IWH
Halle Institute for Economic Research
Member of the Leibniz Association

Discussion Papers

No. 9
May 2019

Predicting Free-riding in a Public Goods Game: Analysis of Content and Dynamic Facial Expressions in Face-to-Face Communication

Dmitri Bershadskyy, Ehsan Othman, Frerk Saxen

Authors

Dmitri Bershadskyy
Corresponding author
Halle Institute for Economic Research (IWH) – Member of the Leibniz Association, Department of Macroeconomics
E-mail: dmitri.bershadskyy@iwh-halle.de
Tel +49 345 7753 863

Ehsan Othman
Otto-von-Guericke University Magdeburg
E-mail: ehsan.othman@ovgu.de
Tel +49 391 6711 231

Frerk Saxen
Otto-von-Guericke University Magdeburg
E-mail: frerk.saxen@ovgu.de
Tel +49 391 6711 231

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of IWH. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors.

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

Editor

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Predicting Free-riding in a Public Goods Game: Analysis of Content and Dynamic Facial Expressions in Face-to-Face Communication*

Abstract

This paper illustrates how audio-visual data from pre-play face-to-face communication can be used to identify groups which contain free-riders in a public goods experiment. It focuses on two channels over which face-to-face communication influences contributions to a public good. Firstly, the contents of the face-to-face communication are investigated by categorising specific strategic information and using simple meta-data. Secondly, a machine-learning approach to analyse facial expressions of the subjects during their communications is implemented. These approaches constitute the first of their kind, analysing content and facial expressions in face-to-face communication aiming to predict the behaviour of the subjects in a public goods game. The analysis shows that verbally mentioning to fully contribute to the public good until the very end and communicating through facial clues reduce the commonly observed end-game behaviour. The length of the face-to-face communication quantified in number of words is further a good measure to predict cooperation behaviour towards the end of the game. The obtained findings provide first insights how a priori available information can be utilised to predict free-riding behaviour in public goods games.

Keywords: automatic facial expressions recognition, content analysis, public goods experiment, face-to-face communication

JEL classification: C80, C92, D91

* The authors would like to thank Joachim Weimann, Ayoub Al-Hamadi and the participants of the 2018 Meeting of the Association for Experimental Economics Research (Paderborn) for their helpful comments.

Financial support from the Halle Institute for Economic Research (IWH Explore project SP2015-03) is gratefully acknowledged.

1. Introduction

Communication is an elementary component of our society. It is absolutely vital in terms of achieving stable cooperation among members of a group. Depending on the specific structure of the problem communication can enhance the coordination towards possible solutions. A very specific structure where communication is known to improve behavior from the perspective of a social planner are Prisoner's Dilemma alike problems (e.g. public goods game) where the socially and individually optimal solutions do not coincide. In order to illustrate this dilemma within the scope of economic laboratory experiments, researchers often focus on the voluntary contribution mechanism (VCM) in which individuals form a group to fund a public good everyone benefits from. However, due to the mechanism design from the individual perspective, it is beneficial to free-ride and simply take advantage of the contributions of the other group members. Comprehensive overviews on the topic were conducted by Ledyard (1995) and Chaudhuri (2011). Hereby, face-to-face communication (FFC) was shown to be a very intuitive hands-on solution against free-riding. Several researchers focused on the question why communication prior to the standardized VCM leads to very high and stable contribution rates. Previous research, by e.g. Cason and Khan (1999), Brosig et al. (2003), Bochet et al. (2006) narrowed down the cause for this effect by the use of different treatments (e.g. audio-communication without video, