally in everyday situations, or one might channel advice through institutions such as mentoring programs.[1]

[1] Two examples of mentoring programs are the one organized by the organization Women in Technology and the one by the Committee on the Status of Women in the Economics Profession (CSWEP; see Blau et al. 2010 for an analysis of the effects of the mentoring program run by CSWEP). Our focus is on the advice component rather than the training or role model component of mentoring programs. It is worth mentioning here that mentoring programs typically have costs associated with them, which would need to be considered in a more complete analysis.

For purposes of maximal comparability, our experimental design closely follows the design in Niederle and Vesterlund (2007). Participants face a real-effort task, which consists of adding up series of five two-digit numbers. They have five minutes to solve as many problems as possible. Subjects complete this task once under a piece rate scheme and once under a winner-take-all tournament scheme (for the tournament they are matched in groups of four). Then, before performing the addition task a third time, they are given the choice between being paid under the piece rate or the tournament scheme. To help with this decision, subjects in our advice treatment receive a message from an advisor, who can either recommend choosing the piece rate or the tournament. Advisors are subjects who had to make the same decision earlier in the session and who received information about the performance levels in the first two rounds of their own group and of their advisee. We chose this design because it reflects in the most natural way how advice is usually given: in most situations, both previous experience with the same decision as well as information about their own generation's and the advisee's performance influence advisors' recommendation. Wozniak et al. (2014), Ertac and Szentes (2011), and Ewers (2012) show that the effect of information feedback of different types on entry behavior is positive. Our focus is not on information feedback. Instead, we want to see how people react to advice from another (experienced) person who is somewhat better, but not perfectly informed. In particular, advisors do not know how the advisee performs compared to the advisee's group members, but only compared to a small sample of the advisor's own generation, meaning that the information content is lower than in the design in Wozniak et al. (2014). In situations in which an objective performance ranking is not available in the short run, an experienced and better-informed insider might be able to compare a potential applicant's performance to previous observations and give good advice. Examples for such situations are performance requirements for top-level jobs. In addition, our design can be seen as a first step toward studying advice more in depth, involving matching advisees and advisors of the same or the other gender, more interactive forms of advice such as free-form communication, and people's advice seeking and giving behavior.

Our results show that advice improves the selection into competition by various measures. Without advice, women and men who enter the competition do not perform significantly better than those who do not. With advice those who enter are the ones with significantly stronger performance. We also calculate the opportunity cost of taking the "wrong" entry decision. There are two types of "mistakes," weak-performing subjects (whose expected payoff would be higher under the piece rate) who enter the competition and strong-performing subjects (whose expected payoff would be higher in the tournament) who do not. The forgone earnings from the two types of mistakes are significantly lower in our treatment with advice than in the control group without advice. Examining who are the subjects that improve their entry decisions under advice, we find that it is in particular the strong-performing women who enter significantly more (an increase from about 40% to over 80%) and the weak-performing men who enter significantly less (a decrease from over 60% to about 25%). The gender gap persists with advice, because a gender gap emerges among intermediate performers—men (women) follow more the advice "tournament" ("piece rate")—whereas it disappears among weak and strong performers. We also find that advice changes confidence levels and that this is an important determinant of the entry decision.

The rest of this paper is organized as follows. Section 2 provides an overview of the literature. Section 3 describes the experimental design and presents our main research questions. Section 4 contains all results. Section 5 concludes.

2. Literature Review

We first discuss the relevant literature on gender differences in competitiveness and then review the relevant literature on advice and mentoring.

2.1. Competitiveness and Gender Differences

An important stream of the literature on gender differences in competitive behavior (for a survey, see Croson and Gneezy 2009) has focused on the tendency to select into competition. These studies started with the discussion of how incentive schemes affect women's and men's performance differently (Gneezy et al. 2003). The gender gap in entry into competition found by Niederle and Vesterlund (2007) seems to be quite robust, as the results in a number of papers with similar designs show (see, e.g., Balafoutas and Sutter 2012, Booth and Nolen 2012, Cason et al. 2010, Dargnies 2012, Niederle et al. 2013, Sutter and Glätzle-Rützler 2014, Wozniak et al. 2014). In addition, similar gender gaps were found under a variety of different designs (see, e.g., Gupta et al. 2013, Gneezy et al. 2009, Dohmen and Falk 2011) and in the field (Flory et al. 2010).

One important reason for the gender gap in entry into competition seems to be that men are relatively more overconfident (for a review, see Croson and Gneezy 2009), in particular in a tournament environment (Charness et al. 2011, Reuben et al. 2012). Beliefs about one's performance are important not only on the supply side of the labor market. Evidence shows that hiring decisions can be discriminatory against women due to biased beliefs about their abilities (Reuben et al 2014). Niederle and Vesterlund (2007) underline that, in addition, preferences for competing play an important role. Some studies find that the gender gap remains after controlling for some relevant factors. In their study of young children, Sutter and Glätzle-Rützler (2014) find that the gender gap between boys and girls in entering a competition is robust to controlling for gender differences in risk attitudes, overconfidence, and past performance. In their study with students around 15 years old, Booth and Nolen (2012) also observe a gender gap after controlling for risk attitudes and past performance, but other studies do not find significant gender differences in the willingness to compete when controlling for other factors such as distributional preferences, risk attitudes, and past performance (Balafoutas et al. 2012), or for confidence levels (Charness et al. 2011). All of these studies use math tasks. Grosse and Riener (2010) show that, after controlling for differences in performance, risk attitudes, and overconfidence, the gender gap persists only in a quantitative, but not in a verbal task, suggesting gender task stereotypes as another explanation for the gender gap.

A number of papers have looked at designs to mitigate or even overcome the gender gap in entry. Niederle et al. (2013) show that affirmative action in form of a quota increases female participation in a tournament. Balafoutas and Sutter (2012) also confirm that quotas and other forms of positive discrimination (preferential treatment and repetition of competition if a man wins) encourage women to enter competitions. Both of the aforementioned studies argue that this is achieved with at most a modest efficiency loss and driven by the increased entry of high performance women. Wozniak et al. (2014) show that providing feedback on relative performance can eliminate the gender gap. Shurchkov (2012) shows that reduced time pressure associated with the competitive setting increase female competitive choices. Finally, Dargnies (2012) shows that the gender gap disappears if participants compete as part of a team. This is a consequence of men choosing to compete significantly less often when they have to compete in a team rather than alone.

2.2. Advice and Mentoring

The main conclusions from the experimental economics literature on decision making with naive advice (in the sense of uninformed word-of-mouth advice compared to advice from experts) are summarized in the survey by Schotter (2003). The experimental design employed in the surveyed papers is usually one where participants are split into two nonoverlapping “generations.” Participants of the first generation experience the decision-making situation once and then become advisors to participants of the second generation, who face the same situation. Schotter (2003) highlights five results: (1) subjects tend to follow advice; (2) advice tends to change behavior; (3) decisions with advice are closer to theoretic predictions; (4) if subjects can choose between receiving advice or the information on which the advice is based, they opt for the advice; (5) advice improves decisions because it forces advisees to think about the problem.[2]

In a similar vein, several field experiments and theoretical papers study the effect of mentoring and coaching on performance, success, and behavior. In particular, Blau et al. (2010) suggest that mentoring programs for female assistant professors lead to a significant increase in performance in terms of publications and grants obtainment.[3] Athey et al. (2000) study theoretically how mentoring can lead to different steady states, including a “glass ceiling.” They assume that mentor–mentee matching takes place among employees of the same type (e.g., gender), whereby one type (e.g., men) represents the majority at the upper level. In our experimental design, advice does not affect ability, and all participants receive advice independent of their gender.

Finally, advice has been extensively studied in the organizational psychology literature. Bonaccio and Dalal (2006) provide a review of that literature. In the typical setup, the decision maker is asked for a tentative decision on the problem and the advisor makes a recommendation to the decision maker, who then gets the opportunity to revise his original choice. The main results that pertain to our experiment are that (1) advice improves the accuracy of final decisions (in many setups there is a correct answer to the problem such as estimating the year of a specific event in U.S. history; Gino 2008), and (2) advice is discounted, that is, it is not fully utilized. The following factors may influence advice utilization and the accuracy of final decisions: (a) whether the decision maker is asked to form an initial opinion (in particular, if people display a confirmation bias, this may make a difference); (b) whether the decision maker is given the option to solicit advice; (c) the number of advisors; (d) the type of decision, e.g., whether the problem has a correct answer or it is rather a choice or a judgment problem; (e) the amount and type of interaction between advisor and decision maker. Factors that reduce advice discounting are the advisor’s expertise and amount of information she holds, the quality of the advice, making advice costly, increasing task complexity, and congruence of the goals of the advisor and the decision maker.

3. Experimental Design and Research Questions

We will first describe the basic experimental design with regard to the choice of participating in a competition and then turn to the design of the advice part of the experiment and some further information on the design.

[2] These conclusions are drawn from the results in Celen et al. (2010), Iyengar and Schotter (2008), Schotter and Sopher (2003, 2007), and Chaudhuri et al. (2009).

[3] See also Bettinger and Baker (2011) and Rodrìguez-Planas (2010) for some more loosely related field experiments.

3.1. The Basic Setup

For the purpose of maximal comparability, we keep the experimental design regarding the participation decision as close as possible to the one in Niederle and Vesterlund (2007), which involves three real-effort tasks, two entry decisions, and two self-evaluations. In the real-effort task, participants have five minutes to add up sets of five two-digit numbers without using a calculator. (See the screenshot provided with the instructions in the online appendix, available at http://docs.google.com/file/d/0B7zDSay6dz3SWi0yWmg2OC1OYVE/edit?pli=1) Subjects first perform the task under the noncompetitive payment scheme (task 1, piece rate), then under the competitive payment scheme (task 2, tournament), and finally, before performing a third time, they have to choose between the competitive and the noncompetitive payment scheme (task 3, choice). In addition, in task 4 subjects have to decide whether to apply the competitive or the noncompetitive payment scheme to their (past) task 1 performance. Finally, subjects have to rank their performance in tasks 1 and 2 relative to the group members' performances on a scale from 1 (best) to 4 (worst), respectively.

Under the piece rate payment scheme, subjects receive €0.5 for each correct sum. For the competitive payment scheme, subjects are matched in groups of four, and only the person with the best performance receives payment in form of €2 for each correct sum. If a subject chooses the competitive payment scheme in task 3, her task 3 performance is evaluated against the task 2 performance of her group members. Thus a subject "wins" the tournament in tasks 2 and 3 if she solves more problems correctly than each of her group members in task 2. Ties are broken randomly among the best performers. The fact that subjects in task 3 compete with the performance of subjects in task 2 ensures that a subject's entry decision is not influenced by beliefs about the other subjects' entry decisions.

3.2. Advice

Upon arrival to the experiment, subjects are divided randomly and evenly into two different rooms (separated by a glass window). Subjects in one room have the role of the advisors, and subjects in the other room the role of the advisees (but they do not learn about their roles until the advice stage begins). The group of advisors, who complete tasks 1–4 as in the original Niederle and Vesterlund (2007) setup, i.e., without receiving advice, also serves as our control group to evaluate the effect of receiving advice. We will refer to this group as generation 1. The advisees form generation 2.

The advice stage for advisors follows after they have completed tasks 1–4 and the self-evaluation task, whereas the advice stage for advisees follows after they have completed tasks 1 and 2, that is, immediately before they have to choose the payment scheme for task 3; see Table 2.1. The advice consists of a recommendation as to the choice of the payment scheme for task 3 and (possibly) some reasons for the recommendation. Each advisee is randomly matched to exactly one advisor, and each advisor has only one advisee. Advisor and advisee do not learn the gender of each other. An advisor is paid 50% of her advisee's task 3 earnings. We reward advisors because the main objective of the advice incentive system is to make the advisors give "good" advice. On average, €4.02 was earned for the advice, with payoffs ranging from €0 up to €27. Thus incentives to give "good" advice were substantial. In the real world, the reward of advisors can be nonmonetary in form of building a reputation or in form of a good feeling because of giving good advice to somebody.

The exact sequencing of the advice stage is as follows. The advisee sends information about his task 1 and 2 performances to his advisor.[4] Upon receiving this information, the

[4] A few weak-performing (strong-performing) advisees sent the advisor higher (lower) performance information than the actual performance level. This did not reduce the quality of advice.

advisor sends a message, telling her advisee whether or not she recommends entering the competition. The advisor is then asked to give her advisee a reason for her recommendation.

We provide three preformulated reasons for each recommendation ("tournament" or "piece rate") from which the advisor can select as many as she wishes to. We chose the two-way design of the advice stage to create a feeling of interaction between advisor and advisee.[5] After having received the advisor's recommendation and (possibly) some reasons for this recommendation, the advisee decides whether to enter the competition in task 3.

Table 2.1: Timeline of Tasks and Compensation Schemes in the Experiment

Generation 1 (advisors)	Generation 2 (advisees)
Task 1: Five-minute addition task – Piece rate (€0.5)	
Task 2: Five-minute addition task – Tournament (€2, winner takes all)	
Task 3: Five-minute addition task – Selection of compensation scheme	
Task 4: Selection of compensation scheme task 1	Task 1: Five-minute addition task – Piece rate (€0.5)
Self-evaluation, task 1 and 2 (€1 per correct guess)	Task 2: Five-minute addition task – Tournament (€2, winner takes all)
Performance feedback, own group	
One advisor randomly matched with one advisee (gender not revealed)	
Receive advisee's performance info	*Send own performance info*
Give advice (50% of the advisee's task 3 earnings)	*Receive advice*
Choose up to three reasons (preference for competition, confidence, risk preferences)	*Receive up to three reasons (preference for competition, confidence, risk preferences)*
	Task 3: Five-minute addition task – Selection of compensation scheme
	Task 4: Selection of compensation scheme task 1
	Self-evaluation, task 1 and 2 (€1 per correct guess)
	Performance feedback, own group

Advisors in our experiment are no experts in the task, but they have experienced the situation once and hold some informational advantage. At the end of the self-evaluation task, each participant receives feedback on task 1 and task 2 performances of all her group members. Therefore, advisors have not only made one entry decision, but also have seen how people perform in the addition task in a small sample of four people. Note, however, that when the advisors receive the information about performance levels in their group, they do not yet know about the advice stage. We chose this design because we felt that this was the most natural setup. Usually, a person who has previously participated in a competition task will have some idea about performance levels in that task, but it is not so clear whether she will remember correctly. Finally, an advisor knows that her advisee has just completed tasks 1 and 2.

Similarly, an advisee knows that his advisor has just completed all tasks and has information about the task 1 and task 2 performances of her own group. However, the advisee does not know if, or how, his advisor is compensated for giving advice. We wanted to eliminate the influence of social preferences on the advisee's entry decision and therefore do

[5] Ideally, we would have liked the interaction between advisor and advisee to be less structured, with them being able to communicate in a free-form chat with each other. However, since our experiments were conducted in Spanish, we were concerned that the use of adjectives could reveal a subject's gender to their matched partner, which could have led to effects stemming from gender pairing.

not provide the advisee with information about the monetary consequences for the advisor. In addition, advisees are put in a more natural advice environment if they do not know the advisor's incentives. Advisees do know that their advisor has some informational advantage, but they need to trust that the advisor will advise them correctly. In previous experimental studies on the effect of information on the relative performance on tournament entry, participants learn their relative performance with respect to the other group members (Wozniak et al. 2014, Ewers 2012) or the performance distribution in general (Ewers 2012) and know that this information is correct. Three key differences from these studies are that advisees in our experiment receive some (imperfect) information indirectly through another person who has made the same decision previously, the information is less accurate than in Wozniak et al. (2014), and advisees need to trust in the correctness of the advisor's recommendation.[6]

3.3. Group Composition, Procedure, and Subject Pool

Each group of four participants who compete against each other is composed of two women and two men.[7] We made sure, however, that participants were not aware that we controlled for the gender composition because there is evidence that making this information salient changes people's behavior (Iriberri and Rey-Biel 2013).[8] Instead, we had each group of four sharing the same row in the computer laboratory and told participants that their competitors were seated in the same row as them.

For their participation, subjects receive a show-up fee of €5 plus €4 for completing tasks 1–4. The group of advisees is paid an additional €2 because they have a waiting period of approximately 15 minutes at the beginning of the experiment. This waiting period was necessary to ensure that advisors and advisees reach the advice stage at roughly the same time. During this 15 minutes waiting period, advisees are not yet informed about the content of the experiment because we wanted to ensure that the waiting period has no effect on the choice of the compensation scheme in task 3. We choose one of tasks 1–4 at random and pay participants according to their performance in that task. Finally, we pay subjects for the self-evaluation task and the advisors for giving advice. On average, our participants earned €18.28. The average duration of a session was 1 hour 30 minutes, starting with reading aloud the general instructions and finishing after participants filled out a questionnaire. Table 2.1 provides a timeline and briefly summarizes each task and its compensation scheme.

The experiment was conducted between December 2011 and January 2012 at the Universitat Autònoma de Barcelona (UAB). Subjects were recruited from a pool of subjects via the online recruitment system ORSEE (Greiner 2004) and were mainly undergraduate students from UAB. Students were invited to subscribe to ORSEE via flyers distributed and posted in different departments at UAB. We tried to have an equal distribution of subjects from different departments. UAB has a total of 40,000 students, and our subject pool contains

[6] We decided not to ask advisees for a tentative entry decision, as is common in the organizational psychology literature, because we were concerned of a possible confirmation bias. Without a tentative entry decision we cannot measure advice utilization in a direct, within-subject, fashion, so instead, we measure it in a between-subject comparison, contrasting entry decisions in our control group with entry decisions of those who receive advice.

[7] It has been shown that the gender composition of (potential) competitors can affect the willingness to participate in a competition (see, inter alia, Gneezy et al. 2003, Booth and Nolen 2012, Gupta et al. 2013). We use an equal gender composition because it is the most "neutral" setup and it is the one used in Niederle and Vesterlund (2007).

[8] In the psychology literature, this effect was coined "stereotype threat" (Steele 1997). The idea is that if a task is stereotypically being thought of as one in which one gender is better than the other (though in fact this might not be the case), then somebody from the "weak gender" might underperform simply because he or she is aware of this.

approximately 1,500 students.[9] There is a very low likelihood that participants of the same group knew each other because all 1,500 students of the subject pool received the invitations for the sessions at the same time, and we assigned participants randomly to generation and group. A total of 224 subjects, 112 men and 112 women, participated in the experiment. Of the 224 subjects, 112 were assigned the role of an advisor (generation 1) and 112 the role of an advisee (generation 2). The experiment was programmed and conducted with the experimental software z-Tree (Fischbacher 2007).

3.4. Research Questions and Hypotheses

With our experiment we would like to address the following questions.

Does advice lead to efficiency gains and does advice reduce the gender gap in competition entry? This is the central question we would like to answer. A reduction in the gender gap is desirable only if the reduction is achieved in an efficiency-enhancing way. It would, for example, not be desirable if weak-performing women (strong-performing men) entered the competition more (less) often. We think of efficiency mainly in terms of a good self-selection process in the sense that strong performers enter frequently, whereas weak performers tend to stay out, thus maximizing expected earnings (efficiency gains in economic terms).

Advisors are expected to give "good" advice (maximizing the advisee's expected earnings), and advisees are expected to trust their advisor's entry recommendation and select better into competition. Consequently, we expect forgone earnings to be lower with advice. If advice leads to efficiency gains in economic terms, we should see a reduction of the gender gap in tournament entry with advice.

Does advice affect the entry decision of high-performing women and low-performing men in a correcting way? We will analyze the effect of advice in more detail. We will examine whether the advisees' entry decisions differ systematically from the advisors' entry decisions, and have a closer look at whose entry decisions are affected (men's, women's, those of the strong performers, those of the weak performers, etc.). We would also like to know how men and women react to the specific advice they receive ("piece rate" or "tournament"). Splitting participants into three performance groups (weak, intermediate, and strong), we examine whether and how the gender gap is affected in each of the three subgroups.

Since we expect that advisees improve their entry decision compared to the advisors, the two main errors in tournament entry (strong-performing women entering too little and weak-performing men entering too often) are supposed to become less predominant. Consequently, the gender gap in the weak and in the strong performance groups is expected to shrink substantially or disappear with advice. In the intermediate performance group with advice, we expect no differences in the entry decision between women and men.

How does advice affect confidence levels? All aspects that are considered in an individual's entry decision (confidence in relative performance, preferences for competition, or risk attitudes) are natural candidates for the channels of advice. We have not designed our experiment to pinpoint through which channels advice affects decisions (and neither is our purpose to make a fine distinction of which part of the change in behavior is due to advice as

[9] One might be concerned about the representativeness of our subject pool (undergraduate university students) with respect to the general population and thus about the external validity of our results. We would like to emphasize that our results and conclusions hold for our subject pool, but we do not claim that the same is true for a different population.

such, and which part is an effect of indirect information transmission). Nonetheless, we are able to look at (changes in) confidence levels and their consistency with the entry decisions.

4. Experimental Results

This section presents the results of our experimental study. In §4.1, we report a few preliminary results, comparing performances across gender and generations, stating the entry rates, and how advice is given. Section 4.2 presents our main results, demonstrating how advice leads to efficiency gains in the entry decisions and that, nonetheless, the overall gender gap persists. In §4.3, we examine how advice affects women's and men's entry decisions in three performance groups (low, intermediate, and high) and what happens to the gender gap in entry in each of the three performance groups. We also show that the advice "tournament" increases the confidence level and entry rates significantly, which is particularly important for high-performing women.

Throughout this paper, whenever we mention performance, we mean the number of problems solved, and when we say a subject "solved" a problem, we take it to mean that the subject solved the problem correctly. If not noted otherwise, we use all 112 observations for generations 1 and 2, respectively.

To test for differences in the performance between (within) subjects, we use Mann–Whitney U (Wilcoxon rank sum) tests.[10] Tests show that the performance variables are overall not normally distributed. Since low (high) performers tend to receive the correct advice "piece rate" ("tournament"), we expect that advice improves selection into competition and use one-sided tests to test the effect of advice on tournament entry and two-sided tests otherwise.[11]

4.1. Preliminary Results and Advice Giving

4.1.1. Performance Distributions and Entry Rates

To get a sense for the performance in the addition task, Figure 2.1 shows the probability density distributions of men's and women's performance in task 1 and task 2, respectively, using all 224 observations. On average, in task 1, women solve 6.64 problems and men solve 7.40, and in task 2, women solve 8.59 problems and men solve 9.42. These performance differences between women and men are not statistically significant. For the sake of brevity, we will omit the details of the tests for our preliminary results, but the interested reader can find those and more descriptive statistics in the online appendix. We also compare performances in tasks 1 and 2 in various other ways: across generations, between women and men for each generation separately, and across generations for each gender separately. We find statistical differences in only one instance: Participants in generation 1 perform significantly better than participants in generation 2 in task 1;[12] the difference is not significant for women and men separately. However, differences in task 1 performance do not affect our results, because the more relevant performance level for the tournament entry is task 2 performance, and because we will condition our further analysis on task 2 performance anyway.

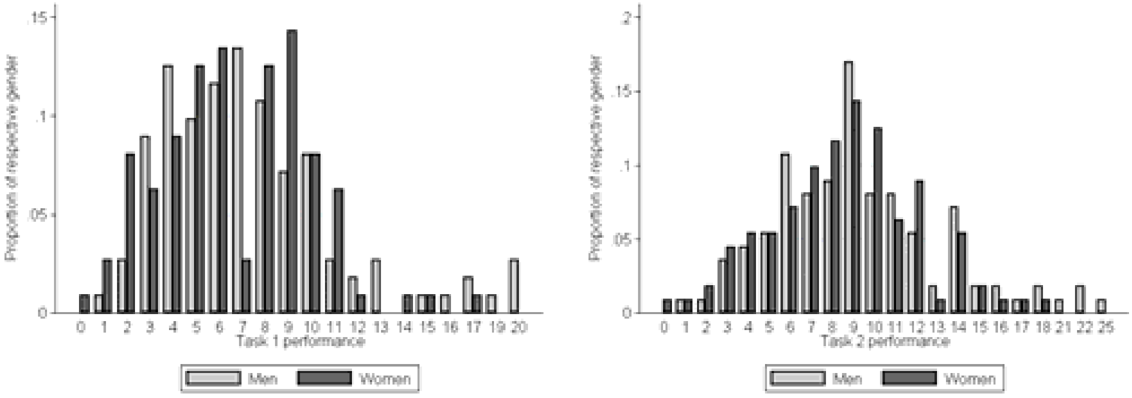
[10] For an easier reading, we will not explicitly describe the Mann–Whitney U test results with "differences in the distribution of the tested variable."

[11] For some subsamples, the number of observations is reduces considerably. Also, variables have partly high variance due to random uncontrollable factors and most analyzed decisions are binomial in nature. We use therefore a significance level of 0.1.

[12] The differences constitute less than one-third of the standard deviations, an indication that this difference is a random event.

Performances in tasks 1 and 2 are highly correlated, with subjects performing significantly better in task 2 than in task 1. This is also the case if we test separately for men and women. It is not clear to what extent this improvement can be attributed to a learning effect or to the change in incentives (moving from the piece rate to a tournament scheme), and we will not attempt to separate the two potential causes.

Figure 2.1: Probability Density Distributions of Number of Correctly Solved Problems by Gender



In generation 1, 58.9% of men (33 of 56) enter the tournament in task 3, whereas only 30.4% of women (17 of 56) do so. In generation 2, 58.9% of men (33 of 56) and 37.5% of women (21 of 56) choose the tournament. We will get back to the comparison of men's and women's entry decisions in §4.2.2.

4.1.2. Advice Giving

The advice given is, overall, in line with the performance of the advisees.[13] Remember that we incentivized advisors to give "good" advice without telling the advisees about the monetary incentives for the advisors (see experimental design).

We need two elements to be able to classify whether an advice given is "good" in economic terms: the number of problems a participant is expected to solve in task 3 and the corresponding probability of winning the tournament in task 3. For the number of problems a participant is expected to solve in task 3, we use the participant's task 2 performance. Table 2.2 summarizes the results of the corresponding probability calculation. Since men and women perform quite similarly, these differences are small. Since the tournament rate (€2) is four times the piece rate (€0.5), expected earnings are the same in the tournament and under the piece rate if the probability of winning is 25%. This is more or less the case if a participant solves 10 problems in task 2; see Table 2.2. We will come back to Table 2.2 later in the results section. The interested reader can find a detailed description of the calculation of the probability of winning in the online appendix.

From the advisors in generation 1, 88.4% (61 of 69) of those whose advisee's number of correct answers is at most nine in task 2 recommend correctly to choose the piece rate in

[13] A few weak-performing (strong-performing) advisees informed the advisor about having performed better (worse) than they actually did. From the nine participants in generation 2 who did not send their actual task 2 performance levels, only one participant with 11 correct answers received advice that did not fit his actual task 2 performance. The participant did, however, enter the competition, which is why we believe that the incorrect information provision by some advisees did not affect the quality of advice.

task 3, whereas 67.6% (23 of 34) of those whose advisee gave at least 11 correct answers correctly recommend tournament entry. The task 2 performance is significantly larger if advisors recommend entering the tournament (p = 0.000, two-sided Mann–Whitney U test).[14]

Table 2.2: Probability of Winning Given a Certain Performance Level in Task 2

Task 2 performance	<5	5	6	7	8	9	**10**	11	12	13	>13
Men (in %)	<0.1	0.3	1	2.8	6.6	15.4	**29.5**	43	56.7	66.3	>74.9
Women (in %)	<0.1	0.2	0.9	2.7	6.1	14.6	**27.7**	40	52.5	60.7	>69.9
Performance group (% of participants)	Weak (26%)			Intermediate (52%)					Strong (22%)		
Most likely (= optimal guessed) rank	4 (worst)			3 or 2*					1 (best)		

Note. With 10 correct answers (in bold), expected earnings for the tournament are similar to those for the piece rate.
* In the intermediate performance group, men with 11 correct answers are ranked second with a probability of 42.0% compared to first with 43.0%. For all other men and women in the intermediate group, the probability of being ranked second or third is highest.

As expected, the information about the performance in the advisor's own group plays an important role. If the advisee's reported task 2 performance is smaller than the number of correct answers of the best-performing subject in the advisor's own group, only 19.5% of the advisors (17 of 87) suggest entering the competition. If the advisee's reported task 2 performance is better than the highest performance level in the advisor's own group, 86.7% of the advisors (13 of 15) recommend choosing the tournament.

4.2. The Effect of Advice on Selection into Competition

4.2.1. Efficiency Gains in the Entry Decision

Our general focus is on whether advice improves the efficiency of people's assignment to jobs. In measuring efficiency gains, one can focus only on what happens for high-ranking jobs, or one can take a more global view and consider changes in both high- and low-ranking jobs. We decided in favor of the more global view and measure efficiency gains in terms of gains in expected earnings due to the changes in choosing the payment scheme.

We could measure gains in earnings only among those participants who win the tournament. In broader terms, this would correspond to looking at earnings of successful applicants for high-ranking jobs who are effectively hired. A second possibility would be to consider gains in expected earnings for all those who enter the tournament, corresponding to all applicants—successful or not—to high-ranking jobs. A third possibility is to measure the gains of expected earnings for all experimental participants, i.e., for the entire labor force, regardless of whether or not they apply for high-ranking jobs. If one is not only concerned about the quality of those who obtain a high-ranking job or who apply for one (i.e., who are available for a high-ranking job), but also wants those who have little or no chance to obtain a high-ranking job to assess their chances correctly, not waste resources on an application for the job, and accept a low-ranking job instead, the entire labor force is the right reference group to look at. Efficient decisions in this group are reflected by "correct" self-selection of all participants in our experiment: weak performers refrain from entering the competition, whereas strong performers do enter. We refer the interested reader to the online appendix for an extended discussion of the efficiency gain analysis.

[14] The improvement in performance from task 1 to task 2 is perceived positively: Advisors who recommend to enter the tournament observe a significantly larger change in advisee performance from task 1 to task 2 than advisors who suggest to select the piece rate (p = 0.030, two-sided Mann–Whitney U test).

To assess the efficiency of the self-selection process among the entire potential labor force, we consider two measures. First, we show that, with advice, the spread between the performance of those who enter the competition and those who stay away widens, indicating that participants take their own performance better into account. Second, we assess whether a participant's entry decision maximizes her expected earnings, and if not, calculate the forgone earnings. We then show that under advice these forgone earnings decrease significantly.[15]

Table 2.3: Performance by Choice of Compensation Scheme in Task 3

	Generation 1 (*without advice*)				Generation 2 (*with advice*)			
Choice of compensation	Piece rate		Tournament		Piece rate		Tournament	
Men								
Task 1 (piece rate)	7.61	(3.4)	8.09	(4.5)	**5.09**	**(2.2)**	**8.18**	**(4.3)*****
Task 2 (tournament)	8.57	(4.0)	10.15	(4.4)	**6.57**	**(2.6)**	**11.27**	**(4.4)*****
Task 3 (choice)	8.78	(3.8)	10.73	(5.9)	**7.00**	**(2.4)**	**11.73**	**(5.2)*****
Women								
Task 1 (piece rate)	6.69	(3.2)	7.71	(2.9)	5.83	(3.1)	7.05	(3.7)
Task 2 (tournament)	8.31	(3.1)	9.41	(3.3)	**7.74**	**(3.1)**	**9.86**	**(4.4)***
Task 3 (choice)	8.97	(2.8)	10.12	(3.4)	**7.94**	**(2.8)**	**10.14**	**(4.2)****

Notes. The average number of solved problems is shown for each subgroup (standard deviation in parenthesis). Bold piece rate-tournament value pairs are statistically significantly different.
* p<0.1; ** p<0.05; *** p<0.01 (two-sided Mann-Whitney U test).

To show that the spread in performance widens between those who enter the competition and those who do not, we use three different measures of performance: performance in tasks 1, 2, and 3. The corresponding average performance values are presented in Table 2.3 for women and men in generations 1 and 2. For each of the three subgroups, the average number of solved problems is calculated separately for those who choose the piece rate and those who choose the tournament in task 3. As expected, the average performance of those who enter the tournament is better in all subgroups for each of the three performance measure. In generation 1, our control group that does not receive advice, women who enter the tournament do not perform significantly better than women who choose the piece rate (p > 00219 for each of the three performance measures, two-sided Mann–Whitney U test). That means that the strong- and weak-performing women do not separate well into those who enter the competition and those who stay away. For men in generation 1, none of the three performance measures of those who enter the tournament is significantly better than the performance measures of those who choose the piece rate even though the gap in performance is larger than among women in generation 1 (p > 00152 for each of the three performance measures). In generation 2 (with advice), women and men make their decisions more in line with their performance: In tasks 2 and 3, women who enter the tournament now perform significantly better than women who choose the piece rate (p < 0.051 for both measures). For task 1 performance, the gap is not statistically significantly different (p = 0.198). For men, the gaps clearly widen and become statistically significant, no

[15] Ideally, we would like to assess efficiency also in terms of actual task 3 earnings. However, actual earnings do not necessarily reflect the "correctness" of the entry decisions. Since we have only few observations for most performance levels and the spread of payoffs is particularly large among those who (correctly) enter the tournament (€2 in case of winning versus zero payoffs in case of at least one better performer in the same group), the random group composition element is not averaged out.

matter which performance measure we consider (p < 0.014 for each of the three performance measures).

Table 2.4 presents the results of logit regressions of the tournament entry decision in task 3—models (1) through (4)—and of an ordered logit regression of the estimated rank in task 2—model (5). In regression models (1) and (2), separate regressions are presented for generations 1 and 2, respectively. The tournament entry decision in task 3 is regressed on a gender dummy (which takes the value 1 for a female participant and 0 for a male participant), task 2 performance, and the performance change from task 1 to task 2. Remember that the number of observations is the same in models (1) and (2). In model (1) without advice, the coefficient estimate of the performance in task 2 is not significant. In model (2) with advice, the performance in task 2 becomes a highly significant predictor for entry into competition (larger coefficient estimate and smaller p-value). This indicates that advice improves selection into competition in that better-performing participants are more likely to enter the tournament than worse-performing participants.

Table 2.4: Logit of Tournament Entry Decision in Task 3 (Models (1)-(4)) and Ordered Logit of Guessed Task 2 Rank (Model (5)) for Generations 1 and 2

	Tournament entry (task 3)				Guessed Rank (task 2)
VARIABLES	(1)	(2)	(3)	(4)	(5)
Female	-1.153***	-0.906**	-1.042***	-1.091***	1.234***
	(0.405)	(0.433)	(0.296)	(0.303)	(0.289)
Task 2 performance	0.092	0.252***	0.0855	0.112**	-0.353***
	(0.063)	(0.0765)	(0.0601)	(0.0509)	(0.0541)
Task 2- Task 1 performance	0.0268	0.092	0.0473	0.0628	-0.280***
	(0.0788)	(0.1116)	(0.0641)	(0.0655)	(0.0673)
Generation 2			-1.476*		
			(0.853)		
(Gen 2)(Task 2 performance)*			0.183**		
			(0.0898)		
Advice "tournament"				1.367***	-1.120**
				(0.494)	(0.465)
Advice "piece rate"				-0.383	0.332
				(0.340)	(0.310)
Observations	112	112	224	224	224
Log-likelihood	-70.53	-63.48	-134.1	-130.4	-201.0

Notes. Coefficient estimates are shown (standard errors in parentheses). The dependent variable *Tournament entry (task 3)* takes the value 1 for tournament and 0 for piece rate. The dependent variable *Guessed rank (task 2)* takes values between 1 (best) and 4 (worst). The variable *Advice "tournament"* (*Advice "piece rate"*) takes the value 1 if the advice received is "tournament" ("piece rate") and 0 otherwise. The constants are not reported. The sample is generation 1 for regression (1), generation 2 for (2), and generations 1 and 2 for regressions (3)–(5).

*p <0.1; **p <0.05; ***p <0.01

The sample for regression models (3)–(5) is generation 1 and generation 2. In model (3), the entry decision is regressed on a dummy variable for gender, the number of correct answers in task 2, the difference in performance between task 2 and task 1, a dummy variable for generation (which takes the value 1 for generation 2 and 0 for generation 1), and an interaction term between generation and task 2 performance. We will discuss regression models (4) and (5) later. At this point note only that in model (3), the positive coefficient estimate for the interaction term of generation and task 2 performance is significant, confirming that participants in generation 2 consider more their own performance than participants in generation 1 when making their entry decision.[16] It may also be noted that, as expected, the coefficient estimate for the dummy variable for gender is negative and highly significant, i.e., women are less likely to choose the competition than men, keeping the other variables constant.

We now examine whether entry decisions maximize expected earnings in task 3. We define the forgone earnings as the difference between expected earnings under the payment scheme the participant did not choose and those under the one she chose if a participant chose the (for her) inferior payment scheme. Forgone earnings are comparable to the difference between the opportunity cost and the actual earnings of a decision and are therefore a measure of efficiency gains in economic terms. We expect that these forgone earnings will be significantly reduced if participants receive advice and will show that this is indeed the case by comparing forgone earnings in generations 1 and 2.

To calculate forgone earnings under each payment scheme, we need the same two ingredients as in the analysis of the correctness of advice: the number of problems a participant is expected to solve in task 3 and the corresponding probability of winning the tournament in task 3. For the number of problems a participant is expected to solve in task 3, we use the participant's task 2 performance. Table 2.2 summarizes the results of the probability calculation. Remember that expected earnings are the same in the tournament and with the piece rate if the probability of winning is 25%. If a subject solves 11 or more problems in task 2 and does not enter the tournament, we count this as underentry. The forgone earnings from underentry are the difference between expected tournament earnings and the expected piece rate earnings. An example for underentry on the job market would be a highly qualified person who does not apply for a high-ranking job and thereby loses potential high earnings. If a participant with nine or fewer correct answers in task 2 enters the tournament, we count this as overentry. The forgone earnings from overentry are the difference between the expected piece rate earnings and the expected tournament earnings. On the job market, a weakly qualified person who has low probabilities of getting a high-ranking job but nevertheless applies, rejecting sure earnings from a low-ranking job, would be an example for overentry.

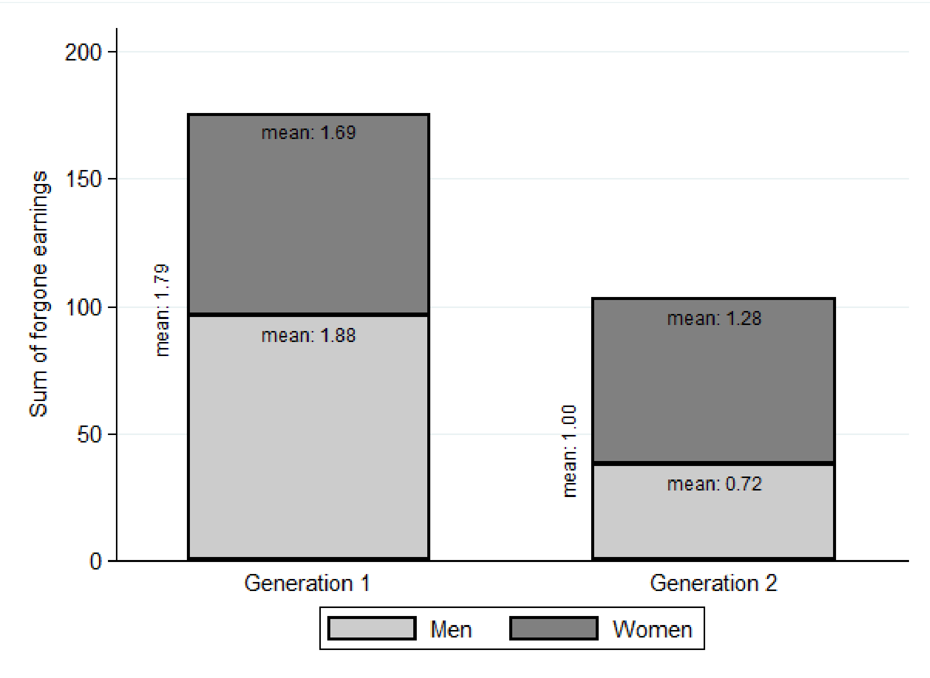
Under- and overentry rates become, as expected, smaller with advice. In generation 1, 23.5% (4 of 17) of men and 57.1% (8 of 14) of women "underenter," i.e., do not enter though they should, whereas the rates are only 10.5% (2 of 19) of men and 33.3% (5 of 15) of women in generation 2. The reduction in the underentry rate with advice is significant for men and women together ($p = 0.091$, one-sided Fisher's exact test). Among those who do enter the tournament though they should not, there is only a reduction among men: 50.0% (17 of 34) of men and 24.2% (8 of 33) of women in generation 1 "overenter," whereas 36.4% (12 of 33) of

[16] We include the interaction term (Generation 2) · (Task 2 performance), because task 2 performance seems to be more important for the entry decision in generation 2 than the performance change; see model (2) of Table 2.4. We ran a logit regression including the interaction term of generation and performance change in addition to the variables in model (3) of Table 2.4. The coefficient estimate of the interaction term (Generation 2) · (Task 2-Task 1 performance) was insignificant ($p = 0.622$), and the coefficient estimate of the interaction term (Generation 2) · (Task 2 performance) was slightly insignificant ($p = 0.103$). Spearman's rank correlation tests showed that task 2 performance and the performance change are highly correlated, and we think that this is the reason for the slightly insignificant interaction term between generation 2 and task 2 performance in the regression with both interaction terms. We also ran the logit regression (3) of Table 2.4 with the interaction term (Generation 2) · (Task 2-Task 1 performance) instead of the interaction term (Generation 2) · (Task 2 performance) and got an insignificant coefficient estimate for the interaction term ($p = 0.183$), underlining that task 2 performance is somewhat more important for the entry decision in generation 2 than the performance improvement.

men and 27.8% (10 of 36) of women in generation 2 do so. The reduction of the overentry rate in generation 2 is not significant (p = 0.314, one-sided Fisher's exact test).

The forgone earnings due to under- and overentry are summarized in Figure 2.2. The bars represent the sum of forgone earnings for generation 1 (175.2) and generation 2 (102.7) and show the forgone earnings of men (96.0) and women (79.2) in generation 1 as well as of men (37.7) and women (65.1) in generation 2. The average forgone earnings are indicated for each subgroup and are calculated over all those who potentially could make a "wrong decision" in each subgroup, respectively. Note that the level of forgone earnings is not interesting by itself since it depends on how much is paid for the experiment. We are only interested in the relative comparison across the generations. Average forgone earnings are 1.79 (1.88 for men, 1.69 for women) in generation 1, and 1.00 (0.72 for men, 1.28 for women) in generation 2. Pooling men and women, advice reduces forgone earnings significantly (p = 0.055, one-sided Mann–Whitney U test). Breaking down the analysis by gender shows that the larger part of the reduction in forgone earnings is due to the improvement of the entry decisions of men. For them, forgone earnings are significantly lower if they receive advice (p = 0.087, one-sided Mann–Whitney U test). For women, forgone earnings become also smaller if they receive advice, but the reduction is not significant (p = 0.220, one-sided Mann–Whitney U test).

Figure 2.2: Forgone Earnings of Tournament Over- and Underentry in Task 3 for Generations 1 and 2 (Based on Task 2 Performance)



In summary, we can provide a positive answer to our expectation that advice leads to efficiency gains in economic terms. Advice improves the self-selection process as evidenced through a widened performance spread between those who enter and those who do not, and lower forgone earnings due to wrong entry decisions. It may be the case that it is optimal for an individual with a strong performance to choose the piece rate, because she is very risk averse or does not like competitions. In that sense, we cannot assess the optimality of an individual's entry decision. However, using the answers to a questionnaire at the end of the experiment, where participants state how much they like to compete on a scale from 1 to 7, we find that, if anything, advice improves the extent to which participants take into account their own preferences. In generation 1, the average answers to the question are 4.5 and 5.2 for those who choose, respectively, the piece rate and the tournament. In generation 2, the average answers are 4.2 and 5.5; that is, the gap widens. Moreover, the general insights from the advice literature we discuss at the end of the literature section lead us to conjecture that in practice efficiency gains could be increased by the right design choice.

4.2.2. The Gender Gap With and Without Advice

After having seen that advice improves the efficiency in tournament selection, we would expect that the commonly found gender gap in tournament entry is reduced with advice. First we have a look at the gender gap in generation 1 (without advice). In generation 1, 58.9% of men (33 of 56) enter the competition in task 3, whereas only 30.4% of women (17 of 56) do so. This difference is statistically significant (p = 0.004, two-sided Fisher's exact test). In generation 2 (with advice), 58.9% of men (33 of 56) and 37.5% of women (21 of 56) choose the tournament, still a significant difference (p = 0.037, two-sided Fisher's exact test). The gender gap, defined as the percentage point gender difference in entry rates, is 28.5 percentage points in generation 1, whereas it is still 21.4 percentage points in generation 2. The gender gap is also confirmed in the regression models (1) and (2) in Table 2.4, where the coefficient estimates for the female dummy are negative and highly significant. The overall gender gap is not changed considerably through advice, a surprising result after having confirmed efficiency gains in tournament entry through advice. We will first look at how the advice received changes participants' entry behavior and then analyze in more detail whose decision is changed with advice depending on the gender and the performance level.

4.2.3. Effects of the Two Types of Advice

First of all, let us note that the advice that men and women receive does not differ (p = 0.840, two-sided Fisher's exact test): 37 men and 39 women (of 56, respectively) receive the advice to choose the piece rate; the remaining 19 men and 17 women are advised to enter the competition. For our analysis of how participants react to the advice, we compare entry rates of those who receive a certain piece of advice, for instance "piece rate," with entry rates of those who do not receive any advice (i.e., generation 1). A difficulty with this comparison is that the performance of those who receive the advice "piece rate" ("tournament") is, on average, worse (better) than the performance in generation 1. To account for these performance differences, we compare entry rates of those in generation 2 who received a certain piece of advice (e.g., "piece rate") with the expected entry rate of a reference group in generation 1 whose performance distribution is the same as in the subgroup of generation 2 who received a certain piece of advice (e.g., "piece rate").[17]

Looking at the data in Table 2.5, we see that participants react in the expected way to the advice they receive, i.e., they enter, on average, less if they receive the advice "piece rate" (a reduction from 38.5% to 33.3%), and they enter, on average, more if they receive the advice "tournament" (an increase from 49.4% to 79.4%). The reaction to the advice "tournament" is considerably stronger than the reaction to the advice "piece rate" (an increase of 59.3% versus a decrease of 13.4% with respect to the greatest possible change,

[17] For illustration, let us stick with the advice "piece rate." We calculate for each task 2 performance level that we observe among those who received the advice "piece rate" in generation 2, the corresponding entry rate in generation 1. If there is no observation in generation 1, we eliminate the corresponding observation in generation 2 as well. By doing so, we lose at most two observations in each subgroup. We then calculate what would be the expected entry rate in a subgroup of generation 1 that has the same size and exact same performance distribution as our subgroup of generation 2.

respectively). This is consistent with findings in the literature that show that individuals react more to positive feedback than negative feedback (see, e.g., Möbius et al. 2013).

The results from the logit regression in model (4) of Table 2.4 confirm this finding as well. There, we replace the generation 2 dummy and the interaction term by two variables: The dummy variable Advice "tournament" takes the value 1 if a participant receives the advice "tournament" and 0 otherwise. The dummy variable Advice "piece rate" takes the value 1 if a participant receives the advice "piece rate" and 0 otherwise. All 224 observations from generations 1 and 2 are included in the regression. The positive and significant coefficient estimate of the variable Advice "tournament" shows that a participant is more likely to enter the tournament if she receives the advice "tournament" compared to receiving no advice (and controlling for performance). For the variable Advice "piece rate" we see that the coefficient is negative and not significant. If anything, a participant is less likely to select into competition if she gets the advice "piece rate" compared to receiving no advice. The absolute value of the coefficient estimate is smaller (and the corresponding p-value is larger) for Advice "piece rate" than for Advice "tournament," which means that participants react more to the advice "tournament" than to the advice "piece rate."

Table 2.5: (Expected) Entry Rates for Generations 1 and 2 (in %)

	Gen 1*	Gen 2	% Change wrt max**
Pooled (men and women)			
Advice "piece rate"	38.5	33.3	-13.4
Advice "tournament"	49.4	79.4	59.3
Men			
Advice "piece rate"	44.9	41.7	-7.3
Advice "tournament"	53.3	93.3	85.7
Women			
Advice "piece rate"	25.8	25.0	-3.1
Advice "tournament"	45.0	62.5	31.8

[a] Expected entry rate in a subgroup of generation 1 with the same task 2 performance distribution as in the corresponding group of generation 2 that received the advice "piece rate" ("tournament").
[b] "% Change wrt max" refers to the change in the (expected) entry rate between generations 1 and 2 with respect to the greatest possible correcting change.

The ordered logit regression model (5) in Table 2.4 provides evidence that advice affects confidence levels and that this is one reason why advice—in particular the advice "tournament"—has an impact on entry decisions.[18] As in Niederle and Vesterlund (2007) and Niederle and Vesterlund (2011), preference for competition, risk attitudes, and self-confidence are three important factors for the entry decision among our participants.[19] In

[18] The general features of confidence levels we find are consistent with findings in the previous literature. In particular, in both generations, participants overestimate their rank in task 2 ($p < 0.001$ for generations 1 and 2 separately, two-sided Wilcoxon signed-rank test) and also when testing separately for men and women ($p < 0.029$ for both genders and generations separately, two-sided Wilcoxon signed-rank test). Men are more overconfident about their task 2 performance than women in generations 1 and 2 ($p < 0.012$, two-sided Mann–Whitney U test).

[19] We add three possible explanations for the tournament entry decision to the logit regression model (3) of Table 2.4: (1) guessed rank for task 2 (where 1 stands for the best rank and 4 for the worst rank), which we use as an inverse measure for confidence; (2) the entry decision in task 4, which we use as a proxy for risk attitudes; and (3) a variable from the questionnaire at the end of the experiment, which asks for the preference for competition on a scale from 1 to 7. The differential impact of task 4 entry as an explanatory variable should be (mainly) driven by risk attitudes because we control for all other known variables that could potentially affect task 4 entry. All three added variables have significant coefficient estimates (at the 5% and 1% levels) and are

model (5) in Table 2.4, the task 2 guessed rank is regressed on a dummy for gender, task 2 performance, the performance change from task 1 to task 2, and the dummy variables Advice "tournament" and Advice "piece rate." All 224 observations from generations 1 and 2 are included in the regression. Participants become significantly more confident if they get the advice "tournament" compared to receiving no advice (and controlling for performance), and become (insignificantly) less confident if they get the advice "piece rate" compared to receiving no advice. The coefficient estimate (and the corresponding p-value) for the advice "tournament" is larger (smaller) than for the advice "piece rate," which means that participants' confidence levels change more with the advice "tournament" than with the advice "piece rate." This parallels our result of the larger change in entry rates through the advice "tournament."

Breaking down the analysis of the reaction to advice by gender, we see that men react more strongly to the advice they receive, taking into account the ceiling problem that the minimum and the maximum entry rates are 0% and 100%, respectively. The entry rate of men (women) who receive the advice "piece rate" decreases from the expected rate of 44.9% (25.8%) in generation 1 to 41.7% (25.0%) in generation 2, a decrease with respect to the greatest possible decrease of 7.3% (3.1%). If the advice received is "tournament," the (expected) entry rates are 53.3% (45.0%) for men (women) in generation 1 and 93.3% (62.5%) in generation 2. Here, the increases in the entry rate with respect to the greatest possible increase are 85.7% for men and 31.8% for women.

4.3. The Effect of Advice Depending on Performance

4.3.1. Whose Entry Decisions Are Affected?

To examine more closely how advice improves the entry decision and why the gender gap in tournament entry persists with advice, we look at the entry decisions of women and men conditional on their performance levels. We use again the task 2 (tournament) performance to create the performance groups because it is arguably more informative about the expected task 3 performance than task 1 performance. We split participants into three groups according to the performance quartiles: (1) "weak" performers solve 6 or less problems (26% of all 224 participants), (2) "intermediate" performers give between 7 and 11 correct answers (52%), and (3) "strong" performers solve 12 or more problems (22%). For an overview, see also Table 2.2.

Another rationale for this split is the most likely rank in the competition given a certain task 2 performance level.[20] Weak (strong) performers are most likely ranked fourth (first) without exception. For weak (strong) male and female performers, the probability of rank 4 (1) is at least 48.1% (52.5%). Remember that the fourth (first) rank stands for worst (best) performer in task 2. In the intermediate performance group, men with 11 correct answers are ranked second with a probability of 42.0%, compared to first with 43.0%. For all other men and women in the intermediate group, the probability of being ranked second or third is highest. The rank a participant is most likely to obtain given her performance is at the same time the optimal guessed rank in the self-evaluation of task 2 performance.

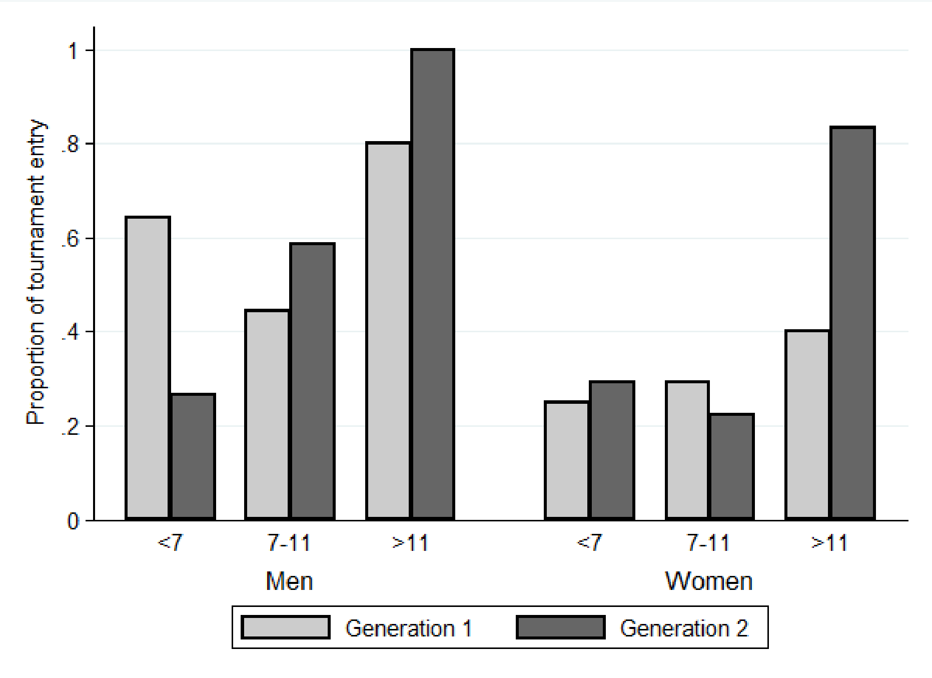
---

thus important determinants of the entry decision: the likelihood of entering the tournament goes up with smaller self-assessed rank, lower degrees of risk-aversion, and increased preferences for competition.

[20] Using the actual distribution of task 2 performances among our participants, we calculated the probability of obtaining each of the possible four ranks for each performance level, assuming that the participant is randomly matched in a group of two men and two women. For example, the chances of a female participant with 12 correct answers in task 2 obtaining ranks 1, 2, 3, and 4 are 52.5%, 37.9%, 8.9%, and 0.7%, respectively. Her most likely rank is thus rank 1.

Ideally, one would want advice to increase entry rates among strong performers and decrease them among weak performers (whereas for the intermediate performers, it is not as clear whether one decision is superior to the other). Figure 2.3 shows the proportions of men and women who choose the tournament for each of the three performance groups and for generations 1 and 2. Clearly, advice reduces the two common errors in tournament entry: The effect of advice is particularly strong for the group of high-ability (i.e., strong-performing) women. In this group, only 40% of women who do not receive advice enter the tournament (4 of 10), as opposed to 83% of women who receive advice (10 of 12); the increase in the entry rate is statistically significant ($p = 0.048$, one-sided Fisher's exact test). This is particularly important because the group of strong-performing participants is the one that should ideally enter the tournament and "be available for high-ranking jobs." Also the entry rate of strong-performing men if anything increases (from 80% to 100%, $p = 0.156$, one-sided Fisher's exact test). The other common error in tournament entry (weak-performing men entering the tournament too often) is also reduced significantly as expected: entry by weak-performing men is reduced from 64% (9 of 14) to 27% (4 of 15, $p = 0.048$, one-sided Fisher's exact test). There are no significant effects in the other subgroups. Note that there are different effects for men and for women in the intermediate group: men enter more and women enter less when they receive advice. We will discuss this in the next subsection.

Figure 2.3: Proportions of Men and Women Who Enter the Competition for a Given Range of Correct Answers in Task 2 by Generation



We find some indication that advice changes confidence levels particularly among strong performers. Recall that in the self-evaluation after tasks 1–4, we elicit beliefs about one's own rank in tasks 1 and 2, ranging from 1 (best) to 4 (worst). Note that, by design, rank 4 (weak-performing) and rank 1 (strong-performing) candidates can only over- and underestimate their performance, respectively. In Table 2.6, the self-assessment for task 2 is presented separately for men and women in generations 1 and 2 for each performance group. Remember that the optimal guessed ranks for weak, intermediate, and strong performers are

4, 3 or 2, and 1, respectively. We expect that strong-performing women become more confident with advice (the guessed rank decreases), and that weak-performing men become less confident with advice (the guessed rank increases). As expected, strong-performing women and also strong-performing men become significantly more confident ($p = 0.053$ for men and $p = 0.020$ for women, one-sided Mann–Whitney U test). Surprisingly, the confidence level of weak-performing men does not decrease ($p = 0.455$, one-sided Mann–Whitney U test); weak-performing women adjust their confidence level slightly downward ($p = 0.092$, one-sided Mann–Whitney U test). Among intermediate performers, the self-assessment of women does not change with advice ($p = 0.4892$, two-sided Mann–Whitney U test), whereas men become more confident with advice ($p = 0.098$, two-sided Mann–Whitney U test).

Table 2.6: Average Guessed Task 2 Ranks of Men and women in Generations 1 and 2

	Weak		*Intermediate*		*Strong*	
Performance group	Men	Women	Men	Women	Men	Women
Men	2.64 (1.2)	2.60 (1.0)	**1.89 (0.9)**	**1.52 (0.7)***	**1.20 (0.4)**	**1.00 (0.0)***
Women	**2.58 (0.9)**	**3.00 (0.7)***	2.15 (0.6)	2.22 (0.6)	**1.70 (0.5)**	**1.25 (0.5)****
Optimal guessed rank	4		3 or 2		1	

Notes. Standard deviations are in parentheses. Bold generation 1–generation 2 value pairs are statistically significantly different. Except for men with 11 correct answers, participants' optimal guessed task 2 rank in the intermediate group is second or third.

* $p < 0.1$; ** $p < 0.05$ (one-(two-)sided Mann–Whitney U test in the weak and strong (intermediate) performance groups).

4.3.2. Changes in the Composition of the Gender Gap

Across performance groups (see Figure 2.3), there are interesting differences regarding the gender gap and how advice affects it. Remember that the strong and the weak performers represent 22% and 26% of all observations, respectively, and the intermediate performers account for 52%. We first note that, in generation 1, men enter more often than women in all three groups, and the effect is significant among the strong and the weak performers ($p = 0.087$ and $p = 0.062$, two-sided Fisher's exact test), but not among the intermediate performers ($p = 0.287$, two-sided Fisher's exact test). The typical gender entry errors become very clear when participants do not receive advice, i.e., strong-performing women enter too seldom and weak-performing men enter too often. The overall gender gap in generation 1 is thus mostly driven by the differences in entry behavior among strong and weak performers. In contrast, the gender gap in generation 2 is now driven by the differences in entry behavior among intermediate performers, where 59% (17 of 29) men and 22% (6 of 27) women enter ($p = 0.007$, two-sided Fisher's exact test). The gender gap becomes insignificant among participants with high performance levels (100% of men and 83% of women enter, p D 00478, two-sided Fisher's exact test). Among weak-performing participants, women enter even slightly more than men if they receive advice ($p = 1.000$, two-sided Fisher's exact test). The effects of advice in the strong- and weak-performing groups are as expected; however, the emergence of a gender gap in the intermediate performance group is a surprise. To find an explanation for the gender differences in tournament entry in generation 2, we look at men's and women's reactions to the advice they receive.

4.3.3. Reactions to the Advice Received

First of all, we confirm that the two types of advice are equally spread between men and women in all three performance groups: Among weak performers in generation 2, 6.7% (1 of 15) of men and 5.9% (1 of 17) of women receive the recommendation to enter the competition. For intermediate performers, the corresponding rates are 27.6% (8 of 29 men) and 25.9% (7 of 27 women), and for strong performers they are 83.3% (10 of 12 men) and 75% (9 of 12 women). None of these differences are statistically significant.

Breaking down the analysis of the reaction to advice by gender, we see that the emerging gender gap in the intermediate group is due to differences between men and women in their reactions to the advice they receive. We find no gender difference in reaction among the weak or strong performers. Table 2.7 summarizes the entry rates and average guessed ranks of men and women separately for each performance group and depending on the advice received. Overall, the entry rates for men are larger than for women for both pieces of advice: With the advice "piece rate," 41% (15 of 37) of men and 26% (10 of 39) of women enter the tournament; with the advice "tournament," 95% (18 of 19) of men and 65% (11 of 17) of women do so. The difference is significant only for the advice "tournament" ($p = 0.037$, two-sided Fisher's exact test). Testing for gender differences in the reaction to advice within performance groups reveals that the gender gap in the reaction to advice depends on the performance level: There is no gender gap in the weak-performing group with the advice "piece rate" or the strong-performing group with the advice "tournament" ($p > 0.474$ for the two groups separately, two-sided Fisher's exact test), but there is one in the intermediate group ($p = 0.026$ for the advice "tournament" and $p = 0.085$ for the advice "piece rate," two-sided Fisher's exact test).[21]

Table 2.7: Entry Rates and Average Guessed Task 2 Ranks of Men and Women in Generation 2 Depending on the Advice Received

	Weak		*Intermediate*		*Strong*	
Performance group	Men	Women	Men	Women	Men	Women
Entry rate						
Advice "piece rate"	29%	31%	**43%**	**15%***	100%	67%
	(4 of 14)	(5 of 16)	**(9 of 21)**	**(3 of 20)**	(2 of 2)	(2 of 3)
Advice "tournament"	0%	0%	**100%**	**43%****	100%	89%
	(0 of 1)	(0 of 1)	**(8 of 8)**	**(3 of 7)**	(10 of 10)	(8 of 9)
Guessed task 2 ranks						
Advice "piece rate"	2.64 (1.0)	3.06 (0.7)	**1.71 (0.8)**	**2.25 (0.7)****	1.00 (0.0)	1.33 (0.6)
Advice "tournament"	2.00 (0.0)	2.00 (0.0)	**1.00 (0.0)**	**2.14 (0.4)*****	1.00 (0.0)	1.22 (0.4)

Note. Bold men–women value pairs are statistically significantly different.
* p < 0.1; * p < 0.05; *** p < 0.01 (two-sided Fisher's exact test for entry rates, two-sided Mann–Whitney U test for guessed ranks).

Interestingly, the finding that, among intermediate performers, women are more reluctant in following advice to enter the competition, whereas men are more reluctant in following advice to choose the piece rate, is also mirrored in the time participants need for their entry decisions upon receiving advice. The longer a participant needs to choose a compensation scheme after having received advice, the more likely the advice has produced a conflict between the recommendation and the individual's own idea of whether to enter the competition. Thus we interpret the time a participant needs as a proxy for his or her initial decision (which we do not elicit explicitly for reasons discussed earlier). Men (women) in generation 2 need, on average, 20.9 (21.0) seconds to reach a decision. Men who receive the advice "piece rate" need, on average, longer for their decision (23.4 s) than men who receive

[21] We also find gender differences in the reaction to the reasons provided to support the advice given. If the advice "tournament" is supported by emphasizing potentially higher earnings/encouraging to trust in one's own ability, men enter the competition more often than women ($p < 0.041$, two-sided Fisher's exact test), particularly in the intermediate group ($p < 0.055$, two-sided Fisher's exact test).

the advice "tournament" (16.1 s), and also longer than women who receive the advice "piece rate" (19.9 s). For women, we find the reverse: Upon receiving the advice "tournament," they need longer to make up their mind (23.6 s) than if they receive the advice "piece rate" (19.9 s), and also longer than men who receive the advice "tournament" (16.1 s). The difference in response times between men who receive the advice "piece rate" and men who receive the advice "tournament" is significant ($p = 0.043$, two-sided Mann–Whitney U test). The other time differences are not significant.[22] The time differences become larger for women if we restrict the analysis on the intermediate performers.

Comparing the self-assessment of men and women in the three performance groups for each type of advice with the corresponding entry decisions indicates that self-confidence is an important reason for the entry decision. Table 2.7 summarizes the average guessed rank for task 2 performance of men and women depending on the advice they receive. Similar to the tournament entry decision where men show higher entry rates, men are more confident than women in generation 2 independent of the advice they receive, i.e., their guessed rank is lower. With the advice "piece rate," men (women) guess a rank of 2.03 (2.51); receiving the advice "tournament," the guessed rank of men (women) is 1.05 (1.65) on average ($p < 0.023$ for both types of advice separately, two-sided Mann–Whitney U test). Analyzing the self-assessment by performance group, we find that men are significantly more confident than women in the intermediate group independent of the advice they receive ($p < 0.030$ for each type of advice, two-sided Mann–Whitney U test). Although there is a slight difference among strong-performing participants who are advised to enter the competition ($p = 0.125$, two-sided Mann–Whitney U test), men and women in the weak-performing group do not differ in their relative self-assessment ($p = 0.239$, two-sided Mann–Whitney U test), which mirrors the results of the effect of advice on entry decisions.

Summarizing the last three subsections, we confirm significant improvements in self-selection; in particular, strong-performing women enter significantly more and weak-performing men enter significantly less often. We have seen that the persistence of the gender gap in tournament entry is due to a gender gap in following the advice received among intermediate performers: whereas women more often follow the advice "piece rate," men more often follow the advice "tournament."

5. Conclusion

We propose advice as a "soft intervention" to improve the (self-)selection into competition and overcome the gender gap in entry into competition. Although we have shown that advice indeed improves the efficiency of the selection process (strong-performing individuals enter more, weak-performing individuals enter less), the gender gap in entry is, at least on the surface, unchanged. A closer analysis has shown that the gap goes away among weak and strong performers, but a gap emerges under intermediate performers. This emerging gap is due to different reactions of men and women to the advice they receive.

Overall, our results suggest that advice significantly increases efficiency in the entry behavior, although it may not be suitable to overcome the gender gap in tournament entry entirely. Advice (given equally to women and men or particularly to strong-performing women) might therefore be a soft alternative to affirmative action, such as quotas or other forms of positive discrimination. Advice has the potential of improving efficiency, rather than only "not hurting" it. The improvement is possible because, as we showed, advice tends to be good and therefore helps individuals adjust their perception of their relative performance, in

[22] Variances of response times are generally quite high.

addition to helping them think about the decision more carefully. Comparing more closely the changes in entry rates in our experiment to those found by others, we find that among strong-performing women, the increase in entry through advice (more than 100%) is roughly the same as through the affirmative action "quota" (Balafoutas and Sutter 2012), and more than through the provision of relative performance feedback (Wozniak et al. 2014). With advice, the entry rate of strong-performing men increases (insignificantly) from 80% to 100%, whereas with the quota and with relative performance feedback, high-ability men even reduce their entry rates slightly. Comparing the impact on the entry decisions of weak-performing men, we see a decrease in entry through advice (about 50%), which is similar to the decrease through the provision of relative performance feedback, but much better than the effect of the policy intervention "quota," which leads to a slight increase in entry. We can conclude that advice—especially the advice "tournament" for strong-performing participants—improves the self-evaluation.

One might argue that advice can be profitable not only for the advisee, but also for the advisor (as it is by design in our experiment). The profit can be monetary and nonmonetary: In a firm, the senior can profit from a high-ability employee in the future through building up a good relationship with a potential star. Academic advisors benefit from advising high-ability students to enter academia in form of good coauthors in future research projects. In the personal environment, more experienced relatives and friends might benefit from giving advice because they care about the wellbeing of younger family members and friends and feel good for giving appropriate advice. Building up a reputation can play a role in the professional and personal environment.

For the moment, we can only speculate whether it matters that the advice in our experiment is personalized. Having somebody (exclusively) by the side could trigger an improvement in confidence, and the motivational part of advice that goes beyond the informational content could be important. This remains to be shown. The results from the organizational psychology literature lead us to conjecture that a more personalized form of communication, for example, a free-form chat, would have an even higher potential for efficiency gains. Our design can easily be extended to allow for free-form communication, as, for example, in Brandts and Cooper (2007) and Brandts et al. (2012). Additionally, analyzing the roles of the intensity of the interaction between advisor and advisee, the repetition of advice, and the willingness of advisors to provide advice would be interesting questions. Other naturally arising questions in this context are how the results would be if only (high-ability) women received advice, whether individuals solicit advice, and, if so, from whom they ask advice if it is an option (maybe even making it costly to obtain it).

6. References

Athey S, Avery C, Zemsky P (2000) Mentoring and diversity. Amer. Econom. Rev. 90(4):765–786.

Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the laboratory. Science 335(6068):579–582.

Balafoutas L, Kerschbamer R, Sutter M (2012) Distributional preferences and competitive behavior. J. Econom. Behav. Organ. 83(1):125–135.

Bettinger E, Baker R (2011) The effects of student coaching in college: An evaluation of a randomized experiment in student mentoring. Working Paper 16881, National Bureau of Economic Research, Cambridge, MA.

Blau FD, Currie JM, Croson RTA, Ginther DK (2010) Can mentoring help female assistant professors? Interim results from a randomized trial. Amer. Econom. Assoc. 100(2):348–352.

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. Organ. Behav. Human Decision Processes 101(2): 127–151.

Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? J. Econom. Behav. Organ. 81(2):542–555.

Brandts J, Cooper D (2007) It's what you say, not what you pay: An experimental study of manager-employee relationships in overcoming coordination failure. J. Eur. Econom. Assoc. 5(6): 1223–1268.

Brandts J, Charness G, Ellman M (2012) Let's talk: How communication affects contract design. CESifo Working Paper 3883, Institut d'Anàlisi Econòmica (CSIC), Bellaterra, Spain.

Cason TN, Masters WA, Sheremeta RM (2010) Entry into winnertake- all and proportional-prize contests: An experimental study. J. Public Econom. 94(9–10):604–611.

Celen B, Kariv S, Schotter A (2010) An experimental test of advice and social learning. Management Sci. 56(10):1687–1701.

Charness G, Rustichini A, van de Ven J (2011) Self-confidence and strategic deterrence. Discussion Paper 11-151/1, Tinbergen Institute, Amsterdam.

Chaudhuri A, Schotter A, Sopher B (2009) Talking ourselves to efficiency: Coordination in inter-generational minimum effort games with private, almost common and common knowledge of advice. Econom. J. 119(534):91–122.

Croson R, Gneezy U (2009) Gender differences in preferences. J. Econom. Literature 47(2):1–27.

Dargnies MP (2012) Men too sometimes shy away from competition: The case of team competition. Management Sci. 58(11):1982–2000.

Dohmen T, Falk A (2011) Performance pay and multi-dimensional sorting: Productivity, preferences, and gender. Amer. Econom. Rev. 101(2):556–590.

Ertac S, Szentes B (2011) The effect of information on gender differences in competitiveness: Experimental evidence. Working Paper 1104, Koç University, Istanbul, Turkey.

Ewers M (2012) Information and competition entry. Discussion Paper 6411, Institute for the Study of Labor (IZA), Bonn, Germany.

Fischbacher U (2007) z-tree: Zurich toolbox for ready-made economic experiments. Experiment. Econom. 10(2):171–178.

Flory JA, Leibbrandt A, List JA (2010) Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions. Working Paper 16546, National Bureau of Economic Research, Cambridge, MA.

Gino F (2008) Do we listen to advice just because we paid for it? The impact of advice cost on its use. Organ. Behav. Human Decision 107(2):234–245.

Gneezy U, Leonard KL, List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. Econometrica 77(5):1637–1664.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. Quart. J. Econom. 118(3):1049–1074.

Greiner B (2004) The online recruitment system ORSEE 2.0—A guide for the organization of experiments in economics. Working Paper Series in Economics 10, University of Cologne, Cologne, Germany.

Grosse ND, Riener G (2010) Explaining gender differences in competitiveness: Gender-task stereotypes. Jena Economic Research Paper 2010-017, University of Jena, Jena, Germany.

Gupta ND, Poulsen A, Villeval MC (2013) Gender matching and competitiveness: Experimental evidence. Economic Inquiry 51(1): 816–835.

Iriberri N, Rey-Biel P (2013) On women's underperformance in competitive environments: When and why. Working paper, University of the Basque Country, Bilbao, Spain.

Iyengar R, Schotter A (2008) Learning under supervision: An experimental study. Experiment. Econom. 11(2):154–173.

Möbius MM, Niederle M, Niehaus P, Rosenblatt TS (2013) Managing self-confidence: Theory and experimental evidence. Working paper, Microsoft Research, Cambridge, MA.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? Quart. J. Econom. 122(3): 1067–1101.

Niederle M, Vesterlund L (2011) Gender and competition. Annual Rev. Econom. 3(1):601–630.

Niederle M, Segal C, Vesterlund L (2013) How costly is diversity? Affirmative action in light of gender differences in competitiveness. Management Sci. 59(1):1–16.

Reuben E, Sapienza P, Zingales L (2014) How sterotypes impair women's career in science. Proc. Natl. Acad. Sci. USA 111(12): 4403–4408.

Reuben E, Rey-Biel P, Sapienza P, Zingales L (2012) The emergence of male leadership in competitive environments. J. Econom. Behav. Organ. 83(1):111–117.

Rodrìguez-Planas N (2010) Mentoring, educational services, and economic incentives: Longer-term evidence on risky behaviors from a randomized trial. Discussion Paper 4968, Institute for the Study of Labor (IZA), Bonn, Germany.

Schotter A (2003) Decision making with naive advice. Amer. Econom. Rev. 93(2):196–201.
Schotter A, Sopher B (2003) Social learning and coordination conventions in inter-generational games: An experimental study. J. Political Econom. 111(3):498–529.

Schotter A, Sopher B (2007) Advice and behavior in intergenerational ultimatum games: An experimental approach. Games Econom. Behav. 58(2):365–393.

Shurchkov O (2012) Under pressure: Gender differences in output quality and quantity under competition and time constraints. Journal of the European Economic Associaction 10(5):1189–1213.

Steele CM (1997) A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist 52(6): 613–629.

Sutter M, Glätzle-Rützler D (2014) Gender differences in the willingness to compete emerge early in life and persist. Management Science, Forthcoming.

Wozniak D, Harbaugh WT, Mayr U (2014) The menstrual cycle and performance feedback alter gender differences in competitive choices. Journal of Labor Economics 32(1):161–198.

7. Appendices

Appendix A. Experimental Design

We used the following procedure to control for the gender composition of each group while minimizing the possibility that participants would take note. In our online recruitment system ORSEE (Greiner, 2004), we created two separate experiments, one for women and one for men, which for them looked like exactly the same experiment. By doing so we ensured that roughly the same number of women and men showed up to each session. Upon arrival to the session, each male participant was given an odd number and each female participant was given an even number (in a way that looked random to them). We told them we used the numbers mainly to do a lottery, determining who could participate and who could not since we needed multiples of eight to participate. We numbered the computer terminals with 1-4 for the first group in the first row, 5-8 in the second row, etc, and asked participants to sit where they saw their number.

Appendix B. Performance in Task 1 (Piece Rate) and Task 2 (Tournament)

The performances of women and men do differ, neither under the piece rate nor under the tournament payment scheme. We can pool the data across generations for the analysis of task 1 and 2 performances because there are no treatment differences for these two tasks. Pooling the data for generations 1 and 2, women solve on average 6.64 and 8.59 problems in tasks 1 and 2, respectively; the corresponding standard deviations of performances in task 1 and 2 are 3.22 and 3.48. The average number of problems solved by men is 7.40 in task 1 and 9.42 in task 2; the corresponding standard deviations of performance in tasks 1 and 2 are 3.99 and 4.33. The distribution of performance is not significantly different between women and men ($p > 0.358$ for either task, two-sided Mann-Whitney U test).[23] Looking at each generation

[23] The distribution of the performance change between task 1 and 2 and the distribution of task 3 performance are not significantly different between women and men either ($p > 0.558$ for either of the two performance measures, two-sided Mann- Whitney U test).

separately, there are no significant gender differences in task 1 and 2 performances ($p > 0.425$ for all four tests, two-sided Mann-Whitney U test).

Comparing performance in generations 1 and 2, we find that, overall, the performance is not very different. In generation 1, participants solve on average 7.45 and 9.07 problems in tasks 1 and 2; the corresponding standard deviations are 3.63 and 3.79 in tasks 1 and 2. In generation 2, participants solve on average 6.60 and 8.94 problems in tasks 1 and 2; the corresponding standard deviations are 3.60 and 4.10 in tasks 1 and 2. Participants in generation 1 outperform participants in generation 2 significantly in task 1 ($p = 0.036$, two-sided Mann-Whitney U test). An indication that this performance difference is a random (though somewhat unlikely) event is that it constitutes less than one third of the standard deviation. The distributions of task 2 performance do not differ significantly between generations 1 and 2 ($p = 0.618$, two-sided Mann-Whitney U test).

If we look at each gender separately and examine differences across generations, we find that the differences in task 1 performance are small (women: 7.00 correct answers in generation 1 and 6.29 in generation 2; men: 7.89 in generation 1 and 6.91 in generation 2) and slightly insignificant ($p > 0.117$ for women and men separately, two-sided Mann-Whitney U test). The task 2 performance does not differ significantly between generation for each gender separately ($p > 0.594$ for women and men separately, two-sided Mann-Whitney U test). The small differences that we find across generations are not a problem as we use task 2 performance mainly (where we find no generation differences in performance). Furthermore, the performance differences in task 1 between generations are mainly driven by participants in the intermediate performance group ($p = 0.034$, two-sided Mann-Whitney U test), but not by participants in the low ($p = 0.217$) or high ($p = 0.864$) performance group. Men and women in the intermediate performance group in generation 2 do not perform differently in task 1 ($p = 0.496$, Mann-Whitney U test). All together, we conclude that there are no relevant performance differences between women and men (and generations) so that, all else equal, one would expect them to enter the tournament at similar rates.

As in NV07, performance in task 1 and 2 is highly correlated (Spearman rank correlations of 0.741 for women and 0.649 for men) and subjects perform significantly better in task 2 than in task 1 ($p = 0.000$ for women and men separately, two-sided Wilcoxon signed-rank test). The same is true if we test for correlation and performance differences in generation 1 and 2, and for women and men in generation 1 and 2 separately (Spearman rank correlations between 0.574 and 0.800 for all six tests; $p < 0.001$ for all six tests, two-sided Wilcoxon signed-rank test). This improvement can be due to a learning effect but also be caused by the change in incentives when moving from the piece rate to a tournament scheme, see also NV07.

Appendix C. Calculation of the Probability of Winning

To calculate forgone expected earnings under each payment scheme, we need two ingredients: the number of problems a participant is expected to solve in task 3 and the corresponding probability of winning the tournament in task 3. A participant's expected tournament earnings are equal to the expected number of solved problems times the expected probability of winning the tournament times €2. A participant's expected piece rate earnings are equal to the expected number of problems solved times €0.5. For the number of problems a participant is expected to solve in task 3, we use the participant's task 2 performance.[24] Since one cannot know the ex-post performance in task 3 before deciding on the tournament entry (using the ex-post performance is a somewhat theoretical analysis), the ex-ante performance in task 2

[24] We could also use task 1, but evidently task 2 performance is a better predictor for task 3 performance in case of entry into competition.

should be the main indicator for the entry decision. We can interpret ex-ante performance as experienced performance under similar circumstances and ex-post performance as the actual performance when entering a competition. We determine the second ingredient, the probability of winning the tournament, as follows. Using the sample of all 224 participants, the probability calculation is done assuming that a participant with a given task 2 performance is randomly grouped with one participant of the same gender and two participants of the other gender. Thus the composition of each possible group is the same as in the experiment: two women and two men. We can then calculate the probability that this performance level is higher than the task 2 performances of three other randomly drawn participants. We use this approach not only for calculating the probability of winning, but also for the probability of obtaining each of the four possible ranks given a certain performance. Because differences in performance between women and men are small and insignificant, the probabilities of winning the tournament conditional on a certain performance level are similar for women and men. NV07 use a bootstrap method to calculate the probabilities of winning. We do not expect the two approaches to lead to relevant differences.

Because the tournament-rate (e2) is four times the piece rate (e0.5), expected earnings are higher in the tournament if the probability of winning given a specific performance level (i.e. a specific number of problems solved) is larger than 25%. A female participant who solves ten problems in task 2 wins the tournament with a probability of 27.7% and a male participant with ten correct answers wins it with 29.5%. For nine solved problems, the probabilities of winning decrease to 14.6% for a woman and 15.4% for a man, whereas they increase to 40.0% for a woman and 43.0% for a man if the participant gives eleven correct answers in task 2. For lower and higher performance levels in task 2, the probabilities decrease and increase, respectively: With eight or less correct answers, the probabilities of winning are less than 7% for both women and men; with twelve or more correct problems, the probabilities of winning are higher than 52% for women and men.

A participant should enter the tournament if she solves at least eleven problems in task 2. If a subject solves eleven or more problems in task 2 and does not enter the tournament, we count this as under-entry. The corresponding forgone expected earnings from under-entry are the difference between expected tournament earnings and the expected piece rate earnings. The probability of winning the competition is lower than 25% if a participant gives nine or less correct answers in task 2. If such a participant nonetheless enters the tournament, then we count this as over-entry. The forgone expected earnings from over-entry are the difference between the expected piece rate earnings and the expected tournament earnings in the tournament.

Appendix D. Efficiency Discussion

We suggest three different reference groups to evaluate efficiency changes. 1) The “winner pool” consists of all participants who enter the tournament and win it. On the job market, these are the applicants who successfully apply for a high-ranking job. If one cares only about the quality of successful candidates, as maybe employers would do, a natural measure of the efficiency of the entry decisions taken by our participants would be the performance distribution (as a measure of ability) within this winner pool. 2) The “applicant pool” consists of all participants who enter the tournament (who "apply for the high-ranking job"). If one is concerned about the quality of the applicants (for example because it is difficult to discern the good from the bad ones, or because one wants to minimize the chance of a complete mismatch), a good measure of efficiency would be the performance distribution within this applicant pool). 3) The “labor pool” consists of all our participants - in the job market analogy, we think of them as the labor force. If one is not only concerned about the quality of those who obtain a high-ranking job or who apply for one, but also wants those who have little or no chance to obtain the high-ranking job to assess their chances correctly, not waste resources on an application, and accept a low-ranking job, this is the right pool to look at. Efficient decisions in this group are reflected by "correct" self-selection: weak performers refrain from entering the competition, while strong performers do enter. Efficiency gains in each of the three reference groups are related. In particular, an improved self selection process in the labor force implies efficiency gains among applicants. And normally (that is without the artificial feature of competing against the task 2 performance), efficiency gains among applicants should lead to efficiency gains among the hired employees.

While the under-entry of strong performing women is a major concern, we would like to draw attention to the inefficient decisions of those who have little chance of winning, but still enter the tournament. In reality, this is not only a waste of resources for the applicant, but may also harm the potential employer who in turn has to provide more resources to select the best candidate among the applicants. Therefore, we will be mainly concerned with the efficiency of the entire labor pool. An additional reason why, in contrast to Balafoutas and Sutter (2012) and Niederle et al (2010), we do not examine efficiency in the winner pool is owed to our experimental design. Recall that we have adopted from NV07 the feature that a participant who enters the tournament in task 3 competes against the task 2 performance of her group. As a consequence, those who enter the tournament do not compete for a fixed number of "openings." To see this, note that it is possible that everybody who enters the tournament wins it (or that everybody loses it). Comparing the quality of winners can therefore be misleading. For instance, suppose that without advice only the top performer among all our participants enters (and wins) the tournament. If with advice many other strong performers enter the tournament (as should be desirable), the quality of the winners must decrease. Therefore we think that this measure does not reflect well efficiency in our context. Furthermore, the conditions, under which winners are determined, change with the introduction of affirmative action programs. With affirmative action, the performance of winners can be substantially lower compared to no intervention, in particular, if women perform worse than men. Comparing the quality of winners is therefore a reasonable efficiency analysis. With advice, the conditions, under which winners are determined, do however not change.

Appendix E. Instructions

General Instructions

Only Generation 1:
In the experiment today you will be asked to complete six different tasks. The method we use to determine your earnings varies across tasks. Before each task, we will describe in detail how your payment of that task is determined. Your total earnings at the end of the experiment is the sum of the following components: (1) A e5 show up fee; (2) e4 for completing Tasks 1-4; (3) In addition, for Tasks 1-4, we will randomly select one of the four tasks and pay you based on your performance in that task; (4) You will be paid for Tasks 5 and 6. Once you have completed all tasks we determine which of the first four tasks counts for payment by drawing a number between 1 and 4. At the end of the experiment, we ask you to stay seated. We will come to you and pay you in private. During the duration of the experiment the use of cell phones is prohibited.

Only Generation 2:
The experiment today will begin with a waiting period of approximately 15 minutes. After these 15 minutes we will instruct you about the next steps. We are asking you to spend the waiting period silently at your assigned seats, without talking to each other or on the phone.

You may read or engage in any other quiet activity as you wish. At the end of the experiment you will be paid e2 for having waited quietly.

In the experiment today you will be asked to complete five different tasks. The method we use to determine your earnings varies across tasks. Before each task we will describe in detail how your payment of that task is determined. Your total earnings at the end of the experiment is the sum of the following components: (1) A e5 show up fee; (2) e2 for the waiting period; (3) e4 for completing Tasks 1-4; (4) In addition, for Tasks 1-4, we will randomly select one of the four tasks and pay you based on your performance in that task; (4) you will be paid for Task 5. Once you have completed all tasks we determine which of the first four tasks counts for payment by drawing a number between 1 and 4. At the end of the experiment, we ask you to stay seated. We will come to you and pay you in private. During the duration of the experiment the use of cell phones is prohibited.

Both Generations:
It is important that you do not talk with one another for the duration of the experiment. We also ask you that you do not look at the screens of the other participants. You can ask us at any point in time. If you have a question, please raise your hand and one of the experimenters will come to you.

Task 1 – Piece rate
In Task 1 you have to calculate a series of sums of five two-digit numbers (see "Screenshot Task 1"). You will be given 5 minutes to calculate the correct sum of a series of these problems. You cannot use a calculator to determine this sum. However, you are welcome to write the numbers down and make use of the provided scratch paper. You submit an answer by clicking the button "Next" with your mouse. When you submit an answer, the computer will immediately tell you whether the answer is correct or not and a new problem is generated. Your answers to the problems are anonymous.

If Task 1 is the one randomly selected for payment, then you earn 50 cents per problem you solve correctly in the 5 minutes. Your payment does not decrease if you provide an incorrect answer to a problem. We will refer to this payment scheme as the piece rate payment.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 1 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task. After clicking the "Continue" button, the task will begin immediately.

Suppose you have solved 2 problems correctly and 3 problems incorrectly, what is your payment for Task 1 if it is chosen for payment?

Task 2 – Tournament

As in Task 1 you will be given 5 minutes to calculate the correct sum of a series of five 2-digit numbers. However, for this task your payment depends on your performance relative to that of a group of other participants. Each group consists of four people; the three other members of your group are located in the same row as you. If Task 2 is the one randomly selected for payment, then your earnings depend on the number of problems you solve compared to the three other people in your group. The individual who correctly solves the largest number of problems will receive e2.00 per correct problem, while the other participants receive no payment. If there are ties the winner will be randomly determined. We refer to this payment scheme as the tournament payment. You will not be informed of how you did in the tournament until you have completed all five tasks.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 2 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task. After clicking the "Continue" button, the task will begin immediately.

Suppose you have solved 2 problems correctly and 3 problems incorrectly, and that everybody else in your group solved 1 problem correctly. What is your payment for Task 2 if it is chosen for payment?

Suppose you have solved 2 problems correctly and 3 problems incorrectly, and that one person in your group solved 3 problems correctly. What is your payment for Task 2 if it is chosen for payment?

Task 3 – Choice

As in the previous two tasks you will be given 5 minutes to calculate the correct sum of a series of five two-digit numbers. However, now you have to choose which of the two payment schemes, piece rate or tournament, you prefer to apply to your performance in the third task.

If Task 3 is the one randomly selected for payment, then your earnings for this task are determined as follows. If you choose the piece rate you receive 50 Cents per problem you solve correctly. If you choose the tournament your performance will be evaluated relative to the performance of the other three participants of your group in the Task 2-tournament. The Task 2-tournament is the one you just completed. If you correctly solve more problems than the other three members of your group in Task 2, then you receive e2.00 for each correct sum, which is four times the amount from the piece rate. You will receive no earnings for this task if you choose the tournament and do not solve more problems correctly now, than the other three members of your group in the Task 2-tournament. If there are ties the winner will be randomly determined. You will not be informed of how you did in the tournament until all five tasks have been completed.

Only Generation 1:
The computer screen following the control question will ask you to choose whether you want the piece rate or the tournament applied to your performance. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Only Generation 2:
(a) Advice
Before deciding on a payment scheme, you will receive some advice as to which one to choose. Your advisor is a person from the group next door who has already completed Tasks 1-4 and who knows how the members of his own group performed in Tasks 1 and 2. Each member of your group will be randomly assigned a different advisor. First, you are asked to send your advisor information on the number of problems you solved correctly in Tasks 1 and 2. Your advisor will then tell you whether he or she thinks you should enter the tournament and probably also give you a reason for his/her advice.

The next computer screen will ask you to enter the numbers of correct problems you solved in Tasks 1 and 2. You will then have to wait for a moment to receive a message from your advisor.

(b) Entry decision
The computer screen that informs you about the advice you received will ask you to choose whether you want the piece rate or the tournament applied to your performance. You will then be given 5 minutes to calculate the correct sum of a series of five randomly chosen two-digit numbers.

Both Generations:
Are there any questions?
Control question
To ensure you correctly understood, how the payment for Task 2 is calculated, please answer the following question. Note that the numbers used in the question are not indicative of what constitutes a good performance in this task.

Suppose you have chosen the piece rate and that you solved 3 problems correctly and 1 problem incorrectly. What is your payment for Task 3 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that everybody else in your group solved 1 problem correctly in Task 2. What is your payment for Task 3 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that another person in your group solved 3 problems correctly in Task 2. What is your payment for Task 3 if it is chosen for payment?

Task 4 – Payment scheme for Task 1
You do not have to add any numbers for the fourth task of the experiment. Instead we will pay you again for the number of problems you solved in Task 1 – Piece Rate. However, you now have to choose which payment scheme you want applied to the number of problems you solved. You can either choose being paid according to the piece rate, or according to the tournament.

If the fourth task is the one selected for payment, then your earnings for this task are determined as follows. If you choose the piece rate you receive 50 Cents per problem you solved in Task 1.

If you choose the tournament your performance will be evaluated relative to the performance of the other three members of your group in the Task 1-piece rate. If you correctly solved more problems in Task 1 than the other three members of your group did then you receive four times the earnings of the piece rate, which is e2.00 per correct problem. You will receive no earnings for this task if you choose the tournament and did not solve more problems correctly in Task 1 than the other members of your group. If there are ties the winner is determined randomly.

The next computer screen will tell you how many problems you correctly solved in Task 1, and will ask you to choose whether you would like to apply the piece rate or the tournament rate to your performance.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 4 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose you have chosen the piece rate. Suppose further that you have solved 2 problems correctly and 3 problems incorrectly in Task 1. What is your payment for Task 4 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you have solved 2 problems correctly and 3 problems incorrectly in Task 1, and everybody else in your group solved 1 problem correctly in Task 1. What is your payment for Task 4 if it is chosen for payment?

Suppose you have chosen the tournament. Suppose further that you solved 2 problems correctly and 3 problems incorrectly, and that another person in your group solved 4 problems correctly in Task 2. What is your payment for Task 4 if it is chosen for payment?

Task 5 – Self-evaluation

In this task you are asked to guess your ranks of your performances in Tasks 1 and 2. Since there are four members in your group your rank may be between 1 and 4, with 1 being your rank if you (correctly) solved the largest number of problems in your group and 4 being your rank if you solved the lowest number.

For each correct guess you will receive e1. If your guess is not correct, you will receive no earnings for this guess. In case of ties in the actual ranks, we count every answer that could be correct as correct. For example, if the performance in the group was 5, 5, 4, 4, then an answer of "last position" and "third position" is correct for somebody who solved 4 problems correctly, and an answer of "first position" and "second position" is correct for somebody who solved 5 problems correctly. Note that the numbers used in this example are not indicative of actual performances in Tasks 1 and 2.

Are there any questions?

Control question

To ensure you correctly understood, how the payment for Task 5 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose that in Task 1 you solved 3 problems correctly and the other members of your group solved, respectively, 1, 2, and 3 problems. Suppose further that you estimated your rank to be "second position". What is your payment for this estimate?

Only Generation 1:

Task 6 – Advice

In the room next to us there are other groups who also complete Tasks 1-4 (the same ones you just completed). At this point they have completed Tasks 1 and 2, but have not yet started with Task 3, that is, their next task is to decide between the tournament and the piece-rate. You will be randomly matched to one of them, whom we will refer to as your "advisee", and your task is to advice your advisee in his or her choice between tournament and piece rate. Before you give your advice, your advisee will send you information on the number of problems he or she solved correctly in Tasks 1 and 2.

The first step is that you send your advisee a message telling him or her whether you recommend entering the tournament. In a second step you may give a reason for the advice you choose. For this purpose we provide you with a list of reasons. You may select as many reasons as you wish (including none, in case you don't wish to select any of the reasons provided).

As a payment for this task you will receive 50% of the Task 3 earnings of your advisee. This means that if your advisee chooses the piece rate you receive 25 Cents (50% of 50 Cents) per problem he/she solves correctly. If your advisee chooses the tournament and his/her performance is better than the Task 2 performance of his/her group members, you receive e1.00 (50% of e2.00) for each problem he/she solves correctly. Finally, if your advisee chooses the tournament and his/her performance is not better than the Task 2 performance of his/her group members, you will receive no earnings. Note that you will be paid even if your advisee does not receive a payment for Task 3 (because Task 3 was not the one randomly selected for payment).

<u>Are there any questions?</u>

<u>Control question</u>

To ensure you correctly understood, how the payment for Task 5 is calculated, please answer the following questions. Note that the numbers used in the questions are not indicative of what constitutes a good performance in this task.

Suppose your advisee has chosen the piece rate. Suppose further that your advisee solved 3 problems correctly and 3 problems incorrectly. What is your payment for Task 6?

Suppose your advisee has chosen the tournament. Suppose further that your advisee solved 2 problems correctly and 1 problem incorrectly, and everybody else in his/her group solved 1 problem correctly in Task 2. What is your payment for Task 6?

Suppose your advisee has chosen the tournament. Suppose further that your advisee solved 2 problems correctly and 3 problem incorrectly, and another person in his/her group solved 3 problems correctly in Task 2. What is your payment for Task 6?

CHAPTER II: NOT JUST LIKE STARTING OVER – LEADERSHIP AND REVIVIFICATION OF COOPERATION IN GROUPS

1. Introduction

A general problem with cooperation in groups is that it may be high at times but then decay to a large extent. Once trapped in a low cooperation situation it is hard to escape from there. Even if the benefits of improved cooperation are self-evident, any process designed to bring about an improvement of the situation faces substantial obstacles. All individuals involved could be better off if all cooperated more, but any one individual who unilaterally starts doing so will feel taken advantage of if others do not do the same. Cooperation decays may happen to occur in institutions and firms at times. The important question is: how can groups of individuals escape such a low cooperation trap?

Two elements may be useful to escape a low cooperation situation: a (mental) new start and a leader who triggers the new start. How can a leader take advantage of an exogenous change like a new start and trigger cooperation in the group effectively *after* cooperation has failed? Some exogenous changes in the circumstances of the groups, like the start of a new week or a new season, may create a sense of a new beginning and may lead to a revivification of cooperation. Moreover, companies, organizations and other human groups with leaders have access to instruments that can facilitate a turnaround. When cooperation failure has occurred it is one of leaders' natural roles to take action to reinforce a new beginning. An additional challenge is to revive cooperation in an enduring way, that is, to trigger a cooperation increase that is not short-lived, but is sustained over time.

Our experiment builds on two important results of earlier experimental work. A common observation in experimental studies of public goods games with voluntary contributions is that in environments with a finite horizon cooperation levels are initially rather high but then decrease steadily over time.[25] We use this pattern to create the experience of decreasing cooperation in a group. At the same it has been shown that in fixed groups the level of cooperation can be driven up again by simply restarting the game after the horizon has been reached. In the experiments reported in Andreoni (1988) participants play the voluntary contribution game in the finitely repeated form and, after the initially announced ten rounds are over, they are informed that there will be some additional rounds of the same game. Contributions go up again after the prolonged experiment is announced. In this experiment play was suspended after three additional rounds and during these rounds the cooperation level stayed up. This effect is called the "restart effect."[26]

Croson (1996) follows up on Andreoni (1988) with public goods experiments in which, after the initial ten rounds, ten additional rounds are announced. The results confirm that the restart leads to an initial increase of cooperation in fixed groups. However, after the initial increase in cooperation, the decline in cooperation begins again and play ends up at an even lower level than at the end of the first ten rounds. Cooperation can be revived by starting

[25] See Davis and Holt (1994) and Ledyard (1995) for reviews. See, e.g., Isaac et al. (1984), Andreoni (1988, 1995), Weimann (1994), Laury et al. (1995), Croson (1996), Burlando and Hey (1997), Gächter and Fehr (1999), Ockenfels and Weimann (1999), Sonnemans et al. (1999), Keser and van Winden (2000), Fehr and Gächter (2000), Park (2000), Masclet et al. (2003), Croson et al. (2005), Carpenter (2007), Sefton et al. (2007), Egas and Riedl (2008), Gächter et al. (2008), Herrmann et al. (2008), Nikiforakis and Normann (2008), Neugebauer et al. (2009) and Fischbacher and Gächter (2010).

[26] The work by Andreoni (1988) is aimed at distinguishing between the learning hypothesis, which suggests that subjects learn the incentives of the game throughout the experiment, and the strategies hypothesis, which suggests that participants play with the objective to influence the other group members' actions, i.e. they take into account the repeated play and use contributions as signals about future contributions.

over, but the effect is short-lived. The restart effect works in the short-run, but not in the long run. This pattern of behavior is the starting point for our study.

We use experiments to study how, after a history of decay in cooperation levels, cooperation in groups can be revived in an enduring way by using various managerial strategies. That is, are there ways to avoid that the positive short-run restart effect vanishes over time? We study this issue in the context of a public good game involving a leader. We choose a structure involving a leader, because we are mostly motivated by issues of successful teamwork in organizations. Almost all types of institutions, firms, departments and (sport) teams are organized in some kind of hierarchical structure and guided by a leader. Societies are lead by politicians or ideological leaders, companies by managers, departments by directors and sports teams by coaches.

In our set-up, leadership takes the form of leading-by-example used in the studies by Güth et al. (2007), Rivas and Sutter (2009), Gächter et al. (2010) and Potters et al. (2007) among others. The game is sequential and each group is composed by one leader and three followers. The group composition is constant over time, i.e. the group members are the same over the entire experiments. First, the leader decides on his contribution to the public good. The followers are informed about their leader's decision and simultaneously choose their contribution levels. Both, leaders and followers of a group influence the group outcome through their contribution to the public good. Leading by example can be a conscious or subconscious form of leadership being present in a broad range of situations. The importance of leadership, and in particular of leading by example, becomes clear when thinking about outstanding business leaders like Steve Jobs or Jack Welch. But also in every-day situations, this kind of leadership is a key feature of the organization and coordination of a group of individuals.

If, in an environment with leading by example, cooperation decays after some time, there are several ways in which things can change. Here we study two interventions and a combination of both interventions that a priori can be expected to lead to a stronger revivification of cooperation than that following a pure restart and that are interesting from a managerial point of view. Two interventions involve a restart, but add another element aimed at avoiding that the increase in cooperation is only short-lived. A third treatment combines the two elements and compares the joint effect with the effect of one intervention only.

We have four treatments. The first is the *restart* treatment, a control treatment in which the restart is pure, that is not accompanied by any other change in the environment. Our second treatment is the *comprehension/advice* treatment, a restart with the provision of a detailed explanation of the causes of the decrease in cooperation and of advice for future contributions. Our third treatment is the *communication* treatment, a restart with a one-way free form message sent by the group leader to the followers. In the fourth treatment, the *comprehension/advice/communication* treatment, we combine the second and third treatment. After participants have received the detailed explanation of the causes of the decrease and the advice for future contributions, the group leader can send a one-way free form message to the followers. All three interventions involve a restart in the sense that, after a number of experimental rounds, additional rounds are played, but they all also involve one or two element that go beyond the pure restart. We think that the two interventions and their combination are central for the issue at hand.

Our comprehension/advice treatment is inspired by the common practice in every-day business of obtaining expert analysis and advice from a consultancy firm for instance. McDonald and Westphal (2003) find that CEOs tend to seek advice when performance deteriorates, which in our context corresponds to decreasing cooperation. The effect of external consultancy and advice on performance is however rather inconclusive as a number of field experiments with micro-, small and large organizations in developing countries obtain different results.[27] Compared to the advice provided in Chaudhuri et al. (2006), where common knowledge advice from a previous generation of participants increases cooperation significantly, our advice has the nature of an exogenous expert advice. In our context, participants first receive an expert explanation of the cooperation decay followed by an advice on how to prevent such a decay. We give participants insights into the difficulties of cooperating over time based on the analysis of this problem contained in the influential paper by Fischbacher and Gächter (2010). Our design allows us to study whether a careful analysis of what goes wrong and constructive advice on how to improve team performance can cause a change in participants' cooperation, beyond the one that comes from a pure restart. A priori, a better understanding of the causes behind the decay of cooperation appears to be a good basis for improvement. In this sense, we provide information on the nature of the problem and on how group members can contribute to an improvement (see McGuire, 1985). Our design allows us to shed light on the conjecture that receiving a rational explanation by outsiders can lead to an improvement. To further strengthen this point, we conclude the explanation with a concrete advice on how a decrease in cooperation can be prevented. We expect that the advice will make participants update their beliefs about others' contributions over time.

Communication between manager and co-workers has been shown to be a crucial element of the successful performance of a firm. Brandts and Cooper (2007) show that in a coordination game communication between manager and employees is quite effective in improving performance of groups. In the experimental public goods literature, there is wide evidence that communication from the very start enhances cooperation (LIT). Koukoumelis, Levati, and Weisser (2012) show that one-way communication by one group member increases cooperation significantly in the simultaneously played game, where communication is possible from the outset and not just after cooperation has broken down. The crucial difference to our design is that individuals do not have any (negative) cooperation experience with their group members before communication takes place. Our case is one in which cooperation levels end up at a low level and we ask whether communication at this point makes it possible to escape from such a situation. It is not clear whether cheap-talk communication works after the group *has experienced* cooperation failure with the same group members. To our knowledge, this has not been studied in previous experiments. Moreover, Olson and Zanna (1993) report evidence that information from in-group sources will have more impact than information from external sources.

The combination of an outsider expert explanation and advice and the leader's communication with the followers is expected to yield the highest contribution levels with the effect of comprehension and advice being two-fold. In the first place, the *direct* effect of comprehension and advice is expected to be positive as pointed out above. The *indirect* effect of comprehension and advice is expected to work through the leader's communication with the followers. We expect that leaders (and followers) understand the game better after having read the explanation and that they have a good idea about what would be the best thing to do after having received the advice. The quality of their communication should thus improve and have a stronger effect on cooperation than in the communication only treatment.

Among other contributions to the existing literature on cooperation, our design involves two restarts to allow us to study to what extent the effect of the different interventions becomes stronger over time. In the first part of the experiment, we let participants play the game without any intervention. The purpose of the first part is to create the experience of decreasing cooperation in the group and to provide an interesting situation for a restart. Our contribution to the existing literature on cooperation is fourfold: First, we analyze the pure restart in a sequential form of the voluntary contribution game. Second, we

[27] See, e.g., Drexler et al., 2010; Karlan and Valdivia, 2011; Bruhn and Zia, 2011; Bruhn et al., 2012; Karlan et al., 2012; Bloom et al., 2013.

study the effect of communication *after* having possibly experienced cooperation failure. Third, "expert" advice in the context of a voluntary contribution game has not been studied previously to our knowledge. Previous experiments have shown that communication *from the start* (Koukoumelis et al. 2012) and commonly known advice from another experienced (non-expert) person outside the group (Chaudhuri et al. 2006) increase cooperation significantly. It is not clear whether the same is true for communication and expert advice after participants have experienced decreasing cooperation. Fourth, the repeated restart allows us to study whether, if the first effect is positive, repeated interventions can further strengthen and lead to sustained cooperation levels. It is likely that the repeated intervention can strengthen a positive experience after a negative cooperation experience.

We find that the effects of the pure restart and comprehension/advice do not differ significantly in the long-run, suggesting that exogenous expert consultancy revives cooperation to the extent of a pure restart, but not beyond the effect of a pure restart. The informational content and therefore the understanding of the game are supposed to be highest in the comprehension/advice treatment and, in addition, participants receive advice on how to act. However, the message sent by the leader to the followers in the communication treatment revives cooperation significantly compared to the pure restart and also compared to the comprehension/advice treatment. The combination of comprehension/advice and communication does not outperform the effect of pure communication. If any, communication by the leader without the expert's explanation and advice results to work in the repeated implementation: *repeated* communication further reinforces the reviving effect on cooperation by eliminating the decline in contributions over time almost completely. Communication (of the leader) is the most effective managerial instrument in our experiment. The positive effect of repeated communication with limited frequency is in consonance with other research in the psychological literature (McGuire, 1985).

2. Experimental Design

In section 2.1., we explain the sequential voluntary contribution game used in our experiment and the theoretical predictions assuming selfish players. We also provide some general information on the procedure of the experimental session is provided. In section 2.2., the control treatment and the intervention treatments are discussed in detail.

2.1.The game and general procedure

In the leading-by-example setting we study, a voluntary contribution game is played repeatedly by fixed groups of four participants. Group members are matched randomly at the beginning of the experiment. There are two roles: leader and follower. The role of the leader is assigned to one of the group members and the remaining group members are followers. The roles are assigned randomly at the beginning of the experiment and are the same throughout the entire experimental session.

The payoff function is the same for both, leaders and followers. The individual endowment is $E = 40$, the return rate of the private good is $r_P = 1$, and the return rate of the public good is $r_V = 0.5$ yielding the following payoff function of individual i in round t:

$$\pi_{i,t} = \underbrace{\left(40 - h_{i,t}\right)}_{\text{Payoff from private good}} + \underbrace{0.5 \cdot \sum_{j=1}^{4} h_{j,t}}_{\text{Payoff from public good}}$$

An individual i's contribution in round t to the public good is denoted by $h_{i,t}$, the contributions by all group members are denoted by $h_{j,t}$ with $j=1,\ldots,4$. The game is played sequentially by the four players over a total of 36 rounds and the group composition does not change over time. All rounds have three stages. In the first stage of the game, the leader of each group decides how much of the endowment to contribute to the public good. In the second stage, followers are informed about their leader's decision and decide how much of their individual endowment to contribute to the public good. In the third stage, all players are informed about the average contribution by the other group members, the sum of contributions by all group members and the individual payoff.[28]

The equilibrium contribution of leaders and followers in the sequential structure of the game is the same as in the simultaneous game, i.e. zero. This holds for the stage game as well as for the finitely repeated game, which can be shown by backward induction. Therefore, the equilibrium contribution in the finitely repeated sequential voluntary contribution game is zero, too. The socially optimal solution is just the same as in the finitely repeated simultaneous game: Each group member $j = 1,...,4$ contributes in each round the entire individual endowment E to the public good leading to an individual round payoff of $r_V \cdot 4 \cdot E = 80$.

The general instructions are handed out to the participants on paper and read aloud by one of the experimenters at the beginning of the experiment. In the general instructions in the appendix A.1, the chronological order of an experimental session and the three stages of each round are represented. They are the same for the control treatment and the three treatments. Participants get the information about the total number of rounds as well as the structure of the rounds before the experiment starts. Also, they are informed that the 36 rounds of the repeatedly played voluntary contribution game are divided into three parts with 12 rounds each and that they would get part-specific instructions at the beginning of each part. Participants at the same time know about the overall horizon and the break-up into blocks of 12 rounds, which are meant to represent exogenous moments of re-start akin to week, seasons etc.

Additional part-specific instructions (see appendix) including the information that the group composition would remain the same over the 12 rounds of the subsequent part are shown on the computer screen just before the corresponding part starts and also announced aloud by one of the experimenters. The restart and the interventions take place at the beginning of part 2 (before round 13) and part 3 (before round 25). Thus, the part-specific instructions differ for the control treatment and the three intervention treatments. Having three parts of twelve rounds allows us to investigate the effects of the two restarts instead of just one. We discuss this in more detail in the results section. A twelve-round part can be seen as a work-period (week, month, quarter, year), a season, the time a particular project lasts or any other length of time after which there is a natural break in the interaction. After the experiment finishes, participants are required to fill out a questionnaire and are paid the earnings in private.

The experimental sessions were conducted at the Universitat Autònoma de Barcelona (UAB, Spain) and programmed with the experimental software z-Tree, Fischbacher (2007). Participants were mainly undergraduate students from the UAB and were recruited using the online recruitment system ORSEE, Greiner (2004). A total of 208 participants took part in twelve experimental sessions composed by 123 women and 85 men. The average earnings per person were 19.70 Euro (including a show-up fee of 5.00 Euro). The average duration of a session was 2 hours 30 minutes.

[28] We inform participants only about the average contribution by the other group members instead of the individual contributions to create some slightly imperfect information which we thought would facilitate a decrease in contributions over time.

2.2. Treatments

In the control treatment with the pure restart, neither the group composition changes nor do participants get any additional information or have to take any new type of action. Both at the beginning of rounds 13 and 25 participants are informed in the part-specific instructions that they will continue playing in the same group composition as before during the subsequent twelve rounds. Note that the effects of a pure restart were studied by Andreoni (1988) and Croson (1996) in a simultaneous voluntary contribution game and the restart was a surprise for participants. Hence, our control treatment is an extension and not a pure replication of previous work. To our knowledge, the restart effect as such has not yet been studied in a sequential form of the game.

Table 3.1: Overview over treatments

Treatment	Characteristics	Intervention	Repetitions	Observations
(Control) Treatment R	Restart	Before part 2 and 3	36 rounds	15 groups
Treatment CA	Comprehension and advice text	Before part 2 and 3	36 rounds	13 groups
Treatment C	One-way free form communication from leader to followers	Before part 2 and 3	36 rounds	12 groups
Treatment CAC	Comprehension and advice text; Subsequently one-way free form communication from leader to followers	Before part 2 and 3	36 rounds	12 groups

In the *comprehension/advice* treatment, we explain to participants, before the start of part 2, how contributions usually evolve in related experiments and give an explanation of why they typically decline. Then we provide some advice on what to do to avoid the decline and to reach and maintain high earnings from the public good. The idea of this treatment is that of a working group receiving external expert analysis, explanation and advice. Following psychological research on attitude change and persuasion (McGuire, 1985) we provide participants with a rational analysis of the causes of cooperation decay and with an evidence-based advice on how the process of decay can be prevented. The better understanding of the problem at hand (McGuire, 1985), i.e. the better understanding of the (possibly experienced) decay in cooperation, and increased beliefs about others' contributions caused by the advice lead us to expect that contributions will be higher in the comprehension/advice treatment than in the control restart treatment.

The content of the information is the following: We first inform participants that we observed a decline in average contributions over part 1 in previous sessions driven by followers undercutting previous contributions on average. We then explain that a study showed that the decline in contributions in the repeated game occurs because participants are on average imperfect conditional contributors (Fischbacher and Gaechter, 2010). Finally, we state that it is recommendable that followers contribute at least as much as the leader of their group to reach and maintain high earnings from the public good. Note that we use in the advice part the Spanish plural form of you ("vosotros") referring to the total group earnings from the public good. In the rest of the comprehension text as well as in the remaining instructions, we use the singular form of you ("tu"). Before part 3, we give a short reminder of the explanation and the recommendation. The text of the comprehension/advice instructions for part 2 and part 3 can be found in the appendix. We wanted to make sure that participants understood well what was going on in the game and we wanted to give a clear comprehensive recommendation of what to do to avoid the decline. We thought carefully about the information we put in the explanation and advice and let non-economists proofread it for understandability. Also, we gave participants enough time to read the information again after we had read it out aloud and asked if anyone had a question before proceeding.

In the *communication* treatment, the leader of a group sends a one-way free form text message to the followers before part 2 and part 3 begin, respectively. Except for standard rules for free form communication in experiments, leaders are free to write whatever they want. Koukoumelis et al. (2012) show that one-way free form communication by one group member increases contributions in the simultaneous voluntary contributions game significantly. We are interested in studying behavior in the sequentially played voluntary contribution game and after a decrease in contributions; our emphasis is on reviving cooperation after it has died down, which is a crucial difference to previous experiments on communication and cooperation. We think that it is an interesting context since a negative cooperation experience is particularly crucial that leaders find the right words to get out of the trap. Still we expect that communication will increase cooperation by more than the pure restart. Note that the informational content and understanding contained in the message participants get in the comprehension/advice treatment can be considered to be at least as precise and deep as in the communication treatment. The advice in the comprehension/advice treatment (although transmitted in a soft way to not be perceived as an order) is supposed to be clear and comprehensive.

In the *comprehension/advice/communication* treatment, all participants receive the exactly same explanation and advice as in the comprehension/advice treatment before part 2 and part 3 begin, respectively. On the subsequent screen, leaders can then send a one-way free form message to the followers, respectively. The instructions and the procedure are identical to the communication treatment. We expect cooperation to be highest because participants are supposed to understand the game and receive advice. On top, the externally provided expert information is supposed to improve the communication content of leaders' messages.

We have one control treatment and three intervention treatments. In the following, we will denote the restart control treatment by "treatment R," the comprehension/advice intervention by "treatment CA," the communication intervention by "treatment C," and the comprehension/advice intervention in combination with the communication by "treatment CAC." Table 3.1 provides a summary of the characteristics and the number of observations for each treatment. We have a total of 15 (independent) group observations for treatment R, 13 group observations for treatment CA, 12 group observations for treatment C, and 12 group observations for treatment CAC.

3. Experimental Results

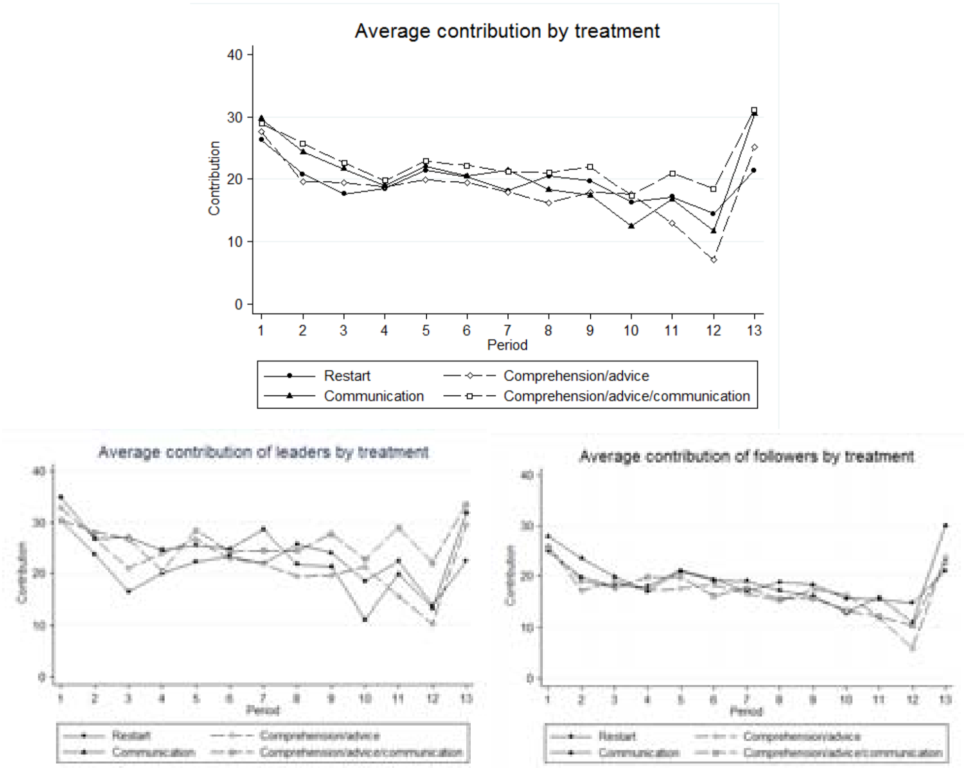
We start the presentation of our analysis with some preliminary results. In section 3.1, we confirm the expected contribution decline over time and a short-run restart effect in cooperation in our sequential form of the voluntary contribution game and in section 3.2, we confirm that overall contributions do not differ in part 1 and that leaders contribute significantly more than followers. We are particularly interested in the long-run effect of the three interventions that can be added to a pure restart and that are interesting from the managerial point of view. Section 3.3 shows that the leader's communication with the followers (with and without the comprehension/advice stage) outperforms the pure restart, and the comprehension/advice intervention both at the beginning and overall in part 2. Section

3.4 draws a similar picture for part 3 underlining the strong enduring effect of repeated communication.

3.1. Decline in contributions and restart

We start our data presentation by confirming that contributions decline over part 1 (rounds 1 through 12) and that there is a restart effect in round 13. In Figure 3.1, the average contributions are depicted for rounds 1 through 13 and for control treatment R and treatments CA, C, and CAC, separately. In the upper panel of Figure 3.1, the average contributions of all participants are shown. In the lower panels of Figure 3.1, the average contributions of leaders only and followers only are presented on the left and right, respectively. Average contributions in all four treatments decrease in part 1, as expected, and evolve similarly. The average contributions and corresponding standard deviations of all participants, leaders and followers, in part 1 are very similar as shown in Table 3.2. We will get back to the part 1 average contributions in the next section.

Figure 3.1: Average contributions in control treatment R and treatments CA, C and CAC (round 1 through 13) for all participants (upper panel), leaders only (lower left panel), and followers only (lower right panel).



The decline in cooperation is confirmed in pooled OLS regressions clustering for groups, see Table 3.3. The observations are those of all 208 participants of control treatment R, and treatments CA, C, and CAC. We cluster for group to control for the correlation of contributions within a group. In regression models (1a), (1b), (2a), (2b), (3a), and (3b), observations are those from part 1 (round 1 through 12), part 2 (round 13 through 24), and part 3 (round 25 through 36), respectively. In models (1a), (2a), and (3a), the individual contributions are regressed on a round variable taking values between 1 and 12, a dummy variable for each of the three interventional treatments CA, C, and CAC and a dummy variable, which takes the value one if the individual is leader and zero if the individual is follower. The reference treatment is thus treatment R. In models (1b), (2b), and (3b), an interaction term between the round variable and each of the three treatments CA, C, and CAC is added to the corresponding model. In model (1b), the coefficient estimate for the round variable is negative and highly significant at the one percent level (just as in the other regression models) indicating that contributions in control treatment R decrease over the rounds of part 1 by 0.59 EMU per round on average.

Table 3.2: Descriptive statistics of contributions by treatment and on the group, leader and follower level.

		Group		Leaders		Followers	
Average contributions	**N**	**mean**	**(sd)**	**mean**	**(sd)**	**mean**	**(sd)**
Treatment R							
Part 1 (round 1-12)	15	19.28	(7.442)	21.86	(9.557)	18.41	(7.334)
Part 2 (round 13-24)	15	18.20	(7.588)	22.24	(9.443)	16.85	(7.847)
Part 3 (round 25-36)	15	16.03	(9.769)	22.31	(10.47)	13.94	(10.03)
Treatment CA							
Part 1 (round 1-12)	13	17.86	(7.086)	21.86	(7.321)	16.53	(7.106)
Part 2 (round 13-24)	13	18.51	(9.675)	22.83	(9.919)	17.07	(9.886)
Part 3 (round 25-36)	13	17.12	(10.97)	22.15	(11.78)	15.44	(11.48)
Treatment C							
Part 1 (round 1-12)	12	19.62	(6.068)	23.32	(6.770)	18.39	(6.414)
Part 2 (round 13-24)	12	26.56	(8.364)	28.10	(9.810)	26.04	(8.108)
Part 3 (round 25-36)	12	29.31	(10.32)	30.56	(11.01)	28.89	(10.24)
Treatment CAC							
Part 1 (round 1-12)	12	21.93	(6.714)	25.73	(7.316)	20.67	(7.642)
Part 2 (round 13-24)	12	27.50	(9.725)	30.69	(9.814)	26.44	(10.68)
Part 3 (round 25-36)	12	26.13	(12.16)	28.18	(11.40)	25.44	(12.84)

In model (1b), the dummy variables for the three treatments CA, C, and CAC are positive and not significant, though the coefficient estimate of the treatment C-dummy (4.4) has a p-value of 0.114. All three interaction terms of the treatment and the round variable are negative. For treatment C, the interaction term is significant at the five percent level in part 1. Compared to control treatment R, contributions start somewhat higher in treatment C in round 1 and the contribution decrease is steeper by 0.63 EMU per round in part 1. Since there are no treatment differences in part 1, there should be no differences in contributions between treatments. Note that the coefficient estimates of the treatment dummy variables in model (1a) are all insignificant indicating that there are no treatment differences in part 1.[29]

The contribution decline over rounds is also confirmed in regression models (4a) and (4b), where observations are those from round 1 through 36. In model (4a), the individual contributions are regressed on a round variable taking values between 1 and 36, a dummy variable for part 2 and part 3, respectively, a dummy variable for each of the three interventional treatments CA, C, and CAC, and a dummy variable for the role “leader.” In model (4b), interaction terms between each part dummy and each interventional treatment dummy are added to model (4a). For the moment, note only that the round variable in both models is negative and highly significant meaning that contributions decrease on average over all four treatments by 0.8 EMU per round. We will get back to the remaining results of Table 3.3 later.

[29] We don’t have an explanation for the difference in the slope, since there are no treatment differences in part 1. It should be random.

Table 3.3: Pooled OLS regression (Data: treatments R, CA, C, and CAC).

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)	(4a)	(4b)
	Dependent variable: Contribution							
VARIABLES	(Part 1)	(Part 1)	(Part 2)	(Part 2)	(Part 3)	(Part 3)	(Part 1-3)	(Part 1-3)
Part round (1-12)	-0.882***	-0.594***	-0.848***	-0.754***	-0.679***	-0.673**		
	(0.137)	(0.220)	(0.167)	(0.251)	(0.148)	(0.261)		
Round (1-36)							-0.803***	-0.803***
							(0.101)	(0.101)
Part 2							12.37***	8.557***
							(1.721)	(2.018)
Part 3							21.35***	16.03***
							(3.005)	(3.985)
Comprehension/advice	-1.413	1.924	0.311	2.886	1.086	5.288	-0.00513	-1.413
	(2.676)	(2.604)	(3.233)	(4.385)	(3.846)	(5.146)	(2.675)	(2.676)
Communication	0.345	4.405	8.358***	10.68**	13.28***	9.895**	7.328***	0.345
	(2.529)	(2.743)	(3.020)	(4.282)	(3.793)	(4.923)	(2.390)	(2.528)
Comp./advice/communication	2.658	3.107	9.305***	6.849*	10.10**	9.089*	7.353**	2.658
	(2.653)	(2.858)	(3.323)	(3.827)	(4.197)	(4.807)	(2.910)	(2.653)
(CA)*(Part round)		-0.513		-0.396		-0.646*		
		(0.335)		(0.380)		(0.368)		
(C)*(Part round)		-0.625**		-0.357		0.521*		
		(0.299)		(0.534)		(0.308)		
(CAC)*(Part round)		-0.0691		0.378		0.155		
		(0.426)		(0.368)		(0.449)		
(CA)*(Part 2)								1.723
								(2.801)
(CA)*(Part 3)								2.499
								(4.196)
(C)*(Part 2)								8.013**
								(3.007)
(C)*(Part 3)								12.94***
								(4.382)
(CAC)*(Part 2)								6.647**
								(2.741)
(CAC)*(Part 3)								7.440*
								(3.936)
Leader	4.631***	4.631***	4.453***	4.453***	5.107***	5.107***	4.730***	4.730***
	(0.804)	(0.805)	(0.928)	(0.929)	(0.960)	(0.961)	(0.768)	(0.768)
Constant	23.85***	21.98***	22.60***	21.99***	19.17***	19.13***	20.27***	23.31***
	(1.952)	(2.047)	(2.336)	(2.929)	(2.809)	(3.507)	(1.932)	(1.935)
Observations	2,496	2,496	2,496	2,496	2,496	2,496	7,488	7,488
R-squared	0.073	0.077	0.129	0.134	0.157	0.165	0.110	0.128

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Pooled OLS (clustering for group), observations from round 1-12 (regression models 1a and 1b), round 13-24 (regression models 2a and 2b), round 25-36 (regression models 3a and 3b), round 1-36 (regression models 4a and 4b)

Dependent variable (contribution) takes values between 0 and 40.

Having confirmed the decline in contributions over the rounds of part 1, we now move to the short-run restart effect for the *sequential* voluntary contributions game, which was found in previous studies with the *simultaneous* voluntary contribution game, see Andreoni (1988) and Croson (1996). Therefore, we compare contributions in round 13 with contributions in round 12 (first restart). Contributions in round 1 through 13 are shown separately for each treatment in Figure 3.1 (upper panel: all participants; lower panel: leaders on the left, followers on the right). We will analyze the contributions on the group level, for leaders only, and for followers only. For comparisons on the group level, we calculate the average over the contributions of the leader and the three followers of a group resulting in 15 (control treatment R), 13 (treatment CA), 12 (treatment C), and 12 (treatment CAC) independent observations. For followers, one independent observation is given by the average over the contributions of the three followers of a group. The increase from round 12 to 13 is very clear for all four treatments and the increase is confirmed by non-parametric tests ($p < 0.061$ separately for each treatment and for average contributions, leaders' contributions, and average followers' contributions, two-sided Wilcoxon signed-rank test, N = 15 for control treatment R, N = 13 for treatment CA, N = 12 for treatment C, N = 12 for treatment CAC). Also the highly significant positive coefficient estimate of the part 2-dummy in regression model (4a) in Table 3.3 confirms an average increase of 11.57 EMU from round 12 to round 13 (part 2-coefficient estimate: 12.37 EMU; marginal round change: -0.80 EMU).

3.2. Contributions in part 1 and leaders' and followers' contributions

Even though the slope of the contributions in the first part is somewhat steeper in treatment C than in control treatment R, the contributions in part 1 do not differ between the four treatments. In the summary statistics in Table 3.2, the average contributions are summarized for each treatment and for each part. Average contributions of a part are the average over the group contributions in the twelve corresponding rounds resulting in 15 (control treatment R), 13 (treatment CA), 12 (treatment C), and 12 (treatment CAC) independent observations. Average contributions (standard deviations) in part 1 are 19.28 (7.442) in treatment R, 17.86 (7.086) in treatment CA, 19.62 (6.068) in treatment C, and 21.93 (6.714) in treatment CAC. As expected, the null hypothesis "no treatment differences in contributions in part 1" cannot be rejected (chi2(3df) = 1.906, $p = 0.592$, Kruskal-Wallis test). Also the pair-wise comparison of the part 1-contribution distributions does not reveal differences between treatments R, CA, C, and CAC ($p > 0.210$, pair-wise two-sided Mann-Whitney U test).

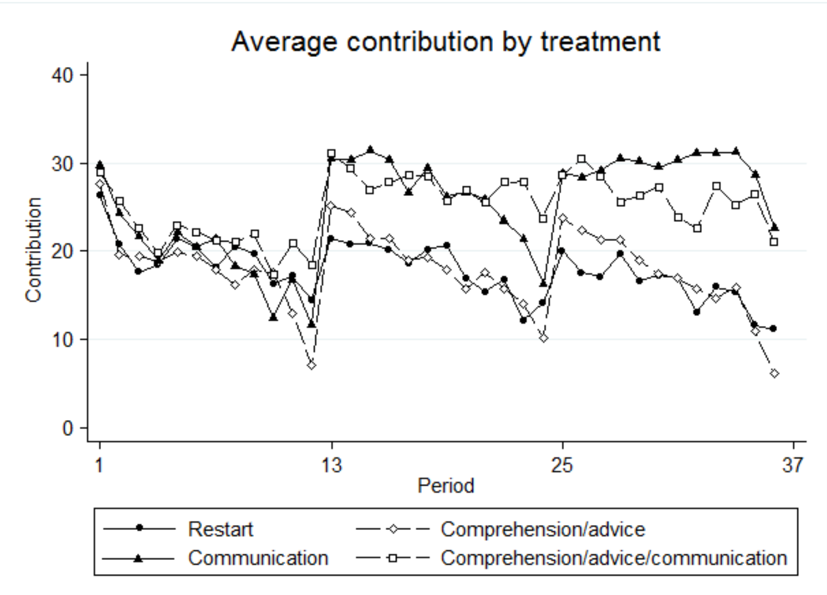
Contributions in part 1 are also the same when analyzing leaders and followers separately. For leaders, the part contributions are calculated taking the average over the contributions in the twelve rounds of a part on the individual level. For followers, the average part contributions are calculated over the average of the three group followers in the twelve rounds of a part leading to 15 (control treatment R), 13 (treatment CA), 12 (treatment C), and 12 (treatment CAC) independent observations. Neither for leaders (chi2(3df) = 1.390, $p = 0.708$, Kruskal-Wallis test; $p > 0.255$, pair-wise two-sided Mann-Whitney U test) nor for followers (chi2(3df) = 2.000, $p = 0.573$, Kruskal-Wallis test; $p > 0.191$, pair-wise two-sided Mann-Whitney U test), there are significant treatment differences in contributions in the first part of the experiment.

Comparing leaders' and followers' contributions, we find that leaders contribute significantly more than the followers of the corresponding group in each treatment and part ($p < 0.084$ for each treatment and part separately, two-sided Wilcoxon signed-rank tests, N = 15 for control treatment R, N = 13 for treatment CA, N = 12 for treatment C, N = 12 for treatment CAC) with two (slight) exceptions: in part 2, the difference between the leaders' and the followers' contribution is not significant in treatment C ($p = 0.170$, two-sided Wilcoxon signed-rank test), and in part 3, the difference is not significant in treatment CAC (p=0.182, two-sided Wilcoxon signed-rank test). For the statistical tests, we compare a leader's average contribution with the average contribution of all followers of the same group in a part. We will get back to the exceptions in the next sections. The larger contributions of leaders are confirmed in all regression models in Table 3.3. On average, leaders contribute between 4.5 and 5.1 EMU more than followers with the coefficient estimate being statistically significantly different from zero at the one percent level. This replicates an earlier finding by Güth et al. (2007) among others.

3.3. The reviving effect of communication in part 2

Up to now we have documented that in all treatments there is a restart effect in round 13. Now we ask whether there are treatment differences in the restart as such. The increase in group contributions from round 12 to round 13 is on average (with the corresponding standard deviation) 7.02 EMU (10.8), 18.13 EMU (9.8), 18.81 EMU (15.0), and 12.60 EMU (11.1) in treatment R, CA, C, and CAC, respectively. The increase is significantly larger in the treatments CA and C than in the control treatment R (p = 0.015 and p = 0.038, respectively, two-sided Mann-Whitney U test). For leaders, this is only the case when we compare the contribution increase in treatment CA (19.15 EMU) with the rise in control treatment R (8.73 EMU) (p = 0.074, two-sided Mann-Whitney U test). For followers only, contributions in treatments CA (17.79 EMU) and C (18.92 EMU) rise more than in control treatment R (6.44 EMU) (p = 0.020 and p = 0.043, respectively, two-sided Mann-Whitney U test).

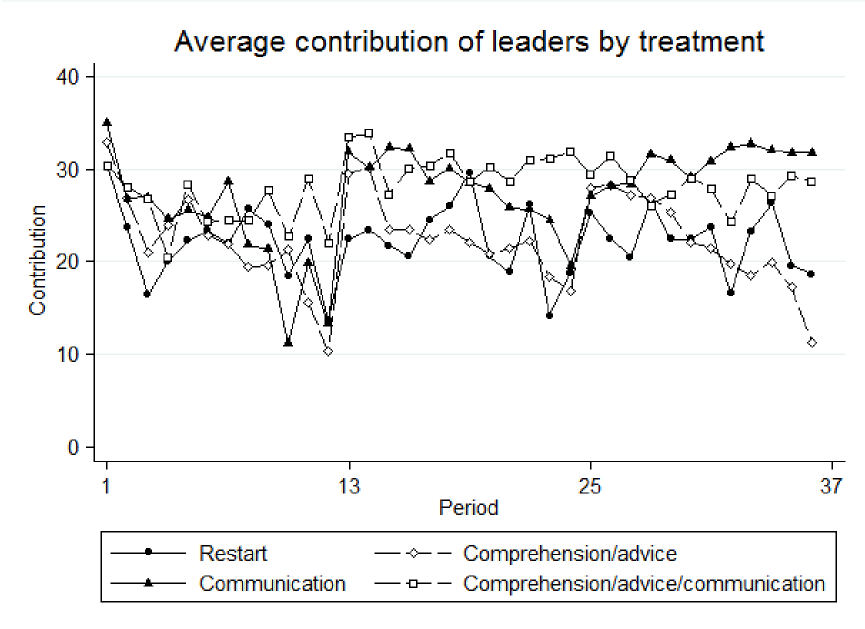
Figure 3.2: Average contributions in control treatment R and treatment CA, C, and CAC (round 1 through 36).



What happens in the rest of part 2 (rounds 13 through 24)? We find that contributions in part 2 are highest when the leader communicates with the followers irrespective of the additional comprehension/advice text, whereas they are similar between the pure restart and the comprehension/advice intervention, see Figure 3.2 and Table 3.2. Average contributions by all participants, leaders, and followers are summarized in Table 3.2 for treatment R, CA, C, and CAC and part 1, 2, and 3 separately. In part 2, there are no significant differences in the distribution of group contributions between control treatment R (18.20 EMU) and treatment CA (18.51 EMU) (p = 0.695, Mann-Whitney U test), nor are there between treatment C (26.56 EMU) and treatment CAC (27.50 EMU) (p = 0.773, Mann-Whitney U test). However, contributions in part 2 are significantly higher than in treatment R and CA if the leader communicates with the followers in either treatment with communication (p<0.045, pair-wise two-sided Mann-Whitney U test).

Separate analyses for leaders and followers draw a similar picture; see also Figure 3.3, Figure 3.4, and Table 3.2. There are no differences between control treatment R and treatment CA/treatment C and treatment CAC for leaders only (p = 0.982/p = 0.339, pair-wise two-sided Mann-Whitney U test) and for followers only (p = 0.730/ p = 0.730, pair-wise two-sided Mann-Whitney U test). For leaders, contributions in treatment CAC in part 2 are larger than in treatment R and CA (0.014 < p < 0.041, pair-wise two-sided Mann-Whitney U test) indicating that leaders try to push contributions in treatment CAC. Leader contributions in treatment C in part 2 are somewhat larger than in treatment R and CA, but not significantly (0.143 < p < 0.211, pair-wise two-sided Mann-Whitney U test). Followers contribute significantly more after receiving a message from their group leader at the beginning of part 2 than in treatments R and CA, independent of the comprehension/advice text (p < 0.039; pair-wise two-sided Mann-Whitney U test).

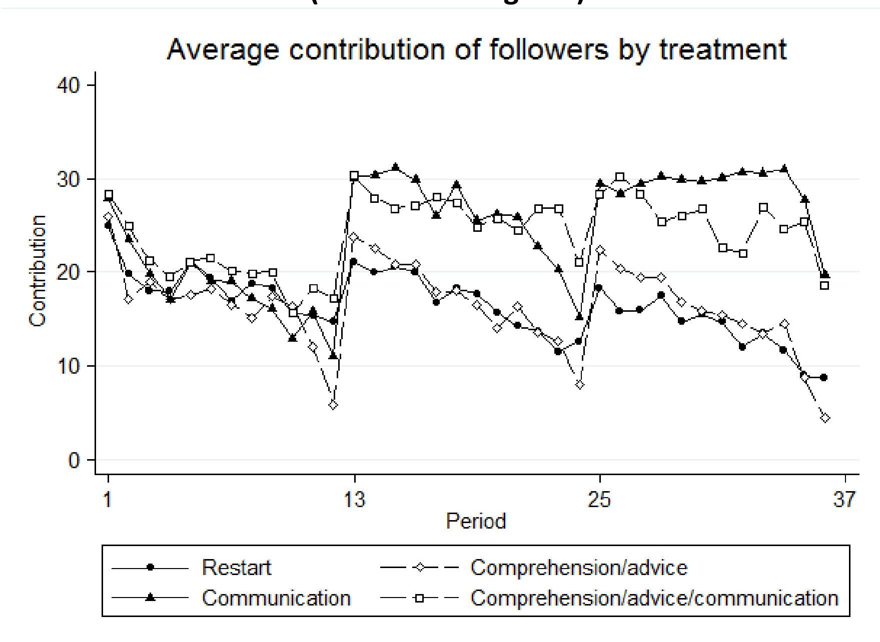
Figure 3.3: Average contributions of leaders in control treatment R and treatment CA, C, and CAC (round 1 through 36).



Another way to look at the long-run effect of the restart and the three interventions on cooperation is to compare part 2-contributions with part 1-contributions within each treatment. We do the analysis again for average contributions (leaders and followers), leaders only, and followers only, see Table 3.2 for the respective average contributions and standard deviations. The rise in cooperation from part 1 to part 2 is only significant with communication; both without (p = 0.050, two-sided Wilcoxon signed-ranks test) and with (p = 0.060, two-sided Wilcoxon signed-ranks test) comprehension/advice. In both treatments, contributions increase by around 35%. The increase in contributions from part 1 to part 2 in treatments C and CAC is also significant for leaders and followers separately (p = 0.071 and p = 0.071 for leaders, respectively; p = 0.050 and p=0.028 for followers, respectively, two-sided Wilcoxon signed-ranks test). All other contribution changes from part 1 to part 2 are not significant, neither on the group level nor for leaders and followers separately (p > 0.256 for each treatment separately, two-sided Wilcoxon signed-ranks test). This is particularly interesting because the informational content and understanding is supposed to be higher with the external explanation and advice (treatment CA) than with communication (treatment C).

Yet another way of analyzing treatment differences is the comparison of the contribution changes between treatments (Diff-in-Diff analysis). Here, the question is whether there is a long-run reaction to a particular treatment controlling for initial contribution levels in part 1. The rise in cooperation from part 1 to part 2 is significantly larger in treatment C and CAC compared to control treatment R (p=0.032 and p=0.017, respectively, two-sided Mann-Whitney U test). The boosting effect of communication compared to the pure comprehension/advice intervention is (slightly) insignificant (p = 0.135 and p = 0.115, respectively, two-sided Mann-Whitney U test). Looking at leaders only, there are no significant differences in the long-run contribution reaction to any of the three interventions or to the pure restart (p > 0.231, pair-wise two-sided Mann-Whitney U test). Leaders' contributions increase however slightly more in treatment CAC than in control treatment R (p=0.107, two-sided Mann-Whitney U test). The change in cooperation is significantly larger among followers who receive a message from the leader compared to the pure restart and compared to the pure comprehension/advice intervention (p = 0.015 for treatment R, p = 0.082 for treatment CA; two-sided Mann-Whitney U test). Adding communication to the comprehension/advice text does not increase the followers' contribution significantly (p=0.157, two-sided Mann-Whitney U test) nor does adding the comprehension/advice text to the communication that followers receive from the leader (p=0.954, two-sided Mann-Whitney U test). There are no significant differences between treatment R and CA among followers (p = 0.461, pair-wise two-sided Mann-Whitney U test).

Figure 3.4: Average contributions of followers in control treatment R and treatment CA, C, and CAC (round 1 through 36).



The long-run cooperation reaction to communication without the comprehension/advice text is particularly strong among followers. Remember that leaders' contributions are in general significantly larger than followers' contributions except for treatment C in part 2 (p = 0.170, two-sided Wilcoxon signed-rank tests), see section 3.1. The average contribution gap is cut to more than half from 4.93 EMU in part 1 to 2.06 EMU in part 2 (Table 3.2) meaning that, with communication, leaders manage to make followers go more after them. Leaders in treatment CAC try equally hard to push cooperation in part 2 by increasing contributions to an average of 30.69 EMU, but they do not manage to convince followers to go after them as much as in treatment C. Unexpectedly, the comprehension/advice text seems to make it harder for leaders to convince followers.

The regression models (2a), (2b), (4a) and (4b) in Table 3.3 confirm the effect of communication beyond the restart effect. The dummy variables for treatment C and CAC are significant at the ten to one percent level and show that contributions in the communication treatments in part 2 are on average 8 EMU (treatment C) and 7-9 EMU (treatment CAC) larger than in the control treatment with the pure restart, see models (2a) and (4b). The coefficient estimates of the dummy variables for the other interventional treatment CA are insignificant. Note that, in model (2b), the coefficient estimates of the three interaction terms are insignificant. This means that cooperation declines over time similarly in part 2. Communication by the leader revives cooperation in part 2 effectively, but does not prevent a similar decline over time as do neither the pure restart nor the external comprehension/advice intervention. Note, however, that the interaction term of treatment CAC and the round variable is positive indicating that the decline with the comprehension/advice intervention in combination with the leader's communication somehow softens the decline in cooperation in part 2.

3.4. The reinforcing effect of repeated communication in part 3

We start the cooperation analysis in part 3 by comparing contributions in round 25 with contributions in round 24 (second restart). The increase from round 24 to 25 can be seen for all four treatments in Figures 3.2, 3.3, and 3.4 and the augmentation is confirmed by non-parametric tests for the three intervention treatments ($p < 0.084$ separately for each treatment and for average contributions, leaders' contributions, and average followers' contributions, two-sided Wilcoxon signed-rank test). In the control treatment R, the increase is not significant for non-parametric tests ($p > 0.132$ separately for average group, leaders', and followers' contributions, two-sided Wilcoxon signed-rank test). Surprisingly, leaders in treatment CAC do not contribute significantly more after the second restart ($p = 0.652$, two-sided Wilcoxon signed-rank test). However, the highly significant positive coefficient estimate of the part 3-dummy in the regression models (4a) and (4b) in Table 3.3 underline an overall increase in contributions at the beginning of part 3 in all four treatments including the restart control treatment.

At the second restart, the comprehension/advice intervention leads to a new short-run restart effect, while communication does not boost cooperation significantly, in contrast to what happened at the first restart. The increase in group contributions from round 24 to round 25 is on average (with the corresponding standard deviation) 5.87 EMU (12.9), 13.54 EMU (12.3), 12.5 EMU (16.9), and 4.8 EMU (7.9) in treatment R, CA, C, and CAC, respectively. The increase is (not) significantly larger in treatment CA (treatments C and CAC) than in the control treatment R ($p = 0.065$ ($p > 0.231$), two-sided Mann-Whitney U test). For leaders, the short-run reaction is significantly smaller in treatment CAC than in treatment CA and C ($p < 0.081$, two-sided Mann-Whitney U test), which is partly due to the fact that contributions in treatment CAC decreased slightly less over part 2. Among followers, contributions in treatment CA (14.36 EMU) rise more than in control treatment R (5.67 EMU) ($p = 0.029$, respectively, two-sided Mann-Whitney U test).

Concerning the effects throughout part 3, the average contributions in part 3 (rounds 25 through 36) are again highest if the leader sends a communication message to the followers, whereas they are very similar with the pure restart and the comprehension/advice intervention, see Figure 3.2. For part 3, there are no significant differences in contributions between the control treatment R (16.03 EMU) and the comprehension/advice intervention (17.12 EMU) nor are there differences between treatment C (29.31 EMU) and CAC (26.13 EMU) ($p = 0.908$ and $p = 0.453$, pair-wise Mann-Whitney U test). With communication, contributions are significantly larger than in control treatment R and treatment CA ($p < 0.009$; pair-wise two-sided Mann-Whitney U test). To a slightly smaller extend the same is true for communication after the comprehension/advice reminder ($p < 0.107$; pair-wise two-sided Mann-Whitney U test). The regression models (3a) and (4b) in Table 3.3 confirm the repeated effect of communication beyond the restart effect. The coefficient estimates of the dummy variable for treatment CAC (and C) are (highly) significant and show that contributions are on average 7.5-10 EMU (13 EMU) higher than in the control treatment with the pure restart. Separate analyses for leaders and followers draw a similar picture, see also Figure 3.3 and Figure 3.4. There are no significant contribution differences in part 3 between the control treatment R and the treatment CA for leaders ($p > 0.963$, two-sided Mann-Whitney U test) and for followers ($p = 0.982$, two-sided Mann-Whitney U test). Contributions of leaders ($p < 0.074$, pair-wise two-sided Mann-Whitney U test) and followers ($p < 0.005$, pair-wise two-sided Mann-Whitney U test) are significantly higher with communication than with pure restart and with comprehension/advice. The leaders' and the followers' contributions in treatment CAC move somewhere in between the contributions in treatment R/CA ($p < 0.200$, pair-wise two-sided Mann-Whitney U test) and treatment C ($p > 0.462$, pair-wise two-sided Mann-Whitney U test).

Comparing part 3 contributions with part 2 contributions within each treatment, we find that contributions decrease in all treatments except for the treatment where the leader communicates with the followers without the comprehension/advice stage, see Table 3.2. The decrease is only significant for the treatment with the pure restart for all participants ($p = 0.094$, two-sided Wilcoxon signed-ranks test) and for followers only ($p = 0.038$, two-sided Wilcoxon signed-ranks test), but not for leaders only. Note that contributions decrease stronger towards the end of the experiment in treatments R, CA, and CAC leading to lower average contributions in part 3, see figures 2, 3, and 4 and Table 3.2. This so-called last round effect is often observed in repeatedly played voluntary contribution games towards the end of the experimental session. On the contrary, contributions in treatment C do not decrease over the rounds of the last part except for the last two rounds. The average contributions with one-way free form communication *increase* from 26.56 EMU in part 2 to 29.31 EMU in part 3. This is particularly surprising because the *repeated* communication at the beginning of part 3 apparently improves cooperation such that it compensates for more than the last round effect. The rise in contributions from part 2 to part 3 is however significant neither for average contributions nor for leaders and followers separately.

Comparing the long-run reaction to the interventions and the restart after the repeated restart and interventions (difference between contributions in part 3 and in part 2), we find significant differences only for treatment C compared to the restart ($p = 0.083$, two-sided Mann-Whitney U test). Leaders who communicate with the followers contribute slightly more than leaders in treatments CA ($p = 0.103$, two-sided Mann-Whitney U test). Followers react significantly more positively to the text message by the leader than to the pure restart ($p=0.054$, two-sided Mann-Whitney U test). The reaction to the repeated communication is not significantly different with the comprehension/advice message than without the external information input, but the combined intervention CAC has no larger repeated impact on cooperation than the repeated restart or the repeater comprehension/advice message does.

The lasting effect of the leaders' (repeated) communication with the followers on cooperation is also confirmed in regression models (3a) and (4b) where the coefficient estimates of the communication dummy and of the interaction term between communication and part 3 are significant at the one percent level, respectively. The repeated communication at the beginning of part 3 does not only maintain the previous reviving effect of the text message by the leader, but reinforces it: compared to the pure restart, contributions in treatment C are on average 8 EMU higher in part 2, model (2a), and 13 EMU higher in part 3, model (3a) in Table 3.3. Looking at the contribution evolution over time in part 3 in model (3b), we find that the interaction term between treatment CA and the round variable is significantly negative with a value of -0.646. Contributions start somewhat higher, but the decay is stronger after two "expert" explanation and advice interventions than after two pure restarts. This could be due to disappointed higher expectations about the others' contributions. The interesting effect of *repeated* communication can be seen in regression model (3b): the coefficient estimate of the interaction term of the treatment C dummy and the part round variable is positive and significant at the ten percent level. Repeated communication prevents the decrease in contributions over time in part 3 to a large extent: the coefficient estimates of the part round variable and of the interaction term are -0.673 and +0.521 in model (3b), respectively. The combination of "expert" explanation and advice and leader communication increases cooperation also in the repeated intervention, but does not perform as well as the pure communication. *Repeated* communication (without the explanation and advice stage) seems to *reinforce* the reviving effect of communication on cooperation. While the leaders' first text message shifts contributions upwards, the leaders' second communication with the followers results not only in the preservation of the contribution shift, but also in a hardly decreasing cooperation slope over time. Fixed effects regressions (robust standard errors) regressing the individual contributions on the round variable for each part and each treatment separately confirm that part 3 in treatment C is the only case where the contribution decay over rounds is not significantly different from zero (regressions nor reported, available upon request). For all other cases, the decay is significantly different from zero on the 1% level (treatments R and CA separately for parts 1, 2, and 3; treatment C for parts 1 and 2; treatment CAC for part 1) and on the 6% level (treatment CAC for parts 2 and 3).

3.5. How followers follow the leader and communication content

The positive effect of communication on contributions is to a large extent related to the following behavior of followers with respect to the group leader's contribution. In Table 3.4, the followers' contributions are regressed on a variable for each part round (taking values between 1 and 12), the group leader's contribution in the same round (taking values between 0 and 40) and three interaction terms of the group leader's contribution and a dummy for the treatments CA, C, and CAC, respectively (taking the value one for the respective treatment and zero otherwise). We use fixed effects regression with robust standard errors and the omitted dummy variable is treatment R. Regression models (1), (2), and (3) use the observations of part 1, 2, and 3, respectively. As shown previously, contributions decrease significantly over the rounds of a part and the group leader's contribution has a significant impact on the followers' contributions (highly significant coefficient estimates for the round and for the group leader's variable in all three models). We want to know whether the impact of the group leaders' contribution on the followers' contributions increases with any of the three treatments compared to the pure restart. The coefficient estimates of the interaction terms are insignificant in part 1; see regression model (1) in Table 3.4. In part 2, the coefficient estimates of the group leader's contribution with the dummy for treatment C and treatment CAC goes up to 0.220 and 0.209, respectively, but it is only significant for the communication treatment. Followers follow the group leader's contributions thus more when the leader communicates compared to the pure restart without any communication. Regression model (3) confirms this trend in part 3 for treatment C only: the coefficient estimate for the interaction term with the communication dummy is 0.387 and significantly different from zero on the 1% level.

Since the communication is free-form, we are interested in what kind of messages the leaders send and whether they differ between treatment C and treatment CAC. We therefore coded the text message that leaders in treatment C and CAC send in rounds 13 and 25 to their followers. The mean values for some summary statistics and code categories are summarized in Table 3.5 separately for rounds 13 and 25 and treatments C and CAC, respectively. The summary statistics refer to the time in seconds that leaders need until they enter the last part of their text message and to the average number of words per text message. For the communication content analysis, we mostly adopted the coding categories from Koukoumelis et al. (2012) and added some categories that we thought would be important for our design.[30]

Table 3.4: Fixed effects regression with robust standard errors (Data: treatments R, CA, C, and CAC, followers only).

	(1)	(2)	(3)
	Dependent variable: Contribution		
VARIABLES	(Part 1)	(Part 2)	(Part 3)
Part round (1-12)	-0.524***	-0.637***	-0.635***
	(0.0946)	(0.0904)	(0.0916)
Contribution group leader (CGL)	0.386***	0.426***	0.366***
	(0.0599)	(0.0640)	(0.0625)
(Comprehension/advice)*(CGL)	-0.0139	-0.00569	0.141
	(0.0932)	(0.0938)	(0.106)
(Communication)*(CGL)	-0.0277	0.220**	0.387***
	(0.0786)	(0.103)	(0.104)
(Comp./advice/communication)*(CGL)	0.0920	0.209	0.00213
	(0.102)	(0.150)	(0.122)
Constant	12.62***	11.57***	11.68***
	(1.229)	(1.321)	(1.192)
Observations	1,872	1,872	1,872
R-squared	0.226	0.304	0.291
Number of subject	156	156	156

Notes. Robust standard errors in parentheses. Dependent variable Contribution takes values between 0 and 40.

*** p<0.01, ** p<0.05, * p<0.1

The code categories are described in the following. The first five coding categories capture the content of the comprehension/advice message in treatment CA (and CAC). The intention is to see whether leaders mention an observed decline in previous contributions, whether they observed followers undercutting in general, whether they mention one or more possible explanation(s) such as selfishness and consequences of such an undercutting behavior, i.e. others may follow the example. Finally, we code a request for conformity, i.e. the leader's emphasis on the need that all group members conform to the leader's contribution.

The next six categories enclose some payoff-related arguments. In particular, they include the leader's suggestion (point or interval) of how much to contribute to the project; the suggestion, implicit or explicit, must be unambiguous. We code whether the suggestion is efficient, i.e. an implicit or explicit suggestion that everybody in the group (including the leader) contributes the whole endowment. Furthermore, coding categories enclose whether the leader makes explicit payoff calculations associated with the proposal, whether he argues explicitly that the suggested amount maximizes the group payoff, or conjectures that participants are interested in maximizing the group payoff, as well as whether the leader mentions explicitly that the followers benefit from following his suggestion. Finally, we code with the last category in the payoff-related group whether the leader mentions a certain

[30] One of the co-authors did the coding of the text messages and did the coding as objectively as possible.

reaction to defecting. In particular, code 1 stands for announcing the tit-for-tat strategy, 2 stands for two-tit-for-tat, 3 for the grim trigger strategy, and 4 for a random/reducing strategy in case of someone defecting.

Table 3.5: Average of coded values for each summary statistic and communication category in treatments C and CAC in rounds 13 and 25

	Round 13		Round 25	
	Treatment C	*Treatment CAC*	*Treatment C*	*Treatment CAC*
Time for message (in sec.)	303.1	264.2	220.8	192.8
Number of words	72.6	55.1	79.4	72.5
Observation of decline (0=no, 1=yes)	**25.0%**	**9.1%**	41.7%	25.0%
Observation of followers undercutting (0=no, 1=yes)	25.0%	18.2%	41.7%	50.0%
Undercutting reasons (e.g. selfishness) (0=no, 1=yes)	16.7%	18.2%	**50.0%**	**8.3%**
Consequences (Future repercussions of actions) (0=no, 1=yes)	16.7%	18.2%	33.3%	33.3%
Conformity (0=no, 1=yes)	58.3%	72.7%	66.7%	75.0%
Suggestion (0=no, 1=yes)	83.3%	90.9%	66.7%	83.3%
Efficient suggestion (0=no, 1=yes)	41.7%	36.4%	33.3%	33.3%
Payoff calculation (0=no, 1=yes)	41.7%	36.4%	25.0%	41.7%
Group payoff maximization (0=no, 1=yes)	66.7%	72.7%	50.0%	66.7%
Satisfaction (e.g. benefit for each) (0=no, 1=yes)	75.0%	81.8%	66.7%	66.7%
Strategy (0=no strategy, 1=tit-for-tat, 2=two-tit-for-tat, 3=grim trigger, 4=random/reduce a bit)	**3 tit-for-tat, 2 grim trigger, 1 random**	**2 tit-for-tat**	1 tit-for-tat, 1 two-tit-for-tat, 2 grim trigger	2 tit-for-tat, 1 grim trigger, 1 random
Fairness (0=no, 1=yes)	**50.0%**	**18.2%**	**58.3%**	**25.0%**
Team spirit (0=no, 1=yes)	33.3%	27.3%	**25.0%**	**50.0%**
Notification of low contributors (0=no, 1=yes)	**33.3%**	**18.2%**	50.0%	58.3%
Praise (0=no, 1=yes)	0.0%	0.0%	**50.0%**	**16.7%**
Complaint (0=no, 1=yes)	**25.0%**	**9.1%**	33.3%	25.0%
Mood (-1=bad, 0=neutral, 1=good)	**0.33**	**0.18**	**0.33**	**0.17**
Leave contribution decision to followers (0=no, 1=yes)	**16.7%**	**0.0%**	**25.0%**	**0.0%**
Promise (0=no, 1=yes)	25.0%	18.2%	25.0%	16.7%
Willingness to contribute more than followers (0=no, 1=yes)	**25.0%**	**0.0%**	**16.7%**	**0.0%**
Form (-1=informal, 0=neutral, 1=formal)	0.08	0.09	0.08	0.08
Labor notion (0=no, 1=yes)	33.3%	36.4%	**16.7%**	**41.7%**
Strange/nonsense (0=no, 1=yes)	**25.0%**	**9.1%**	**8.3%**	**25.0%**

Notes. The number of analyzed text messages in round 13 in treatment CAC is 11 (due to technical problems, the message of one leader was not saved). In all other cases 12 text messages were analyzed, respectively. The bold value pairs show a considerable difference in communication between the two treatments (circa 50% or more).

The third group of coding categories encompasses social preferences, emotional expression, and own contribution behavior. With fairness, we refer to an explicit or implicit reference to fairness or just behavior, which also includes an explicit rejection of some group member contributing less than the others. Team spirit refers to a statement promoting the willingness to cooperate as part of a team or emphasizing the importance of cooperation in the group. Closely related is the notification of low contributors, implicit or explicit, of those who contributed less than suggested or who started decreasing their contributions. The difference to the category "Observation of followers undercutting" is that here the leader refers to the followers as a whole. We furthermore code whether the leader praises or complains about observed contributions. The mood of the communication is (mostly) independent from the leader's praise or complaint and gives an overall impression of bad, neutral, or positive vibes, which includes the use of "smileys," or other forms of creating a good atmosphere. Furthermore, we code whether the leader leaves the contribution decision explicitly to the followers, promises to contribute some specific amount, or expresses the willingness to contribute more than the followers do.

The last group includes a set of different coding categories. We code whether the form of the text message is rather informal, neutral, or formal, whether the leader uses the labor notion from the instructions, e.g. "director," "workers," or "firm," and whether the communication content is to some extend strange, wrong or does not make any sense. The number of analyzed text messages in round 13 in treatment CAC is 11 (due to technical problems, the message of one leader was not saved). In all other cases 12 text messages were analyzed, respectively.

The comprehension/advice categories are mentioned frequently in treatment C in round 25. Whereas each of the five categories is mentioned between 17% and 58% of the times in round 13 in treatment C, the respective frequencies go up to 33% to 67% of the times in round 25. The communication content in treatment C is thus partly quite similar to the content of the expert explanation and recommendation in treatment CAC.

Besides a request for conformity in contributions, a suggestion and in particular an efficient suggestion from the leader is an important determinant of successful cooperation. Note that in the comprehension/advice communication text, we only recommend conformity to reach high earnings. In both treatments C and CAC, 83% to 91% (67% to 83%) of the leaders make a contribution suggestion in round 13 (25), but only 36% to 42% of the leaders suggest the efficient contribution to the public good in round 13, i.e. everybody contributing the entire endowment. The monetary benefit of cooperating is however stressed by almost all leaders (group payoff maximization and satisfaction).

For some communication categories, we find considerable differences between the communication only and the comprehension/advice *and* communication treatments. In particular for the formulation of a "punishment" strategy for deviations from cooperating, for the notion of fairness, and the expression of emotions, we find interesting differences between treatment C and treatment CAC. When there is no explanation and advice from outside, leaders propose more often less forgiving strategies, in particular when communicating for the first time in round 13. Three leaders announce the tit-for-tat, two leaders the grim trigger, and one leader the random strategy in treatment C compared to two tit-for-tat announcements in treatment CAC in round 13. The gap in tough strategies becomes closer in round 25 but remains (one tit-for-tat, one two-tit-for-tat, two grim trigger in treatment C; two tit-for-tat, one grim trigger and one random strategy in treatment CAC). With communication only, leaders refer more often to fairness reasons (50% and 58% of the cases in rounds 13 and 25, respectively) compared to treatment CAC (18% in round 13 and 25% in round 25).

Overall, the expression of emotions in form of complaint or praise is more frequent without the expert analysis and advice. It is notable that half of the leaders in treatment C praise the observed contributions in round 25 whereas only 17% of the leaders in treatment CAC do so even though the contributions in part 2 are similar in both treatments. This might be the result of a surprisingly positive contribution behavior in treatment C or leaders feeling more responsible for the motivation in treatment C. Even though, in round 13, leaders complain more often about the followers previous contributions with communication only (25% compared to 9% in treatment CAC), the overall mood in the text messages is more positive in both rounds. Compared to none of the leaders in treatment CAC, some leaders leave the contribution choice explicitly to the followers (17% in round 13 and 25% in round 25) or express the willingness to contribute more than the followers (25% in round 13 and 17% in round 25) when they are not influenced by the expert analysis and advice.

4. Conclusion

Cooperation in teams is an important component of successful performance of companies and other organizations. It is a common observation that cooperation decreases over time due to different reasons and we analyzed in this study how a group leader can revive cooperation effectively. A pure restart has been shown to be a powerful mechanism in voluntary contribution games, but only has a short-run effect. We analyzed and compared two wide-spread managerial strategies and a combination of both strategies that organizations may realistically have available and that could create an effect beyond that of a restart.

The results show that leader communication with the followers is by far the most effective intervention for increasing cooperation in the long-run. The effect on cooperation is significantly larger than the effect of a simple restart driven mainly by an increased contribution of followers, i.e. followers pursue the leader better. The effect is also larger compared to an external expert explanation and advice even though the informational content and understanding of the decline in cooperation and a counter-action is supposed to be at least as high with the expert intervention. Also beliefs about the others' contributions are expected to increase with the expert explanation and advice.

A combination of the expert explanation and advice together with the leaders' communication with the followers increases cooperation, but does not outperform the pure effect of communication on cooperation. In addition, *repeated* communication *reinforces* the reviving effect of communication on cooperation. After the leader sends a second text message to the followers, contributions go hardly down over time and significantly less than with the pure restart. Repeated communication after the comprehension/advice intervention does not have a similar reinforcing effect, but maintains high contribution levels.

It is important to note that we do not show that communication by the leader alone has a reviving impact on cooperation. An important feature of our design is the sequential form of the voluntary contribution game, or in other words, the leading-by-example structure meaning that leaders choose the contribution first. Cooperation revival in our design is a combination of the leader's communication with the followers and an exemplary contribution behavior by the leader. We cannot say whether leaders may manage to revive cooperation with communication only nor can we claim that it is the *leader's* communication that matters.

The expert consultancy does not show an effect that goes significantly beyond that of a restart in our experiment nor does it improve the effect of the leader's communication with the followers. We believe that these negative results are as important as the positive one mentioned above. A priori, changing the leader and providing members of failing organizations with analysis and advice would seem as promising ways to strongly revive cooperation. What our results show is that the effect is short-lived and that even the short-run effect does not go beyond that of a pure restart. It is perhaps most surprising that the comprehension/advice treatment has no additional effect, since it would seem that an analysis of the causes of cooperation decline and a clearly formulated advice are the best starting point for not running into the same problem as before. One explanation for this finding may be that not *production oriented communication* matters for cooperation, as it is the case in the comprehension/advice intervention, but *people oriented communication*. Another explanation could be that the communication by the leader is less *formal* than the written communication in the comprehension/advice treatment. Pinto and Pinto (1990) show that high cooperating hospital project teams use informal communication more frequent than low cooperating teams, but there are no differences concerning formal communication. In a similar vein, one could think about the formal, production oriented expert analysis and advice from an external human resource consulting firm as a way to create a short-run restart in the firm. The restart effect should however not be confound with the effect of the intervention itself. Whether the external expert advice has an effect beyond the restart may depend on the content of the analysis and the advice and the communication form.

As to the content of the communication from leader to followers, we do not have enough observations to do a thorough analysis (nor is it the purpose in this study). However, the most commonly mentioned categories are the monetary benefit from cooperation and requesting conditional contribution. Some leaders also threaten to decrease contribution if the followers do not cooperate at the same level, create a feeling of relationship closeness and mention the previous decrease in cooperation and possible reasons thereof. The communication content is thus partly quite similar to the external "expert" explanation and advice we give to the participants adding a personal nuance. It could make a big difference whether the information is transmitted from within the group or from outside the group. Also, the content of the "expert" explanation and advice is purely informative (production oriented) while the leaders can evoke feelings and emotions such as identity, solidarity, or guilt for letting others down (people oriented). Another possibility could be that too much information is not good for changing individuals' behavior. Also the leader can target the previous cooperation in the own group with the free form communication, while the comprehension/advice text is a general statement. It would be interesting to analyze in future work what kind of communication leaders can use to restore cooperation in organizations.

5. References

Andreoni, J. (1988), Why free ride? Strategies and Learning in Public Goods Experiments, *Journal of Public Economics*, 37, 291-304.

Andreoni, J. (1995), Cooperation in Public-Goods Experiments: Kindness or Confusion?, *American Economic Review*, 85(4), 891–904.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, J. Roberts (2013), Does management matter? Evidence from India, *Quarterly Journal of Economics*, Vol. 128(1), 1-51.

Brandts, J. and D. Cooper (2007), It's What You Say, Not What You Pay: An Experimental Study of Manager-Employee Relationships in Overcoming Coordination Failure, *Journal of the European Economic Association* 5, 1223-1268.

Bruhn, M. and B. Zia (2011), Stimulating Managerial Capital in Emerging Markets: The Impact of Business and Financial Literacy for Young Entrepreneurs, *World Bank Working Paper 5642*.

Bruhn, M., D. Karlan, and A. Schoar (2012), The Impact of Offering Consulting Services to Small and Medium Enterprises, *Mimeo World Bank*.

Burlando, R. and J.D. Hey (1997), Do Anglo-Saxons Free-Ride More?, *Journal of Public Economics*, 64(1), 41–60.

Carpenter, J.P. (2007), Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods, *Games and Economic Behavior*, 60(1), 31–51.

Chaudhuri, A., S. Graziano and P. Maitra (2006), Social Learning and Norms in a Public Goods Experiment with Inter-Generational Advice, *Review of Economic Studie*s, 73, 357-380.

Croson, R.T.A. (1996), Partners and strangers revisited, *Economics Letters*, Vol. 53, 25-32.

Croson, R., E. Fatas, and T. Neugebauer (2005), Reciprocity, Matching and Conditional Cooperation in Two Public Goods Games, *Economics Letters*, 87(1), 95–101.

Davis, D and C. Holt (1994), *Experimental economics*, Princeton University Press, Princeton, NJ.

Drexler, A., G. Fischer, and A. Schoar (2010), Financial Literacy Training and Rule of Thumbs: Evidence from a Field Experiment, MIT, *Working paper*.

Egas, M., and A. Riedl (2008), The Economics of Altruistic Punishment and the Maintenance of Cooperation, *Proceedings of the Royal Society B – Biological Sciences*, 275(1637), 871–78.

Fehr, E. and S. Gächter (2000), Cooperation and Punishment in Public Goods Experiments, *American Economic Review*, 90(4), 980–94.

Fischbacher, U. (2007), z-Tree: Zurich Toolbox for Ready-made Economic Experiments, *Experimental Economics*, 10(2), 171-178.

Fischbacher, U. and S. Gächter (2010), Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Good Experiments, *American Economic Review*, Vol. 100, 541-556.

Fischbacher, U., S. Gächter and E. Fehr (2001), Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, *Economics Letters*, 71(3), 397–404.

Gächter, S. and E. Fehr (1999), Collective Action as a Social Exchange, *Journal of Economic Behavior and Organization*, 39(4), 341–69.

Gächter, S., E. Renner and M. Sefton (2008), The Long-Run Benefits of Punishment, *Science*, 322(5907), 1510.

Gächter, S., D. Nosenzo, E. Renner and M. Sefton (2010), Who makes a good leader? Cooperativeness, optimism, and leading-by-example, *Economic Inquiry*, 49: no. doi: 10.1111/j.1465-7295.2010.00295.x.

Greiner, B. (2004), The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics, University of Cologne, *Working Paper Series in Economics*, 10.

Güth, W., M. Levati, M. Sutter and E. van der Heijden (2007), Leading by example with and without exclusion power in voluntary contribution experiments, *Journal of Public Economics*, 91, 1023–1042.

Herrmann, B., C. Thöni and S. Gächter (2008), Antisocial Punishment across Societies, *Science*, 319(5868), 1362–67.

Isaac, R.M., J.M. Walker, and S.H. Thomas (1984), Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations, *Public Choice*, 43(2), 113–49.

Karlan, D. and M. Valdivia (2011), Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions, *Review of Economics and Statistics*, 93, no. 2, 510–527.

Karlan, D., R. Knight, and C. Udry (2012), Hoping to Win, Expected to Lose: Theory and Lessons on Microenterprise Development, *NBER working paper 18325*.

Keser, C. and F. van Winden (2000), Conditional Cooperation and Voluntary Contributions to Public Goods, *Scandinavian Journal of Economics*, 102(1), 23–39.

Koukoumelis, A., M. V. Levati, and J. Weisser (2012), Leading by words: A voluntary contribution experiment with one-way communication, *Journal of Economic Behavior and Organization*, 81(2), 379-390.

Laury, S.K., J.M. Walker and A.W. Williams (1995), Anonymity and the Voluntary Provision of Public Goods, *Journal of Economic Behavior and Organization*, 27(3), 365–80.

Ledyard, J. (1995), Public goods: A survey of experimental research, in: Kagel, J. and A. Roth (eds.), *The handbook of experimental economics*, Princeton University Press, Princeton, NJ.

Masclet, D., C. Noussair, S. Tucker and M.-C. Villeval (2003), Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism, *American Economic Review*, 93(1), 366–80.

McDonald, M. L. and J. D. Westphal (2003), Getting by with the advice of their friends: CEOs' advice networks and firms' strategic responses to poor performance, *Administrative Science Quarterly*, 48, 1-32.

McGuire, W. J. (1985). Attitudes and attitude change. In *Handbook of Social Psychology*, ed. G. Lindzey, E. Aronson, 2:233-346. New York: Random House. 3rd ed.

Neugebauer, T., J. Perote, U. Schmidt and M. Loos (2009), Selfish-Biased Conditional Cooperation: On the Decline of Contributions in Repeated Public Goods Experiments, *Journal of Economic Psychology*, 30(1), 52–60.

Nikiforakis, N. and H.-T. Normann (2008), A Comparative Statics Analysis of Punishment in Public-Good Experiments, *Experimental Economics*, 11(4), 358–69.

Ockenfels, A. and J. Weimann (1999), Types and Patterns: An Experimental East-West-German Comparison of Cooperation and Solidarity, *Journal of Public Economics*, 71(2), 275–87.

Olson, J.M. and M.P. Zanna (1993), Attitudes and attitude change, *Annual Review of Psychology* 44, 117-54.

Park, E.-S. (2000), Warm-Glow Versus Cold-Prickle: A Further Experimental Study of Framing Effects on Free-Riding, *Journal of Economic Behavior and Organization*, 43(4), 405–21.

Pinto, M.B. and J.K. Pinto (1990), Project team communication and cross-functional cooperation in new program development, *Journal of Product Innovation Management*, 7(3), 200-212.

Potters, J., M. Sefton and L. Vesterlund (2007). Leading-by-Example and Signaling in Voluntary Contribution Games: An Experimental Study, *Economic Theory*, Vol. 33, 169–82.

Rivas, M. F., and M. Sutter (2009), The Dos and Don'ts of Leadership in Sequential Public Goods Experiments, University of Innsbruck, *Working Papers in Economics and Statistics*, No. 2008-25.

Sefton, M., R. Shupp and J.M. Walker (2007), The Effect of Rewards and Sanctions in Provision of Public Goods, *Economic Inquiry*, 45(4), 671–90.

Sonnemans, J., A. Schram and T. Offerman (1999), Strategic Behavior in Public Good Games: When Partners Drift Apart, *Economics Letters*, 62(1), 35–41.

Weimann, J. (1994), Individual Behavior in a Free Riding Experiment, *Journal of Public Economics*, 54(2), 185–200.

6. Appendices

Appendix A. Instructions

A.1. Instructions at the beginning of the experiment

General information

Thank you for coming to the experiment. You will receive 5 Euro for the participation in the experiment. You will be assigned to a group and depending on your and your group members' decisions you can earn additional money during the experiment. It is important that you do not talk to any of the other participants until the experiment is over. You can ask questions at any time. If you have a question, please raise your hand and one of us will come to your place to answer.

Role and group matching

You will be randomly assigned to one of two roles: (1) director or (2) employee. This role will be the same throughout the entire experiment.
Participants will be randomly split in groups with 4 members, each composed by 1 director and 3 employees. At no time during the experiment you will know whom you are matched with and your decisions will be anonymous.

Task and stages of each of the 36 rounds

There will be 36 separate rounds. In each round, each group works on a joint project whose payoff will depend on the hours dedicated by all group members. In each round, every participant has an endowment of 40 hours and decides how many of the 40 hours to dedicate to the project. The remaining hours will be automatically dedicated to a private activity.

Each round is independent from the others and develops in the following way:

Stage 1:
Directors: The director of each group decides how many of the 40 hours to dedicate to the project. The rest will be automatically dedicated to the private activity. There will be a simulation area on the lower part of the screen where directors can calculate earnings choosing different hours dedicated to the project by themselves and by the other group members on average (see "Decision screen director"). The calculations are absolutely private. In the upper part of the screen, directors enter the hours that they want to dedicate to the project in the corresponding round.
Employees: The employees do not have anything to do in this stage and wait until the director of their group have taken a decision.

Stage 2:
Directors: The directors do not have anything to do in this stage and wait until the employees of their group have taken a decision.
Employees: The employees of each group are informed about the hours that the director of their group decided to dedicate to the project and decide how many of their own 40 hours to dedicate to the project. The rest will be automatically dedicated to the private activity. There will be a simulation area on the lower part of the screen where employees can calculate earnings choosing different hours dedicated to the project by themselves and by the other group members on average (see "Decision screen employee"). The calculations are absolutely private. In the upper part of the screen, employees enter the hours that they want to dedicate to the project in the corresponding round.

Stage 3:
Directors and employees: All participants are informed about the average hours dedicated to the project by the other group members, the sum of hours dedicated to the project by all group members and about their own earnings. Summaries of previous rounds will also be listed.

After stage 3, a new round starts which develops in the same way.

Additional information

The experiment is split in 3 parts and each part consists of 12 rounds. The specific instructions for each part will be shown on the screen before the corresponding part starts.

Payoff

Your earnings in Experimental Currency Units (ECU) for each round are given by the following function, which is the same for directors and employees:

$$Earnings_{Round} = \underbrace{(40 - Hours_{Project})}_{\text{Earnings from private activity}} + \underbrace{0.5 \cdot \sum\nolimits_{Group} Hours_{Project}}_{\text{Earnings from joint project}}$$

The earnings in ECU are composed by the earnings from the *hours* dedicated to the *private activity* by that person and the earnings from the *sum of hours* dedicated by *all group members* to the *joint project*. That means that each hour that you decide to dedicate to the project gives *each* of the group members (i.e. you and all other group members) earnings of 0.5 ECU. Analogously, each hour that another group member decides to dedicate to the project gives *each* of the group members (i.e. you and all other group members) earnings of 0.5 ECU. Each hour that you decide *not* to dedicate to the project (i.e. to dedicate to the private activity) gives you and only you earnings of 1 ECU.

150 ECU are worth 1.00 Euro. At the end of the session you will receive 5 Euro plus the sum of what you will have earned in all 36 rounds of the experiment. After the experiment finishes we will pay you the earnings in private.

Example and test question

So that everyone understands how decisions translate into earnings we provide an example and a test question. (The number of hours used for the example and test are simply for illustrative purposes. In the experiment the allocations will depend on the actual decisions of the participants.)

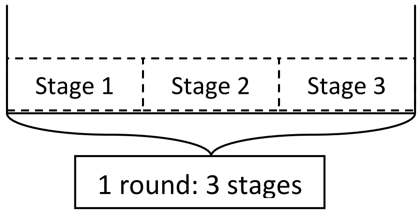
Example: Suppose that you decide to dedicate 31 hours to the project and the other group members decide to dedicate on average 33 hours to the project in one of the 36 rounds.

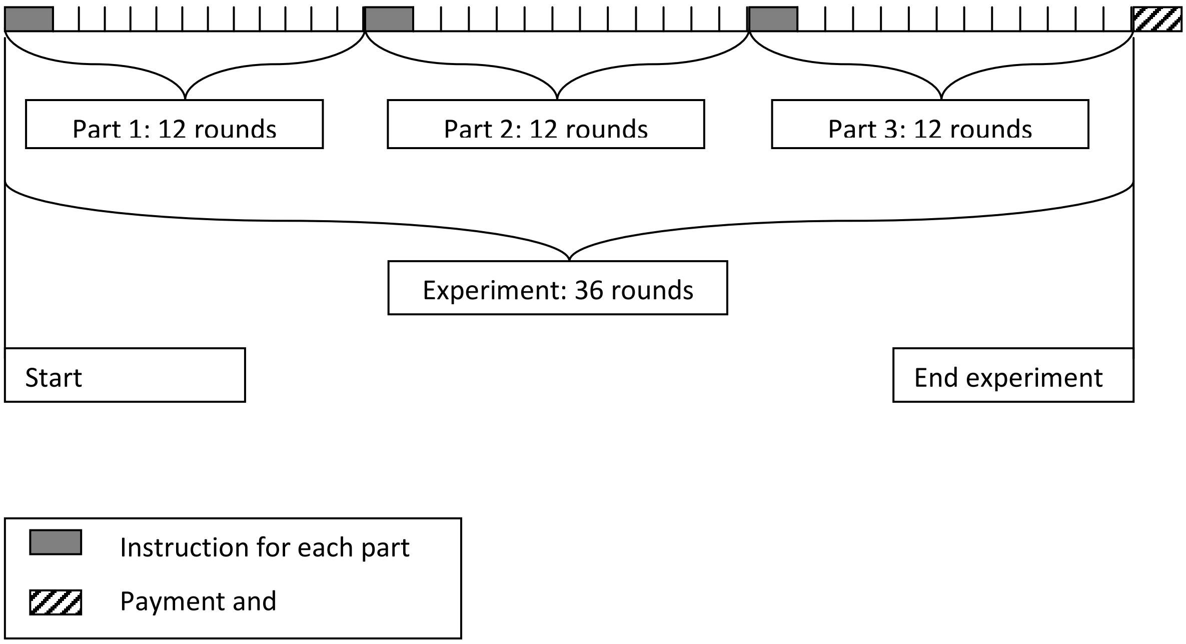
The sum of hours dedicated to the project by all group members is:
31 + 3*33 = 31 + 99 = 130 (hours)
Your earnings in that round are:
(40 – 31) + 0.5*130 = 9 + 65 = 74 (ECU)

Test: Suppose that you decide to dedicate 28 hours to the project and the other group members decide to dedicate on average 24 hours to the project in another of the 36 rounds.

The sum of hours dedicated to the project by all group members is:
____
Your earnings in that round are:
____

Graphical representation of the chronological order of the experiment





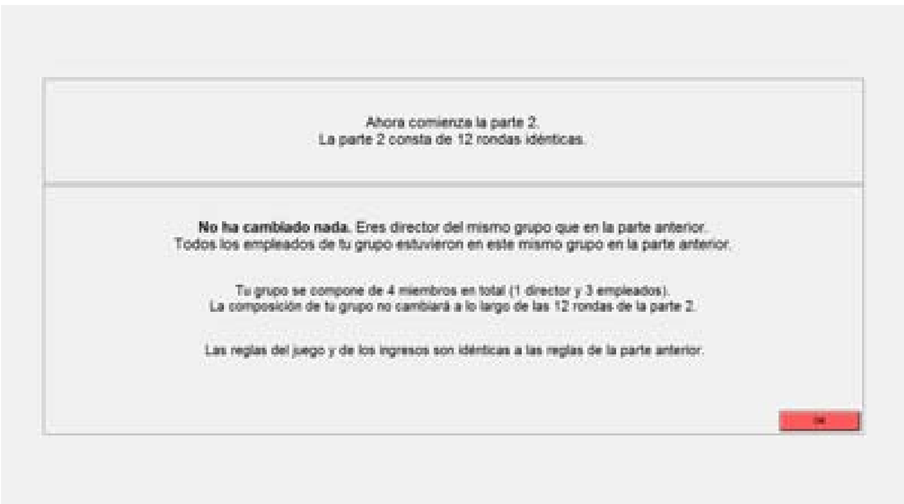
Screenshots

Decision screen director

Decision screen employee
(The number of hours used for the example and test are simply for illustrative purposes. In the experiment the allocations will depend on the actual decisions of the participants.)

A.2. Instructions at the beginning of part 1 (all four treatments R, CA, C, and CAC)

A.3. Instructions at the beginning of part 2 (treatments R, CA, C, and CAC)



A.5. Additional instructions at the beginning of parts 2 and 3 (treatments CA & CAC)

<u>Text at the beginning of part 2</u>

Please read the following text carefully. It gives you some explanation about the game that you are playing in this experiment and some advice.

We observed in previous sessions of this experiment in which you are participating today that the hours dedicated to the common project decrease on average over rounds in this part. You also might have observed that the hours dedicated to the common project in your group decreased over the previous 12 rounds.

We were wondering why contributions decrease and realized that the director's and the workers' hours dedicated to the common project follow similar patterns. That means that directors react to the workers' previous contributions and workers on their turn react to the other workers' and the director's previous contributions.
Workers contribute on average fewer hours to the common project than the other workers of the same group in the previous round and less hours than the director in the same round.

Even though the directors dedicate on average more hours to the common project than the workers in the previous round, they also tend to decrease their contributions compared to the previous round. Therefore, the hours of the directors also decrease over time.
You might have observed this contribution behavior in your group.

A recent study of an experiment similar to ours analyzes more in detail the behavior of the workers only in the experiment (if you want, we can provide you with the reference of the study at the end of the experiment). In that study, the workers are not only asked about how much to contribute, but also about what they believe the other workers will contribute. The study concludes, that "contributions decline because, on average, people […] match others' contributions only partly." That means that, on average, the workers are willing to contribute slightly less than what they believe the other workers will contribute. This leads to contributions being initially lower than expected. Once workers see this the beliefs about the others' contributions will be lower than before. Since the workers contribute on average slightly less than what they believe that the others contribute, this reinforces the process by which average contributions decrease over rounds.
In other words, if the workers start with the idea of undercutting others then others will follow and the contributions to the common project will fall over time.

If you wish to reach and maintain a high earnings level from the common project it is recommendable that all workers dedicate at least the same number of hours to the common project as the director of the group does.

If you have a question, raise your hand and someone of us will come to your place to answer the question.

<u>Text at the beginning of part 3</u>

We do not know how hours dedicated to the common project evolved in your group over the previous part. However, we would like to remind you of the explanation for the decline of contributions to the common project over time and the advice that we gave you previously:

We observed in previous sessions of this experiment that the director's and the workers' hours dedicated to the common project follow similar patterns. Workers contribute on average fewer hours to the common project than the other workers of the same group in the previous round and less hours than the director in the same round. Even though the directors dedicate on average more hours to the common project than the workers in the previous round, they also tend to decrease their contributions compared to the previous round. Therefore, the hours of the directors also decrease over time.

A recent study of an experiment similar to ours concludes that, on average, workers are willing to contribute slightly less than what they believe the other workers will contribute. If the workers start with the idea of undercutting others, this will lead to the decrease of contributions over time.

If you wish to reach and maintain a high earnings level from the common project it is recommendable that all workers dedicate at least the same number of hours to the common project as the director of the group does.

If you have a question, raise your hand and someone of us will come to your table to answer the question.

A.6. Additional instructions at the beginning of part 2 and part 3 (treatment C and CAC after having received the comprehension and advice text)

In the box on their screen, the directors have *now/again* the opportunity to write a message, which will be sent to the employees of their group. After entering the message, you - the director - need to press the Enter key. The written text will appear in the upper part of the box the way it will be sent to the employees and you won't be able to change the entered text once you press the Enter key (just like in chats in Skype or WhatsApp). When you have finished writing the text and are ready to send the message to the employees you may raise your hand and one of us will come to your table to give you the code to get to the next screen. The employees will receive the message of the director of their group and, after that, the *second/third* part of the experiment (rounds *13/25* through *24/36*) will start.

You – the director - are free to send the message you like, including what you think is the best approach to the experiment, what you plan to do, and/or what you would like the others to do and/or why. However, there are two restrictions on the kind of messages that you can send:

1. First, you are not allowed to identify yourself to the others. Thus, you cannot reveal your real name, nicknames, or any other identifying feature such as gender, hair, or where you are seated.
2. Second, there must be neither threats nor promises pertaining to anything that is to occur after the experiment.

The minimum entry of characters is 10. Please, try to finish your message within seven minutes. The remaining time in seconds is shown on the upper right corner of the screen.

If you have a question, raise your hand and someone of us will come to your table to answer the question.

CHAPTER IV: SPOKESPERSON – TO BE OR NOT TO BE?

1. Introduction

Dating back to Sophocles, Plutarch and Shakespeare, the communication of bad news has been considered an unpleasant task. More recently, Bies (2010) shows that business leaders classify the communication of bad news such as corporate downsizing and employee layoffs as one of the most difficult activities. Leaders (have to) make such unpleasant or unfair decisions at times and a possible way to avoid the emotional distress coming along with the communication of bad news (Tesser and Rosen 1975; Folger and Skarlicki 2001) is to delegate the communication to another person. For instance, the *New York Times* reports that companies appoint external consultants to communicate firing decisions to the formers' own workers.[31] In the political arena, the press documents that key politicians avoid communicating unpopular decisions to the media themselves and instead delegate this activity.[32] However, at this point, it is not clear how delegating communication to a spokesperson affects the decision-making process and how individuals affected by such a decision react to the delegation of communication.

Existing studies analyze separately why people delegate decision rights (Oexl and Grossman 2013; Bartling and Fischbacher 2012; Coffman 2011; Hamman, Loewenstein, and Weber 2010; Bolton and Dewatripont 2005; Fershtman and Gneezy 2001) and how communication affects fairness (Andreoni and Rao 2011).[33] Some work suggests that delegation helps shift the responsibility of unkind actions (Oexl and Grossman 2013; Bartling and Fischbacher 2012; Coffman 2011; Hamman et al. 2010).[34] Focusing on delegated communication, our work extends this research by testing the possibility to achieve potential shifts of blame and attention while at the same time not giving away the decision rights. Thus, our setting is particularly important when the decision rights are valuable and the decision-maker wants to keep full control over them.

Bridging these strands of research, our paper analyzes *delegated communication*, i.e. a situation in which a decision-maker takes an action affecting another person and can then communicate this action to the affected person using a spokesperson, i.e. someone who did not make any allocation decision and whose only task is to communicate the decision previously made by the decision-maker. Specifically, we use a laboratory experiment to analyze: (1) how the possibility of delegating the communication of a decision influences the decision-maker's fairness, and (2) whether the persons affected by this decision react

[31] In "Letting a stranger do the firing", *New York Times* (November 10th, 2007). For a similar example, see also "Meet Rebecca. She's here to fire you", *Inc. Magazine* (November 1st, 2007).

[32] On the Spanish context, see for instance "Comparecencias de Mariano Rajoy", *20minutos.es*; "Los silencios de Rajoy," comment by Miguel Ángel Aguilar at *elpais.com* (January 10th, 2012); "Rajoy sigue escondido y tampoco explica el 'caso Bárcenas' tras el Consejo de Ministros", *Publico.es* (February 1st, 2013); "Los cinco silencios más clamorosos de Rajoy", *Eldiario.es* (July 11th , 2013).

[33] Fershtman and Gneezy (2001) suggest that the proposer's payoffs in the ultimatum game are significantly higher if she delegates to an agent who has incentives to make harsh offers. Another related paper is Erat (2013), who shows that a significant fraction of people uses an agent to lie. Moreover, the likelihood of delegating to an agent depends on the harm degree; people delegate more when the lie is particularly harmful to the receiver. Charness, Cobo-Reyes, Jiménez, Lacomba, and Lagos (2012) study the provision of effort by workers when the wage choice is delegated.

[34] Using firm-level data, a different strand of research shows that CEOs are more likely to delegate decision rights when the firm is large and the CEO is overloaded (Graham, Harvey, and Puri 2013), whereas the decisions delegated are often about workforce and capital spending (Colombo and Delmastro 2004). This research exclusively focuses on delegation of decision rights and does not consider the possibility that companies may outsource just the communication of the decisions made. A growing number of firms outsource communication while keeping the decision-making inside the company, see opening examples.

differently depending on whether the decision is communicated to them by the decision-maker or by a spokesperson. By investigating delegated communication, our paper contributes to the debate on the best practices to make a firm's decision-making and communication strategies more effective (Bies 2010). To our knowledge, this study is the first step towards analyzing the separation of making and communicating a decision in a laboratory experiment. Our setting allows for a variety of extensions to study the delegation of communication more in depth such as face-to-face treatments, costly delegation or variations in the reaction forms of the persons affected by the decision.

In our experiment, participants play a modified version of the dictator game. Similar to Bartling and Fischbacher (2012), participants play in groups of four players. In each group, there are three different roles: decision-maker (one player), spokesperson (one player), and receiver (two players).[35] Each decision-maker is asked to allocate a monetary endowment: she can choose a *fair* allocation (same amount to each of the four players), or an *unfair* allocation (substantially larger amount to herself and the spokesperson). In the next step, either the decision-maker or the spokesperson communicates the allocation chosen by the decision-maker to the receivers. Employing a spokesperson is not costly to the decision-maker. Next, a message informs receivers about who, the decision-maker or the spokesperson, communicates the decision-maker's allocation choice, and the allocation decision is communicated accordingly. Finally, receivers can punish by reducing the earnings of the decision-maker and of the spokesperson. Punishment is costly to receivers.

We run a total of six treatments that vary in two dimensions: we adopt three different communication forms, each of which is implemented under two different punishment forms. Let us start with the three communication forms. In the main treatment, the decision-maker can *choose* whether to communicate personally the decision made or delegate the communication to a spokesperson (*voluntary delegation)*. In the control treatments, either the decision-maker is obliged to communicate personally the decision made (*no delegation)* or the spokesperson has to communicate the decision previously made by the decision-maker (*mandatory delegation)*. We run each treatment varying the punishment structure.

To better understand how receivers perceive and punish an unfair allocation, we adopt two alternative punishment structures for each communication form, which allow us to isolate how the perception of shifting blame by delegating communication and deciding on the communication content affects decisions. In the first case (*aligned punishment*), receivers who decide to punish shall do so by using the same monetary amount of punishment for the decision-maker and the spokesperson. By hurting one in reaction to a decision, e.g. by providing less effort in reaction to a wage cut in a firm lead by two partners, the other one is hurt, too. In the second case (*independent punishment*), receivers are free to punish both the decision-maker and the spokesperson (or only one of them) and to assign the same or different sizes of punishment. Here, one person can be hurt, e.g. by lowering the president's or the vice-president's valuation in public opinion polls, without necessarily hurting the other one.

Our results show how important the context is in which the decision-maker employs the spokesperson. While 72% of the decision-makers choose the unfair monetary allocation with aligned punishment, only 49% of the decision-makers do so under the independent punishment and delegation voluntary. In the aligned punishment, the decision-maker and the spokesperson assume the same consequences for any decision that either of them makes, whereas receivers can target their punishment with the independent punishment which opens room for shifting blame that might be perceived negatively. Additionally, the decision-maker

[35] We employed two players C, as in Bartling and Fischbacher (2012), in order to have more observations in each subgroup.

may count on receivers feeling guilty about punishing the spokesperson since she is not responsible for the unfair decision.

We find that, compared to the fair choice where about 20% of decision-makers delegate the communication to the spokesperson, the share more than doubles to over 50% of decision-makers among those who choose the unfair allocation. This indicates that decision-makers avoid the emotional distress and feeling of shame when communicating the unpleasant decision and/or try to shift attention towards the spokesperson with the communication delegation.

The importance of the perception of blame shifting by delegating communication is underlined by the communication strategy that decision-makers and spokespersons choose. While decision-makers explain the unfair allocation primarily with "needs," spokespersons often express "remorse" for the unfair allocation, in particular when decision-makers decide to delegate the communication under the independent punishment form. Whereas the delegation of communication under the aligned punishment might be perceived as task splitting among the "unit decision-maker/spokesperson" it is possibly perceived as an irresponsible attempt of shifting blame under the independent punishment.

Receiving an unfair allocation increases the probability of punishment significantly confirming individuals' inequality aversion. Also, decision-makers are significantly more punished than spokesperson, which indicates that receivers are aware that the decision-maker is responsible for the unfair allocation of money. Overall, the receivers' punishment does not differ for un-delegated and delegated communication under the aligned punishment. Under the independent punishment, the receivers' punishment does however increase significantly when spokespersons communicate the unfair allocation instead of the decision-maker, which is a reaction to the more frequent expression of remorse rather than the delegation itself. All findings hold for the punishment frequency and intensity.

2. Fairness, Delegated Communication and Punishment

Organizations deal every day with the communication of bad news, such as negative performance feedbacks, corporate downsizing and employee layoffs (Bies 2010). Managers are often called to communicate bad news, but they are typically unwilling to do. This is because communicating bad news may cause emotional distress (Tesser and Rosen 1975; Folger and Skarlicki 2001), or because the person delivering the bad news may become the target of anger and retaliation by the recipients of the bad news (Tripp and Bies 2009). Overall, Bies (2010) shows that business leaders classify the communication of a bad news as one of the most difficult tasks.

Employing a spokesperson may provide a solution to this problem. Letting a spokesperson communicate the bad news can allow the decision-maker to avoid the negative emotions entailed in communicating an unpleasant decision to the affected person. The reduction in punishment through the delegated communication of an unfair decision could be driven by a (perceived) shift in responsibility (Bartling and Fischbacher 2012), a shift in blame (Gurdal, Miller, and Rustichini 2012), and/or a shift in attention (Kahneman 1973). Whereas the delegation of the allocation choice leads to a shift in the responsibility for the allocation choice (Bartling and Fischbacher 2012), the spokesperson in our design *cannot be responsible* for the allocation choice. Delegation of communication may however provoke that the spokesperson *is hold responsible* for the unfair choice by the receivers, i.e. is blamed even though she cannot influence the monetary outcome (Gurdal, Miller, and Rustichini 2012). Another possible mechanism of how delegating communication affects the receiver's punishment is the shift in attention to the spokesperson, who is not responsible for the allocation choice. How affected persons *perceive* the decision-maker's *intention* of the delegation decision is another factor that will be discussed below. The decision-maker's intentions of the delegation decision might not only impact the affected persons but also the spokesperson and her communication strategy.

Three models help us understand how affected persons may react to an unfair allocation. First, if they are self-interested payoff maximizer and punishment is costly, they would not give up part of their earnings to punish an unfair decision. Second, if the affected persons have social preferences (Fehr and Schmidt 1999) and act based on intentions and fairness (Rabin 1993), they may punish an unfair allocation even if punishing is costly. Third, if the affected persons interpret the decision-maker's choice based on responsibility, they will punish an unfair decision and the person they consider responsible for the decision (Bartling and Fischbacher 2012).

In elaborating their punishment reaction, the affected persons may take into account whether the unfair decision was communicated by the decision-maker or by the spokesperson and how the unfair decision was communicated. One possibility is that they blame the person who delivers the bad news for the unfair allocation, and thus show her anger by punishing mainly the person in charge of the communication (see the arguments in Tripp and Bies 2009). The idea that spokespersons may be blamed even if they did not make the harsh decision is similar to the one in Gurdal, Miller and Rustichini (2013), who point out that people tend to blame others for events, such as the outcomes of a lottery, for which the latter ones cannot be considered responsible. On the opposite, the communication delegation to the spokesperson may shift the attention of the affected persons from the decision-maker responsible for the unfair decision to another person not responsible for the allocation choice (see Kahneman 1973). In this case, the affected persons could punish the decision-maker less than in the case of un-delegated communication. A third possibility is that the affected persons are indifferent to the delegation decision, and they punish only an unfair allocation.

It is important how the decision-makers' behavior and the reaction of the affected persons are influenced by the way in which the involved parties perceive the connection between decision-maker and spokesperson and the intentions behind their decisions. First, we consider a case in which receivers that want to punish are obliged to punish both the decision-maker and the spokesperson using the same punishment size. In this case, the decision-maker and the spokesperson "assume the same responsibility" by receiving the same punishment and delegating the communication may be perceived as task splitting. Consequently, neither the decision-maker nor the spokesperson benefit from shifting blame or showing particularly courageous assumption of responsibility. Second, we consider a case in which the persons affected by an unfair decision may punish the decision-maker and the spokesperson differently (or punish only one of them). In this case, the decision-maker could be possibly the only one punished for an unfair decision and delegating the communication may be perceived as an irresponsible attempt of shifting blame. The same is true for the communication content, which can aim at shifting blame in this setting in contrast to the previous scenario.

3. Experimental Design

In section 3.1, we explain the one-shot modified dictator game (adding communciation and punishment) used in our experiment, the theoretical predictions assuming selfish players, as well as the general procedure. In section 3.2, the six treatment are discussed in detail. Section 3.3 explains the research questions and discusses alternative hypotheses.

3.1 Experimental Design and Procedure

At the beginning of the experiment, participants are randomly divided in groups of four players.[36] The general instructions of the experiment can be found in Appendix 1. In each group there are three types of players randomly assigned to participants: one decision-maker (dictator, player A); one –spokesperson (player B); and two –receivers (players C).

The experimental procedure can be divided in three main steps: (1) allocation, (2) communication, and (3) punishment decisions. In the first step, each decision-maker has to choose between a fair or unfair allocation of her initial endowment, which corresponds to €20. The *fair* allocation assigns €5 to each member of the group: herself (the decision-maker), the spokesperson and the two receivers; the *unfair* allocation assigns €9 each to herself (the decision-maker) and the spokesperson, and €1 each to the two receivers.

Next, depending on the design, either the decision-maker or the spokesperson has to communicate the allocation choice made by the decision-maker. If the decision-maker decides to delegate the communication, the spokesperson takes over and has the duty to communicate the allocation choice made by the decision-maker to the receivers choosing one out of two pre-formulated sentences reported in Appendix B. The person who communicates the allocation decision learns about this form of communication and the available pre-formulated sentences at the moment he or she reaches this stage of the experiment. The sentences are almost identical for the decision-maker and the spokesperson. We formulated two sentences for the fair and the unfair allocation choice, respectively, following the most used categories in Andreoni and Rao (2011). For the fair choice, the sentence refers (1) to the fairness of the choice or (2) to the reduced earnings of the decision-maker and the spokesperson. For the unfair choice, the sentence expresses (1) remorse (emotional excuse) or (2) need (in combination with you'd do it too, rational explanation). Ideally we would have liked to have free form communication, but because of gender issues and noise in free form messages, we decided in favor of this form of communication. This way, the person communicating the allocation decision is able to decide at least in a very limited way about the communication content.

Finally, the two receivers learn who communicated the allocation choice and the corresponding text message. In response to the allocation and the communication, the two receivers can choose to punish or not the decision-maker and/or the spokesperson, and decide on the size of the punishment. Receivers can decide to give up €1 and reduce the decision-maker's and the spokeperson's payoffs by up to €3.5 per player (values between €0 and €3.5 in intervals of €0.5, total reduction up to €7). They can also decide to punish less than €3.5 per player, and give up the money not used. The €1 fee for punishing is fixed, i.e. independent of the amount by which the receiver decides to reduce the decision-maker's and the spokesperson's earnings.[37] Each receiver decides independently from each other. After the two receivers have made the punishment decision independently, one of the two receivers is chosen randomly and his decision is applied. The other receiver's decision does not affect anybody's earnings. We are interested in the receivers' punishment reaction to an unfair allocation decision. Since we expect that some decision-makers will choose the fair money split we have two receivers per group to get twice as many punishment observations than allocation decisions in order to compensate for the reduced number of interesting punishment decision.

The game is played one-shot and total earnings for the decision-maker and the spokesperson are given by the allocation payoffs, €9 or €5, minus the potential punishment assigned by the randomly chosen receiver. The total earnings of players C are given by the

[36] Players never learn the identity of each other before or after the experiment.

[37] We use a fix punishment fee because we do not want the punishment amount to be influenced by the cost of punishment, but instead by variables as explained in the treatment section.

allocation payoffs, €1 or €5, minus the cost of the punishment of €1 if he decides to punish and his decision is chosen randomly to be applied.

The decision-maker and the receivers' (selfish) payoff maximizing decisions are the following: since the maximum punishment of the decision-maker is €3.5 she is always better off by choosing the unfair allocation, i.e. by assigning €9 to herself and to the spokesperson, respectively, and €1 to each of the two receivers. Payoff maximizing receivers should never punish because it reduces their earnings.

We conducted the experiment at the Universitat Autònoma de Barcelona (UAB, Barcelona, Spain) and Universitat de Valencia (UV, Valencia, Spain) between April 2013 and November 2013.[38] Participants were mainly undergraduate students from Universitat Autonoma de Barcelona and Universitat de Valencia and were recruited by email using the online recruitment system ORSEE (Greiner 2004) where they had voluntarily registered to participate in experiments. Each subject could participate only in one session. A total of 868 participants took part in the experiment. The average earnings per person were €9.43 (including a show-up fee of €5). Subjects were paid privately in cash after the experiment finished. The average duration of a session was slightly less than one hour, including the instructions reading and payment.

3.2. Treatments

We run a total of six treatments that vary in two dimensions: we adopt three different communication forms, each of which is implemented under two different punishment forms. Table 4.1 provides a summary of the characteristics and the number of observations for each treatment.

Table 4.1. Overview of treatments

Treatment	Communication delegation	Punishment	No. of groups (Barcelona)	No. of groups (Valencia)	Total no. of groups
Vol Del-Al Punish	Decision-maker decides who (decision-maker or spokesperson) communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson has to be the same	21	22	43
No Del-Al Punish	Decision-maker communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson has to be the same	12	20	32
Man Del-Al Punish	Spokesperson communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson has to be the same	12	20	32
Vol Del-Ind Punish	Decision-maker decides who (decision-maker or spokesperson) communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson can be different	23	22	45
No Del-Ind Punish	Decision-maker communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson can be different	13	20	33
Man Del-Ind Punish	Spokesperson communicates allocation decision to receivers	Receiver's punishment amount for decision-maker and spokesperson can be different	12	20	32
			93	**124**	**217**

First, we explain the three communication forms. In the main treatment (*voluntary delegation*, treatment *Vol Del*), the decision-maker can *choose* whether to communicate

[38] From the summer period on, we had difficulties running sessions at UAB because of students being on summer break and technical issues with the recruitment system. We therefore ran some sessions at UV and tried to obtain a balanced number of observations for each treatment from UAB and UV in order to avoid treatment effects steaming from different subject pools and/or locations. For an overview of the composition of the treatments see Table 1.

personally the decision made or delegate the communication to a spokesperson. In the control treatments, the delegation of communication is not voluntary, but exogenously assigned by the treatment. Here, either the decision-maker has to communicate personally the decision made (*no delegation*, treatment *No Del*) or the spokesperson is obliged to communicate the decision previously made by the decision-maker (*mandatory delegation*, treatment *No Del*). We run each treatment varying the punishment structure.

We adopt two punishment structures. In the *aligned punishment* (treatment *Al Punish*), the two players C are forced to punish both player A and B by using the same punishment size; in other words, they can either punish both or none. For example, if a player C wants to punish the decision-maker by cutting €2 from her earnings, then she has to cut €2 from the spokesperson's earnings as well. The two players C do not need to agree on whom to punish and by how much. In the *independent punishment* (treatment *Ind Punish*), players C can decide to punish player A and/or B using different amounts. In other words, they can now punish both spokesperson and decision-maker or only one of them, and they can also choose a different amount of punishment.

We have a total of 43 (independent) groups for treatment *Vol Del-Al Punish*, 32 groups for treatment *No Del-Al Punish*, 32 groups for treatment *Man Del-Al Punish*, 45 groups for treatment *Vol Del-Ind Punish*, 33 groups for treatment *No Del-Ind Punish*, 32 groups for treatment *Man Del-Ind Punish*. Observations from Universitat Autonoma de Barcelona and Universitat de Valencia are more or less equally distributed among the six treatments, with Barcelona constituting between 38% and 51% of the number of observations per treatment.

3.3. Research questions and hypotheses

We discuss several determinants of the allocation and delegation decision in section 2: (1) intrinsic motives: avoidance of shame and emotional distress and (2) extrinsic motives: receivers' punishment which may be affected by shifting blame, attention (direct effects of communication delegation), and the receivers' perception of the communication delegation as task splitting versus an irresponsible attempt of shifting blame (indirect effect of communication delegation).

Expectations about the allocation decision

H1.1: On the basis of avoidance of shame/emotional distress and the possibility to reduce the punishment due to shifted attention, we posit that decision-makers choose more often the unfair allocation if the spokesperson communicates the decision (treatments *Man Del-Al Punish* and *Man Del-Ind Punish*) as opposed to the decision-maker communicating (treatments *No Del-Al Punish* and *No Del-Ind Punish*).

H1.2: On the basis of the possibility to reduce the punishment due to shifted blame with (exogenously) delegated communication and independent punishment, we posit that decision-makers choose more often the unfair allocation in treatment *Man Del-Ind Punish* than in treatment *Man Del-Al Punish*.

H1.3a: If decision-makers expect to shift blame with the communication delegation *decision* under the independent punishment, we expect more unfair decisions in treatment *Vol Del-Ind Punish* than in treatment *Vol Del-Al Punish*.

H1.3b: If, however, shifting blame does not affect the decision-makers' allocation choice but instead her expectations about the receivers' perception of the delegation decision, decision-makers will choose more often the unfair allocation in treatment *Vol Del-Al Punish* (delegation decision perceived as task splitting) than in treatment *Vol Del-Ind Punish* (delegation decision perceived as irresponsible attempt of shifting blame).

Expectations about the delegation decision

H2.1: In a similar vein as in H1.1, we posit that decision-makers who choose to split the initial endowment unequally delegate the communication to the spokesperson more often in treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish* (avoidance of shame/emotional distress and the possibility of reducing punishment by shifting attention).

H2.2a: With the independent punishment, the decision-makers can try to shift the blame to the spokesperson, which is not worth doing with the aligned punishment. Decision-makers will be more likely to use a spokesperson to communicate an unfair allocation in treatment *Vol Del-Ind Punish* than in treatment *Vol Del-Al Punish* if they care about shifting blame.

H2.2b: If on the opposite decision-makers care about the perception of the communication delegation (irresponsible with the independent and task splitting with the aligned punishment), decision-makers will delegate the communication more often to the spokesperson in treatment *Vol Del-Al Punish* than in treatment *Vol Del-Ind Punish*.

Expectations about the communication decision

H3.1: Decision-makers (spokespersons) are (not) responsible for the allocation choice and therefore more likely to explain the decision with needs that reveal standing by the allocation decision (remorse that reveals hiding behind the excuse).

H3.2: Spokespersons perceive the delegation *decision* as task splitting as they assume the same punishment consequences as the decision-maker (as an irresponsible attempt of shifting blame) and choose the expression of needs equally (remorse more) often in treatment *Vol Del-Al Punish* (*Vol Del-Ind Punish*) than in treatment *Man Del-Al Punish* (*Man Del-Ind Punish*).

We outline some behavioral factors in section 2 that might influence the receivers' punishment. These are inequality aversion, responsibility for the unfair choice, the direct (shifts in blame and attention) and indirect (perception of the delegation as neutral task splitting versus irresponsible attempt of shifting blame) effects of communication delegation, as well as the direct (change of fairness consideration versus acknowledgement of socially unaccepted allocation) and indirect effects (perception of shifting/reducing blame) of the communication content. H4.1 considers fair and unfair allocations, H4.2-H4.8 formulate hypotheses for unfair allocations only.

Expectations about the punishment of unfair allocation choices

H4.1: Because of inequality aversion receivers are expected to punish unfair allocations more than fair allocations.

H4.2: Receivers punish the decision-maker more than the spokesperson in the independent punishment treatments *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish* because they are aware that the decision-maker is responsible for the allocation decision.

H4.3a: If the receivers' attention is shifted with delegated communication without (with) the delegation *decision* then the punishment is expected to be lower in treatment *Man Del-Al Punish* than in *No Del-Al Punish* (lower with than without delegation in *Vol Del-Al Punish*) for either communication strategy.

H4.3b: If the receivers perceive the communication delegation *decision* not as irresponsible but instead as task splitting (decision-maker and spokesperson "assume the same responsibility" as they get punished by the same amount) and are not influenced by the shifted attention, the punishment is expected to be the same with and without delegation in *Vol Del-Al Punish* for either communication strategy.

H4.4a: If the receivers' blame is shifted with delegated communication without (with) the delegation *decision* then the punishment of the decision-maker will be lower in treatment *Man Del-Ind Punish* than in *No Del-Ind Punish* (lower with than without delegation in *Vol Del-Ind Punish*) for either communication strategy.

H4.4b: If the receivers perceive the communication delegation *decision* as an irresponsible attempt of shifting blame, the punishment of the decision-maker is expected to be higher with and without delegation in treatment *Vol Del-Ind Punish* for either communication strategy.
H4.5a: If the receivers' attention is shifted with delegated communication without (with) the delegation *decision* then the punishment of the spokesperson is expected to be lower in treatment *Man Del-Ind Punish* than in *No Del-Ind Punish* (lower with than without delegation in *Vol Del-Ind Punish*) for either communication strategy.
H4.5b: If the receivers' blame is shifted with delegated communication without (with) the delegation *decision* then the punishment of the spokesperson is expected to be higher in treatment *Man Del-Ind Punish* than in *No Del-Ind Punish* (lower with than without delegation in *Vol Del-Ind Punish*) for either communication strategy.
H4.6: By expressing needs (remorse), fairness considerations may be altered (the unfairness of the decision may be acknowledged). If this is the case the punishment will be lower after the expression of needs than after the expression of remorse in all six treatments.
H4.7: With the aligned punishment, receivers do not perceive that the decision-maker or the spokesperson is particularly responsible by expressing needs (shifts blame by expressing remorse) and the punishment is the same with either communication strategy in treatments *Vol Del-Al Punish*, *No Del-Al Punish*, and *Man Del-Al Punish*.
H4.8: With the independent punishment, receivers punish the expression of remorse more than the expression of need if the former is perceived as a coward way of avoiding assuming responsibility and shifting blame and the later as a straightforward way of assuming responsibility in treatments *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish*.

4. Experimental Results

Using a post-experiment questionnaire, we collect personal details, such as age, and the gender of participants. The total number of participants is 868, out of which 486 are women (56%) and 382 are men (44%). Participants have an average age of about 24.[39] In section 4.1, we will present the results of the decision-makers' allocation decision in all six treatments and her delegation decision in the voluntary delegation treatments. Section 4.2 resumes the decision-maker's and the spokesperson's communication strategy. In section 4.3, we will present and discuss the results of the receivers' punishment choices. Finally, in section 4.4, we conclude with the effects of gender.

4.1. Fairness of the allocation decision and delegation of communication

The shares of decision-makers who decide for the unfair allocation are 72% (31 of 43), 59% (19 of 32), 66% (21 of 32), 49% (22 of 45), 64% (21 of 33), and 63% (20 of 32) in treatments *Vol Del-Al Punish*, *No Del-Al Punish*, *Man Del-Al Punish*, *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish*, respectively. Remember that the fair split assigns €5 to each group member and the unfair split gives €9 (€1) to the decision-maker and spokesperson (the receivers), respectively.

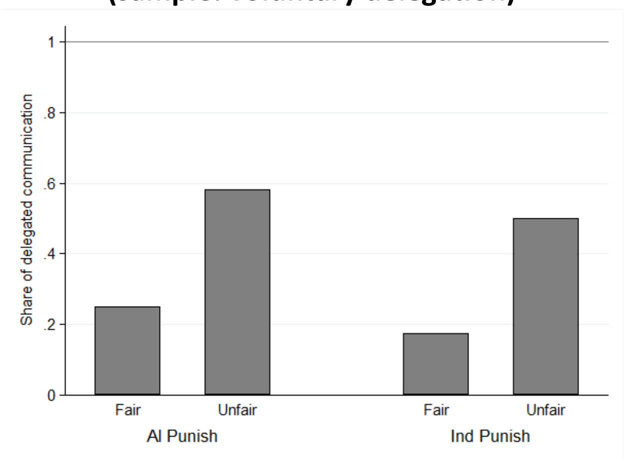
The actual *delegation decision* and the decision-makers' expectations about the receivers' perception thereof results to be important for the fairness of the allocation choice. We find only one significant difference in the allocation choice in the *voluntary delegation* treatments: decision-makers choose significantly more often the unfair allocation in treatment *Vol Del-Al Punish* than in treatment *Vol-Del-Ind Punish* (p=0.031, two-sided Fisher's exact test). We can reject the null hypothesis (no difference in the share of the unfair allocation

[39] We also collect information on education, which is classified using different study fields: economics, psychology, sociology, languages, medicine, computer science, and others. 38% of the participants come from economics.

between treatments *Vol Del-Al Punish* and *Vol-Del-Ind Punish*) and find confirmation for our alternative hypothesis H1.3b. Decision-makers who can be punished with the same amount as the spokesperson only (independently from the spokesperson) seem to expect that others perceive the communication delegation of an unfair allocation as task splitting (as an irresponsible attempt of shifting blame). Apparently, this weights more than the possibility to actually shift blame with the communication delegation under the independent punishment (H1.3a). Consequently, decision-makers choose the unfair (fair) split more often as they may expect not to face (to face) a conflict of interest when deciding on the communication delegation.

All other pairwise allocation comparisons between treatments are insignificant, failing to reject the null hypothesis to H1.1 (p = 0.856, treatments *Man Del* and *No Del*; p = 0.797, treatments *Man Del-Al Punish* and *No Del-Al Punish*; p = 1.000, treatments *Man Del-Ind Punish* and *No Del-Ind Punish*; two-sided Fisher's exact tests) and the null to H1.2 (p = 1.000, treatments *Man Del-Al Punish* and *Man Del-Ind Punish,* two-sided Fisher's exact test). We thus do not find evidence that shifted blame with (exogenously) delegated communication and independent punishment increases the share of unfair allocation decisions (H1.2). The data does neither support the H1.1 that, if the spokesperson communicates, the avoidance of shame and emotional distress as well as the possibility of reduced punishment due to shifted attention lead to more unfairness in the allocation.

Figure 4.1. Share of delegated communication by punishment form and allocation (sample: voluntary delegation)



The possibility to avoid shame and emotional distress and to reduce punishment due to shifting attention does however seem to affect the communication delegation decision. One of our main hypotheses (H2.1) is that decision-makers delegate the communication of the unfair allocation choice more often to the spokesperson than the communication of the fair choice. We use the data from the treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish*, where decision-makers decide on the communication delegation. For both treatments, we observe a significant positive relationship between the unfairness of the allocation and the delegation of communication, see Figure 4.1. The left (right) side of Figure 4.1 depicts the share of decision-makers in treatment *Vol Del-Al Punish* (treatment *Vol Del-Ind Punish*) who delegate communication after having chosen the fair or the unfair monetary split, respectively. Among the decision-makers who choose the fair allocation, 25% (3 of 12) and 17% (4 of 23) in treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish*, respectively, decide that the spokesperson communicates the allocation decision to the receivers. Among those who decide for the unfair allocation, the shares of those who delegate more than double reaching 58% (18 of 31) and 50% (11 of 22) in treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish*, respectively. The relationship between the fairness of the monetary split and the communication delegation is statistically significant (p = 0.088, treatment *Vol Del-Al Punish*; p = 0.029, treatment *Vol Del-Ind Punish*; two-sided Fisher's exact tests).

Table 4.2. Seemingly Unrelated Regression of allocation and communication delegation (S1a)-(S4b) and OLS of allocation (R1)-(R2)

	(S1a)	(S1b)	(S2a)	(S2b)	(S3a)	(S3b)	(S4a)	(S4b)	(R1)	(R2)
	SURE 1		SURE 2		SURE 3		SURE 4		REG 1	REG 2
VARIABLES	Unfair All	Delegation	Unfair All	Delegation	Unfair All	Delegation	Unfair All	Delegation	Unfair All	Unfair All
Ind Punish	-0.232**	-0.0761	-0.237**	-0.0616	-0.224*	-0.155	-0.257**	-0.156	0.0426	0.0164
	(0.101)	(0.164)	(0.0990)	(0.167)	(0.126)	(0.104)	(0.122)	(0.105)	(0.122)	(0.123)
Unfair All		0.331**		0.375**						
		(0.156)		(0.161)						
(Ind Pun)*(Unfair all)		-0.00456		-0.0107						
		(0.208)		(0.208)						
Delegation					0.266*		0.246*		0.0625	0.0416
					(0.137)		(0.130)		(0.123)	(0.122)
(Ind Pun)*(Del)					0.100		0.170		-0.0739	-0.0438
					(0.198)		(0.187)		(0.173)	(0.173)
Constant	0.721***	0.250*	2.501**	-0.369	0.591***	0.488***	2.336**	0.563	0.594***	1.019*
	(0.0725)	(0.133)	(1.182)	(1.218)	(0.0959)	(0.0740)	(1.105)	(1.254)	(0.0867)	(0.593)
Controls	No	No	Yes	Yes	No	No	Yes	Yes	No	Yes
Observations	88	88	88	88	88	88	88	88	129	129
R-squared	0.056	0.126	0.145	0.162	0.156	0.025	0.255	0.047	0.002	0.048

Notes. Coefficient estimates are shown (standard errors in parentheses). The dependent variable *Unfair All* takes the value 1 for the unfair allocation choice and 0 for the fair allocation choice. The dependent variable *Delegation* takes the value 1 for delegated communication and 0 if the decision-maker communicates. Control variables are a dummy for female, a dummy for economics studies, age and age^2 (all insignificant, except for the positive coefficient estimate for economics studies in regression models (S3a) and (S4a)). Sample is treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish* for regressions (S1a)-(S4b), and treatments *No Del-Al Punish*, *Man Del-Al Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish* for regressions (R1) and (R2).

*** p<0.01, ** p<0.05, * p<0.1

Table 4.2 summarizes regressions of the allocation (and delegation) choice on some explanatory variables. Regression models (S1a) through (S4b) show the coefficient estimates (standard errors in parentheses) of Seemingly Unrelated Regressions (SURE) with the dependent variables *Unfair All* and *Delegation*, and regression models (R1) and (R2) the Ordinary Least Square Regression (OLS) results with the dependent variable *Unfair All*. In treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish*, the allocation and the delegation decision are not independent of one another because the data come from the same subjects, i.e. the same decision-maker. The multiple equation model SURE takes into account the correlated errors in the two regression models. The dependent variable *Unfair All* takes the value 1 for the unfair allocation choice and 0 for the fair allocation choice. The dependent variable *Delegation* takes the value 1 for delegated communication and 0 if the decision-maker communicates. Control variables for individual characteristics are a dummy for female, a dummy for economics studies, age and age^2. The sample is treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish* for regressions (S1a)-(S4b), and treatments *No Del-Al Punish*, *Man Del-Al Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish* for regressions (R1) and (R2).

Models (S1a)-(S4b) regress the allocation and delegation decision on a dummy variable for independent punishment *Ind Punish* (taking value 1 for Ind Punish and 0 for Al Punish). In the model of the delegation decision (S1b), the variable *Unfair All* (1 for unfair and 0 for fair) and an interaction term between *Ind Punish* and *Unfair All* are added and the variable *Unfair All* is treated as if it was exogenous. Models (S2a)-(S2b) perform the same regressions as (S1a)-(S1b) adding the before mentioned control variables. In the model of the allocation decision (S3a), the variable *Delegation* (1 for delegated communication and 0 otherwise) and an interaction term between *Ind Punish* and *Delegation* are added and the variable *Delegation* is treated as if it was exogenous. Models (S4a)-(S4b) perform the same regressions as (S3a)-(S3b) adding the before mentioned control variables.

Models (S1a)-(S4b) confirm that about 24% less decision-makers choose the unfair allocation in treatment *Vol Del-Ind Punish* than in *Vol Del-Al Punish*. The positive and significant coefficient estimates for the variable *Unfair All* in (S1b) and (S2b) and for the variable *Delegation* in (S3a) and (S4a) cannot be interpreted as causal effects since both variables are in fact endogenous. The regression results show however that the unfair allocation choice and the delegation of communication are positively correlated, confirming H2.1. We fail to reject the null hypothesis to H2.2 ("difference in the delegation frequency of unfair allocations between Al Punish and Ind Punish") as the interaction terms in (S1b) and (S2b) are not significantly different from zero.

In regression models (R1) and (R2), the variable *Unfair All* is regressed on the dummy variable *Ind Punish*, the dummy variable *Delegation* for (exogenously) delegated communication and an interaction term of the two explanatory variables. The coefficient estimates of both the dummy variable *Ind Punish* and the interaction term are not significantly different from zero and can thus not reject the nulls to H1.1 and H1.2.

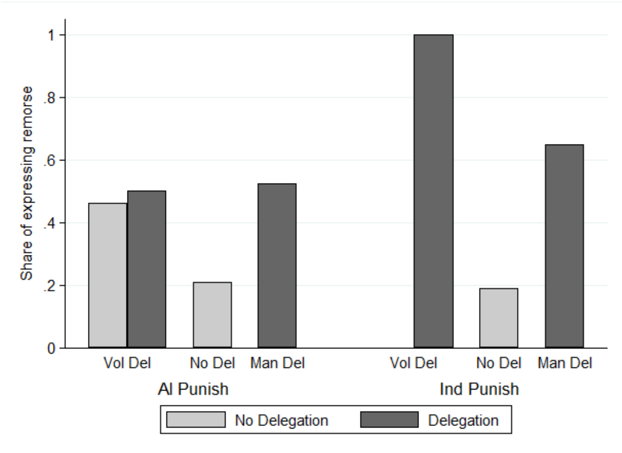
In summary, we find evidence that it is not the possibility of shifting attention (and blame) and of avoiding emotional distress, but the decision-makers expectation about the receivers' perception of the communication delegation that explain the fairness of the allocation choice. While with the aligned punishment (decision-maker and spokesperson share the consequences of responsibility and blame), the delegation of communication is likely to be perceived as task splitting, it may rather be perceived as an irresponsible attempt of shifting blame with the independent punishment (decision-maker and spokesperson do not share the consequences of responsibility and blame). Results support, however, the hypothesis that decision-makers are more inclined to delegating communication when they choose the unfair allocation possibly caused by the possibility of avoiding shame/emotional distress and shifting attention (and blame).

4.2. Communication Strategy

In this section we study how the decision-maker and the spokesperson communicate the allocation choice to the receivers and how the communication strategy is affected by the spokespersons' perception of delegation. As described in Section 3.1 and Appendix B, one sentence for the communication of the unfair allocation is based on "needs" (i.e. decision-maker: "*I chose option 1[the unfair allocation]. I made this decision because I need the money. Probably you would have chosen the same*."; spokesperson: "*Participant A[decision-maker] chose option 1[the unfair allocation]. I did not make the decision but I need the money. Probably you would have chosen the same*."), whereas the other expresses "remorse" (i.e. decision-maker: "*I chose option 1[the unfair allocation]. I feel uncomfortable with this decision. I am sorry.*"; spokesperson: "*Participant A[decision-maker] chose option 1[the unfair allocation]. I feel uncomfortable with this decision. I am sorry."*).

Results indicate that decision-makers and spokespersons have a different communication approach: decision-makers tend to express need while spokespersons rather express remorse. Figure 2 shows the share of decision-makers (un-delegated communication) and spokespersons (delegated communication) who choose the sentence expressing remorse/regret for the unfair allocation choice for each treatment. In the case of un-delegated communication, decision-makers chose the sentence based on needs in 21% (4 of 19) and 19% (4 of 21) of the cases if they are obliged to communicate themselves (treatments *No Del-Al Punish* and *No Del-Ind Punish*, respectively); by contrast, in the case of delegated communication, spokespersons chose the sentence expressing remorse in 52% (11 of 21) and 65% (13 of 20) of the cases if they have to communicate (treatment *Man Del-Al Punish* and *Man Del-Ind Punish*, respectively). The different approaches to communication of decision-makers and spokespersons are significant confirming our H3.1 (p = 0.055 for *Al Punish*, p = 0.004 for *Ind Punish*, two-sided Fisher's exact test).

Figure 4.2. Share of Decision-Makers (no Delegation) and Spokespersons (Delegation) Expressing Remorse



If the decision-maker *decides* whether to delegate the communication to the spokesperson or not, the punishment form (and the spokespersons' perception of the delegation decision) affects strongly the communication strategy of decision-makers and spokespersons: 46% (6 of 13) of the voluntary decision-makers express remorse for the unfair allocation under the aligned punishment similar to the 50% (9 of 18) of spokespersons in treatment *Vol Del-Al Punish* (p = 1.000, two-sided Fisher's exact test). On the contrary, none (0 of 11) of the voluntarily communicating decision-makers says that she feels remorse/regret under the independent punishment whereas all (11 of 11) spokespersons in charge of communicating in treatment *Vol Del-Ind Punish* express remorse (p = 0.000, two-sided Fisher's exact test). The share of spokespersons who express remorse in treatment *Vol Del-Ind Punish* (100%) is significantly larger than the same reference group in treatment *Man Del-Ind Punish* (65%) (p = 0.033, two-sided Fisher's exact test). The same comparison for the aligned punishment is not significant (50% in treatment *Vol Del-Al Punish* and 50% in treatment *Man Del-Al Punish*) (p = 1.000, two-sided Fisher's exact test). This supports our H3.2, which suggests that spokespersons perceive the delegation *decision* as task splitting with the aligned punishment because they assume the same punishment consequences as the decision-maker anyway, and as an irresponsible attempt of shifting blame with the independent punishment.[40]

In summary, decision-makers and spokesperson communicate the unfair allocation decision systematically differently. Decision-makers explain the allocation most of the times with needs and spokespersons express remorse and regret for the unfair allocation decision. While spokespersons in the aligned punishment treatment perceive the communication delegation indifferently (e.g. as neutral task splitting) and do not change their communication

[40] We can only speculate about the reasons driving the difference in the communication strategies among decision-makers. The different communication strategies of decision-makers in the treatments *Vol Del-Al Punish* and *Vol Del-Ind Punish* might be driven by self-selection (or be strategic): while more decision-makers in treatment *Vol Del-Al Punish* choose the unfair allocation, less do so in treatment *Vol Del-Ind Punish* compared to the reference groups with no delegation. The decision-makers who become aware of the protecting "shield" of the aligned punishment that facilitates communication delegation might be more likely to choose the unfair allocation even though they feel somehow guilty for the choice. The independent punishment might drive those decision-makers who feel slightly guilty away from choosing the unfair allocation leaving the selfish and/or needy decision-makers among the pool choosing the unfair split.

strategy, spokespersons in the independent punishment shift completely to expressing remorse for the decision-maker's unfair allocation choice. The later effect can be explained by the spokespersons' perception of the delegation as an attempt of shifting the receivers' blame.

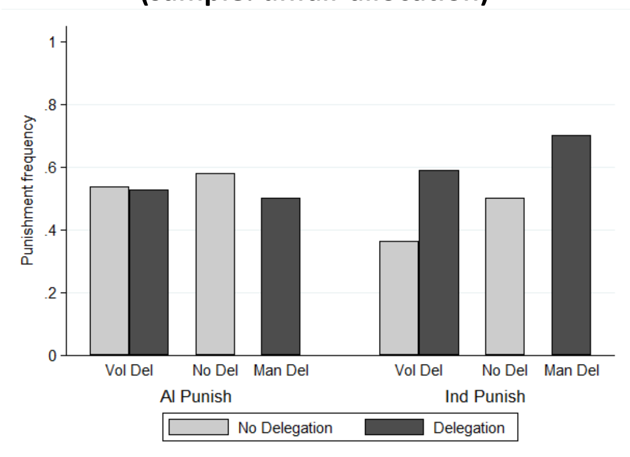
4.3. Receivers' punishment

Selfish payoff-maximizing receivers should never punish the decision-maker and the spokesperson if punishing is costly as in our design. Nonetheless, receivers use the punishment tool and do so significantly more often if the decision-maker chooses the unfair allocation (€9 to herself and spokesperson, €1 to each receiver) than if she chooses the fair split (€5 to everyone). While only 12% (20 of 166) of receivers decide to punish if the decision-maker chooses the fair allocation, 54% (146 of 268) of receivers punish an unfair allocation ($p=0.000$, all receivers, two-sided Fisher's exact test). Separate analysis for each of the six treatments reveal the same picture confirming H4.1 confirming inequality aversion ($p<0.036$ for each treatment separately, two-sided Fisher's exact test). With this result, we confirm previous findings of inequality aversion. Because of the low share of receivers punishing fair decisions, we will restrict the punishment analysis to unfair allocations.[41]

4.3.1. Receivers' punishment frequency

Figure 4.3 presents results on the receivers' punishment reaction to the unfair allocation chosen by the decision-makers for non-delegated and delegated communication and each treatment. 54% (14 of 26) and 53% (19 of 36) of receivers in treatment *Vol Del-Al Punish* decide to punish after learning from the decision-maker and the spokesperson, respectively, that the decision-maker chose the unfair allocation. With the exogenous communication assignment, the corresponding punishment rates are 58% (22 of 38) if the decision-maker communicates (treatment *No Del-Al Punish*) and 50% (21 of 42) if the spokesperson communicates (treatment *Man Del-Al Punish*). Under the independent punishment, the punishment rates are 36% (8 of 22) and 59% (13 of 22) after the decision-maker decides to communicate herself and to delegate the communication to the spokesperson, respectively (treatment *Vol Del-Ind Punish*). With the exogenous communication assignment, the corresponding punishment rates are 50% (21 of 42) if the decision-maker communicates (treatment *No Del-Ind Punish*) and 70% (28 of 40) if the spokesperson communicates (treatment *Man Del-Ind Punish*).

Figure 4.3. Punishment frequency by delegation and treatment (sample: unfair allocation)



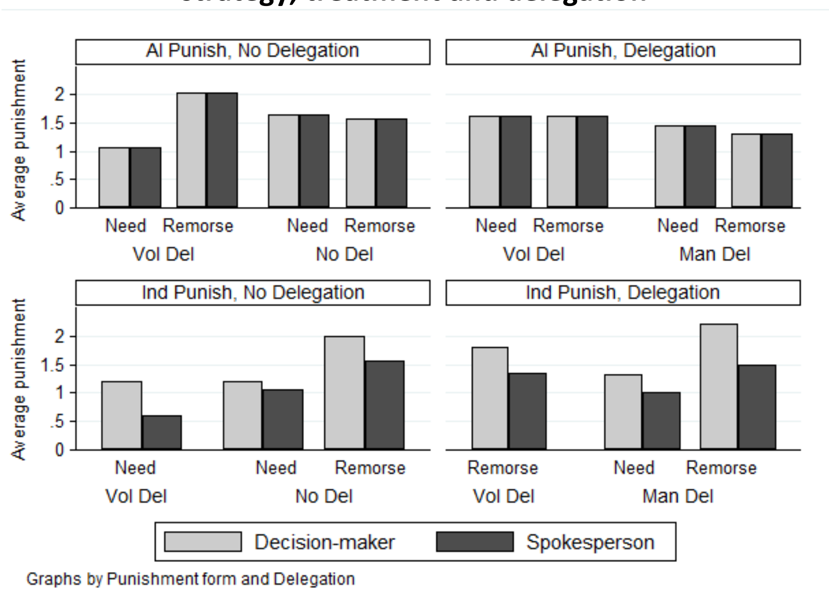
[41] The distribution of punishment after the fair allocation is slightly biased towards treatments with the independent punishment (14 of 94) compared to the aligned punishment (6 of 72). The number is yet too small to show a correlation with the punishment form, the delegation or the communication content.

With the aligned punishment, the delegation of communication has no effect on the punishment frequency of unfair allocations (p=0.613, two-sided Fisher's exact test); the same is true for separate analysis of endogenous and exogenous communication assignment (p=1.000 for treatment *Vol Del-Al Punish*, p=0.509 for treatments *No Del-Al Punish* and *Man Del-Al Punish*, two-sided Fisher's exact test). With the independent punishment in contrast, the delegation of communication increases the punishment frequency of unfair allocations significantly (p=0.021 for treatments *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish*, two-sided Fisher's exact test); the same is true for separate analysis with the exogenous communication assignment only (p=0.075 for treatments *No Del-Ind Punish* and *Man Del-Ind Punish*, p=0.227 for treatment *Vol Del-Ind Punish*, two-sided Fisher's exact test). However, for an analysis of the pure effect of communication delegation on punishment, we need to control for the communication strategies employed by the decision-maker and the spokesperson.

4.3.2. Receivers' punishment of the communication delegation

In the following we analyze the punishment intensity, i.e. the punishment amount that receivers assign to the decision-maker and to the spokesperson. After an unfair allocation choice with the aligned punishment (treatments *Vol Del-Al Punish*, *No Del-Al Punish*, and *Man Del-Al Punish*), receivers assign, on average, a punishment of €1.53 to the decision-maker and to the spokesperson, respectively. With the independent punishment (treatments *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish*), receivers punish the decision-maker (spokesperson) by €1.58 (€1.14), on average; this difference is statistically significant at the 1% level (p=0.000, two-sided Wilcoxon signed-rank test). A separate analysis for treatment *Vol Del-Ind Punish* with and without delegation, and for treatments *No Del-Ind Punish* and *Man Del-Ind Punish* draws a similar picture: the punishment assigned to the decision-maker exceeds in all four cases the amount assigned to the spokesperson ($p < 0.0605$ for each of the four cases separately, two-sided Wilcoxon signed-rank tests). This confirms our expectation that receivers make the decision-maker responsible for the unfair allocation decision and punish her accordingly more than the spokesperson (H4.2).

Figure 4.4. Punishment amounts of decision-maker and spokesperson by communication strategy, treatment and delegation



We now turn our attention to the pure effect of delegation on punishment controlling for the communication content. Figure 4.4 gives a detailed overview of the receivers' average punishment reaction to the expression of remorse and need phrase for each treatment and delegation separately. The light (dark) grey bar is the average punishment amount assigned to the decision-maker (spokesperson). Again, the alternative hypotheses H4.3a-H4.5b formulate the direct effect (shifting attention and eventually blame) versus the indirect effect (perception of the delegation as neutral task splitting or irresponsible attempt of shifting blame) of delegation on punishment, respectively.

With the aligned punishment (upper panel of Figure 4.4), the delegation of communication does not alter the punishment amount of the decision-maker and the spokesperson. We thus fail to reject the null hypothesis to H4.3, which is at the same time the second alternative hypothesis H4.3b. Controlling for the communication content "remorse" or "need," the difference in the punishment amount is not significantly different with and without delegation (p = 0.3437 for remorse, p = 0.7801 for need, treatments *Vol Del-Al Punish*, *Man Del-Al Punish*, and *No Del-Al Punish*; p > 0.3160 for each reason and delegation form (voluntary versus exogenous) separately; two-sided Mann-Whitney U test). We cannot confirm a direct (shift in attention) nor an indirect (perception of the delegation intention) effect of delegation on punishment with the aligned punishment.

With the independent punishment (lower panel of Figure 4.4), the delegation of communication does neither change the punishment amount of the decision-maker (p = 0.9005 for remorse, p = 0.5379 for need, treatments *Vol Del-Ind Punish*, *Man Del-Ind Punish*, and *No Del-Ind Punish*; p > 0.6000 for each reason with exogenous delegation separately; two-sided Mann-Whitney U tests) or of the spokesperson (p = 0.3437 for remorse, p = 0.6021 for need, treatments *Vol Del-Ind Punish*, *Man Del-Ind Punish*, and *No Del-Ind Punish*; p > 0.8180 for each reason with exogenous delegation separately; two-sided Mann-Whitney U tests). Here again we control for the communication content and fail to reject the nulls of H4.4 and H4.5. Neither the direct effect of communication delegation (shifting blame and/or attention) nor the perception of delegation (irresponsible attempt of shifting blame) seems to affect the receivers' punishment of the decision-maker and of the spokesperson.

4.3.3. Receivers' punishment of the communication strategy

How do receivers react to the decision-maker's and the spokesperson's expression of "remorse" or "need"? While the expression of need may change the receivers' fairness considerations, the expression of remorse acknowledges (indirectly) that the unfair allocation is socially unacceptable. With their punishment reaction, receivers show that they do not like the expression of remorse after having learned the unfair allocation choice: overall, 62% (72 of 116) of the receivers who receive a "sorry" for the unfair choice decide to punish, while only 49% (74 of 152) of the receivers who are informed about "needs" pay for punishing the decision-maker and the spokesperson (p=0.035, two-sided Fisher's exact test). The same is true for the punishment amounts on the aggregate level (1.79€/1.51€ with remorse, 1.37€/1.22€ with need) (p = 0.0146 for decision-makers, p = 0.0480 for spokespersons, all six treatments; two-sided Mann-Whitney U tests), but not for a separate analysis of the aligned punishment (p = 0.5308, treatments *Vol Del-Al Punish*, *No Del-Al Punish*, and *Man Del-Al Punish*; two-sided Mann-Whitney U test). We thus fail to reject the null that the communication contents "need" and "remorse" have the same effect on the receivers' punishment (H4.6). This leads us to the last two hypotheses on the perception of the communication strategies depending on the punishment form.

Receivers perceive the expression of needs and remorse differently with the two punishment forms. With the aligned punishment, the decision-maker or the spokesperson do not appear particularly responsible by expressing needs (shift blame by expressing remorse) and the punishment is the same with either of the two communication strategies (p > 0.9450, *Vol Del-Al Punish* with delegation, *No Del-Al Punish*, and *Man Del-Al Punish* separately; two-sided Mann-Whitney U tests). The only exception is the voluntary communication of the decision-maker (p = 0.0761, *Vol Del-Al Punish* without delegation, two-sided Mann-Whitney U test). It seems that the receivers perceive the expression of remorse negatively in this case (maybe because in this particular case the decision-maker does appear to reduce the responsibility for her own decision). Overall, we find some confirmation for H4.7.

With the independent punishment, we can partially confirm that the expression of remorse is perceived as a coward way of avoiding responsibility and shifting blame and the expression of needs as a straightforward way of assuming responsibility (H4.8). The punishment is in all cases larger after the expression of remorse than after the expression of needs, but it is significant only on the aggregate level (mostly) (p = 0.1855, *Vol Del-Ind Punish*, p = 0.0079, *No Del-Ind Punish* and *Man Del-Ind Punish*, for the decision-maker; p = 0.0651, *Vol Del-Ind Punish*, p = 0.0672, *No Del-Ind Punish* and *Man Del-Ind Punish*, for the spokesperson; two-sided Mann-Whitney U test). Note that the punishment amount of the spokesperson is particularly small in treatment *Vol Del-Ind Punish* without delegation (lower left panel of Figure 4.4). This indicates again that the decision-maker draws the receivers' attention to herself and away from the spokesperson by communicating voluntarily herself and possibly even more so by standing by the allocation decision and reasoning with needs.

Table 4.3: OLS of Punishment Amount with Al Punish (R1)-(R23) and Seemingly Unrelated Regression of Punishment Amount with Ind Punish (S1a)-(S3b)

VARIABLES	(R1) REG 1 Punish DM/SP	(R2) REG 2 Punish DM/SP	(R3) REG 3 Punish DM/SP	(S1a) SURE 1 Punish DM	(S1b) Punish SP	(S2a) SURE 2 Punish DM	(S2b) Punish SP	(S3a) SURE 3 Punish DM	(S3b) Punish SP
Delegation	-0.161		-0.205	0.646**	0.511**			0.160	0.173
	(0.272)		(0.279)	(0.285)	(0.247)			(0.376)	(0.329)
Remorse		0.159	0.206			0.875***	0.652**	0.763*	0.531
		(0.282)	(0.290)			(0.294)	(0.257)	(0.394)	(0.344)
Constant	-0.0385	0.178	0.231	3.741**	4.579***	5.595***	5.945***	5.400***	5.734***
	(2.001)	(2.035)	(2.040)	(1.720)	(1.494)	(1.845)	(1.612)	(1.899)	(1.659)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	142	142	142	126	126	126	126	126	126
R-squared	0.026	0.026	0.030	0.054	0.080	0.080	0.095	0.081	0.097

Notes. Coefficient estimates are shown (standard errors in parentheses). The dependent variables *Punish DM/SP*, *Punish DM*, and *Punish SP* take values between 0 and 3.5. *Punish DM/SP* is the punishment amount the receiver assigns to the decision-maker and the spokesperson with aligned punishment, *Punish DM (SM)* is the punishment amount the receiver assigns to the decision-maker (spokesperson) with independent punishment. Control variables are a dummy for female, a dummy for economics studies, age and age^2. Sample is unfair allocations of treatment *Vol Del-Al Punish*, *No Del-Al Punish*, and *Man Del-Al Punish* for regressions (R1)-(R3) and unfair allocations of treatments *Vol Del-Ind Punish*, *No Del-Ind Punish*, and *Man Del-Ind Punish* for regressions (S1a)-(S3b).
*** p<0.01, ** p<0.05, * p<0.1

OLS regressions of the punishment amount (taking values between 0 and 3.5 in discrete steps of 0.5) on *Delegation* (taking the value 1 if the spokesperson communicates and 0 otherwise) or the expression of *Remorse* (taking the value 1 for remorse and 0 for needs) show that the amount of punishment does not change with delegation or the communication strategy under the aligned punishment form, regressions (R1) and (R2) of Table 4.3, respectively. However, the punishment amount of the decision-maker and the spokesperson increases significantly with delegation or the expression of remorse under the independent punishment form, respectively, SURE regressions (S1a)-(S2b) of Table 4.3. The variables *Delegation* and *Remorse* are highly correlated, which is probably why the coefficient estimates in models (S3a) and (S3b) are mostly insignificant. The slightly significant positive coefficient estimate in model (S3a) of the punishment for the decision-maker indicates however that the communication content drives the increase in punishment rather than the delegation.

In summary of the previous subsections, we have shown that it is inequality aversion, responsibility, and the perception of the intention of the communication strategy, that drive the receivers' punishment rather than the communication content itself, the delegation of communication or the perception of communication delegation. Decision-makers are not able to transform their responsibility for the unfair allocation into blame by delegating the communication to the spokesperson. On the contrary, receivers seem to be sensitive to the attempt of shifting blame with the communication strategy and punish it accordingly.

4.4. Variations in gender

Recent papers have shown the presence of gender differences in fairness and communication. In particular, it has been shown that women are more inequality averse than men in the context of a dictator game (Croson and Gneezy 2009; Eckel and Grossman 1998), and that women adopt a more indirect, elaborate and emotional communication style (see von Hippel, Wiryakusuma, Bowden and Shochet 2011 and related references). We investigate whether gender influences the likelihood to choose the unfair allocation, to delegate communication, the communication style and/or to punish. In untabulated regressions (available upon request), we find that male and female decision-makers do not differ in their delegation behavior. Neither do we find any difference in the communication strategy or punishment behavior of male and female receivers. A remarkable (yet insignificant) gender difference results in treatment *Vol Del-Ind Punish*, where the reduced ratio of unfair allocations seems to be driven by male participants: 16 of 26 male decision-makers chose the fair split avoiding the conflict of interests when it comes to the delegation decision, while 12 of 19 female decision-makers choose the unfair allocation, which they then communicate themselves mostly (8 of 12).

5. Conclusion

Imagine a manager that makes a harsh decision, e.g. a wage cut, and that such decision can be communicated to employees either personally by the manager or by a spokesperson. In this thought example, does the manager make harsher decisions when she relies on the spokesperson to communicate? Does the communication strategy influence how employees respond to the managerial decision? And would the employees punish the spokesperson, the manager, or both? We use a laboratory experiment to investigate these questions.

First, a decision-maker allocates an initial endowment choosing between a fair and unfair allocation. Our results highlight a statistically significant relationship between allocation decision and punishment structure. Decision-makers who can decide on delegating communication tend to choose more frequently the unfair allocation when receivers can only punish by subtracting the same amount from spokesperson and decision-maker's earnings. In this setting, the decision-maker may believe that the delegation will be perceived as task splitting as decision-maker and spokesperson "assume the same responsibility" for any decision because they get punished by the same amount anyway. With independent punishment, others might perceive the communication delegation of an unfair allocation as an irresponsible attempt of shifting blame thus creating a conflict of interests when it comes to the delegation decision.

Second, the decision-maker decides whether to delegate or not the communication of the decision made to the spokesperson. Results show that the relationship between allocation and delegation decisions is statistically significant. Decision-makers who choose unfair allocations are significantly more likely to delegate the communication. This result is consistent with decision-makers avoiding the shame and emotional distress when communication an unpleasant decision and believing that, following an unfair allocation, the receivers' attention is shifted towards the spokesperson.

Looking at the spokespersons' reaction to the communication delegation in form of the chosen communication strategy, we find evidence that they perceive the delegation as an unpleasant action (of shifting blame) with the independent punishment and as a neutral decision with aligned punishment. Overall, spokespersons excuse the unfair allocation decision mostly with the expression of remorse/regret, while decision-makers explain it mostly with needs.

Concerning the receivers' behavior, we find that, on average, receivers punish the decision-maker stronger than the spokesperson. Receivers even punish more if the spokesperson communicates the unfair decision instead of the decision-maker under the independent punishment which is a reaction to the more frequent expression of remorse among spokespersons. Note that all our results are obtained in an environment where the person communicating the information does not directly face the receivers and, even stronger, where all decisions are anonymous. Results are likely to be much stronger when taking away these restrictions.

Our findings are relevant for the design of an effective corporate communication strategy and show how important the context is in which the decision-maker employs the spokesperson. If receivers notice a close link between decision-maker and spokesperson and cannot punish them independently, the decision-maker may be more inclined to making unfair decision if she is aware of the protective shield provided by the aligned spokesperson. We argue that the manager may decide to delegate the communication for at least three reasons: to shift attention, to shift blame, and to avoid the intrinsic shame and emotional distress when communicating a bad news. Our results confirm at least one of the three hypotheses; the manager prefers to delegate the communication of an unfair decision in order to avoid the distress of "facing" the person affected by the unfair decision and/or to shift attention. Because of both motives, delegating communication of unpleasant decisions is a way of taking away pressure from a responsible manager in a stressful situation.

Overall, we learn that when a manager has to communicate a bad news, whether she decides to communicate it personally or to employ a spokesperson, the employees' reaction depends crucially on how they perceive the delegation of communication. If manager and spokesperson are perceived as closely linked (sharing possible consequences) and the delegation is perceived as task splitting between the two, employees might not react strongly to the delegated communication. If the manager shows however a lack of courage and responsibility and, instead of communicating the decision herself, employs a spokesperson to communicate harsh policies to shift blame (while keeping full control over decision rights), she should be aware that this could fire back through the reaction of individuals negatively affected by these policies. The effects apparently work step by step: the decision-maker's intention with the delegation decision seems to affect the spokesperson's communication strategy, which on its turn seems to affect the receivers' punishment reaction. While delegated communication and the expression of remorse are not punished when the decision-maker and the spokesperson are closely linked, they are when there is room for shifting blame. The potential shift of blame results to be an important determinant of decisions and reactions.

If managers want to achieve full shifting of responsibility, an effective way is to delegate decision-rights altogether (Bartling and Fischbacher 2012). If decision rights are particularly valuable and managers are reluctant to delegate them to a third party, they will be kept responsible for the decisions made. But, even in this case, spokespersons may be useful as they allow managers to avoid the emotional cost intrinsic in the communication of bad news. A careful presentation of the reasons of the communication delegation and control of the spokesperson's communication strategy are recommendable and open up interesting future research paths.

6. References

Andreoni J., and Rao J.M. (2011) "The power of asking: How communication affects selfishness, empathy, and altruism", *Journal of Public Economics*, 95, 513-520.

Aghion P., and Tirole J. (1997) "Formal and real authority in organization", *Journal of Political Economy*, 105, 1-29.

Bartling B., and Fischbacher U. (2012) "Shifting the blame: On delegation and responsibility," *Review of Economic Studies*, 79, 67–87.

Bies R.J. (2010) "Leading change in the era of bad news: Dealing with anger, distrust, and revenge in the workplace," Working paper, Prosci Global Conference.

Bies R.J. (2012) "The delivery of bad news in organizations: A framework for analysis", *Journal of Management*, 39(1), 136-162.

Bolton P., and Dewatripont M. (2005) "Contract theory", *Cambridge, MA: MIT Press*.

Charness G. (2000) "Responsibility and effort in an experimental labor market", *Journal of Economic Behavior and Organization*, 42, 375-384.

Charness G., Cobo-Reyes R., Jiménez N., Lacomba J., and Lagos F. (2012) "The hidden advantage of delegation: Pareto-improvements in a gift-exchange game", *American Economic Review*, 102, 2358-2379.

Coffman L.C. (2011) "Intermediation reduces punishment (and reward)", *American Economic Journal: Microeconomics*, 3, 77-106.

Colombo M., and Delmastro M. (2004) "Delegation of authority in business organizations: an empirical test", *Journal of Industrial Economics*, 52, 53-80.

Croson R. and Gneezy U. (2009) "Gender differences in preferences", *Journal of Economic Literature*, 47, 1-27.

Eckel C. and Grossman P.J. (1998) "Are women less selfish than men? Evidence from dictator experiments", *Economic Journal*, 108, 726-735.

Erat S. (2013) "Avoiding lying: the case of delegated deception", *Journal of Economic Behavior and Organization*, 93, 273-278.

Fehr E., and Schmidt K. (1999) "A theory of fairness, competition, and cooperation", *Quarterly Journal of Economics*, 114, 817-868.

Fershtman C., and Gneezy U. (2001) "Strategic delegation: an experiment", *RAND Journal of Economics*, 32, 352-368.

Fischbacher U. (2007) "z-Tree: Zurich toolbox for ready-made economic experiments", *Experimental Economics*, 10, 171-178.

Folger R. and Skarlicki D.P. (2001) "Fairness as a dependent variable: why tough times can lead to bad management". In *Justice in the workplace: from theory to practice*. NJ: Lawrence Erlbaum.

Graham J.R., Harvey C.R., and Puri M. (2013) "Capital allocation and delegation of decision-making authority within firms", Working paper.

Gurdal M.Y., Miller J., and Rustichini A. (2013) "Why blame?", *Journal of Political Economy*, forthcoming.

Greiner, B. (2004), The Online Recruitment System ORSEE 2.0 – A Guide for the Organization of Experiments in Economics, University of Cologne, *Working Paper Series in Economics*, 10.

Hamman J.R., Loewenstein G., and Weber R.A. (2010) "Self-interest through delegation: an additional rationale for the principal-agent relationship", *American Economic Review*, 100, 1826-1846.

Kahneman D. (1973) "*Attention and effort*", Prentice-Hall, Englewood Cliffs, New Jersey.

Oexl R., and Grossman Z. (2013) "Shifting the blame to a powerless intermediary", *Experimental Economic*, forthcoming.

Rabin M. (1993) "Incorporating fairness into game theory and economics", *American Economic Review*, 83, 1281-1302.

Schelling T. (1960) "The strategy of conflict", *Cambridge: Harvard University Press*.

Tesser A. and Rosen S. (1975) "The reluctance to transmit bad news". In *Advances in experimental social psychology*. New York: Academic Press.

Tripp T.M. and Bies R.J. (2009) "Getting even: the truth about workplace revenge – and how to stop it. *San Francisco: Jossey-Bass.*

von Hippel C., Wiryakusuma C., Bowden J. and Shochet M. (2011) "Stereotype threat and female communication styles" *Personality and Social Psychology Bulletin*, 37, 1312-1324

7. Appendices

Appendix A. Instructions (Voluntary Delegation)

General information

Thank you for coming to the experiment. You will receive 5 Euro for the participation in the experiment. You will be assigned to a group and depending on your and your group members' decisions you can earn additional money during the experiment. We will pay you in private at the end of the experiment (you will receive your total earnings in a closed envelope). We ask you not to look at the other participants' screens. It is important that you do not talk to any of the other participants until the experiment is over. If at any time during the experiment you have a question, please raise your hand and one of us will come to your place to answer. During the experiment, the use of mobile phones is forbidden.

Role and group matching

At the beginning of the experiment, you will be randomly assigned to a group of four together with three other participants. You will not learn the identity of these participants before, during or after the experiment, nor will they learn your identity. Your decisions will be absolutely anonymous. There are three types of participants in this experiment: participants A, B, and two C.

Task

In this experiment, participant A has to decide how to distribute €20 among the four participants of his/her group. When distributing the Euros, participant A can decide between two possible options:

- Option 1: Participants A and B receive €9 each and the two participants C receive €1 each. (9, 9, 1, 1)
- Option 2: Each participant receives €5. (5, 5, 5, 5)

After participant A has made the decision, it will be communicated to the participants C. Participant A can decide whether he or she wants to communicate the decision or, on the contrary, that participant B communicates it.
Whoever has been chosen to communicate the decision previously made by participant A, participant A himself or participant B, will send a message to the participants C.
If participant A decides to communicate the option chosen him-/herself, the participant B will not take any action. In this case, participant A makes the decision and communicates the decision. If participant A decides that participant B will be the one communicating the option chosen by A, participant A cannot take any further action. In this case, participant B communicates the decision made previously by participant A. The table below provides a summary of the Euro that every participant receives depending on the options that participant A can choose.

	Participant A	Participant B	Participant C	Other participant C
Option 1	€9	€9	€1	€1
Option 2	€5	€5	€5	€5

After participant A has chosen one of the two options, participant A or participant B (if he/she has decided that participant B will communicate) will send a message about the decision made. Participants C will learn whether participant A chose B to communicate the decision or not, the text message, and also the option chosen by A. Both participants C will have the possibility of giving up 1 of their Euro to assign up to a total of 7 negative Euro to the earnings of participant A *and (and/or)* participant B. The number of Euro that the participant C wants to reduce *has (does not have)* to be the same for participant A and B (maximum of €3.5 per person, values between 0 and 3.5 in intervals of 0.5). The total of negative Euro assigned can also be inferior to 7.
Subsequently, one of the participants C will be chosen randomly and his/her decision will be implemented; on the other side, the decision of the other participant C will not affect any earnings.

Example: Option 1 (9, 9, 1, 1) is chosen (by participant A) and communicated by A or B, and the randomly chosen participant C gives up €1 to reduce A's and B's earnings by €2. The final payments would be as follows:

Participant A	Participant B	The chosen participant C	The other participant C
9-2=€7	9-2=€7	1-1=€0	€1

Please answer the following practice questions in order to get familiar with the experiment. The decisions and numerical values in the practice questions are chosen on a purely random basis and are not to be considered as a hint or suggestion as to how you could decide. Your answers to the practice questions will have no effect on your payment at the end of the experiment.

1. Suppose Option 2 (5, 5, 5, 5) is chosen (by participant A), and the randomly chosen participant C gives up €1 to reduce A's and B's earnings by €3.5. ¿What would be the final Euro of each participant?

Participant A	Participant B	The chosen participant C	The other participant C
€	€	€	€

2. Suppose Option 1 (9, 9, 1, 1) is chosen (by participant A), and the randomly chosen participant C does not give up €1 to reduce the other participants' earnings. ¿What would be the final Euro of each participant?

Participant A	Participant B	The chosen participant C	The other participant C
€	€	€	€

Appendix B. Communication

The decision-maker and the spokesperson communicate the former's allocation choice to the receivers by choosing one of two pre-formulated sentences showed on the screen.

I) The decision-maker chooses the unfair allocation.

I.a) The decision-maker has to choose between the following two sentences:

1. *I chose option 1. I feel uncomfortable with this decision. I am sorry.*
2. *I chose option 1. I made this decision because I need money. Probably you would have chosen the same*.

I.b) The spokesperson has to choose between the following two sentences:

1. *Participant A chose option 1. I feel uncomfortable with this decision. I am sorry.*
2. *Participant A chose option 1. I did not make the decision but I need the money. Probably you would have chosen the same*.
3.

II) The decision-maker chooses the fair allocation.

II.a) The decision-maker has to choose between the following two sentences:

1. *I chose option 2. I think this is an equal split because we all receive the same*.
2. *I chose option 2. Please, take into account that the spokesperson's earnings and mine will be affected by this decision. I hope you are fine with this decision*.

II.b) The spokesperson has to choose between the following two sentences:

1. *Participant A chose option 2. I think this is an equal split because we all receive the same amount.*
2. *Participant A chose option 2. Please, take into account that the decision-maker's earnings and mine will be affected by this decision. I hope you are fine with it*.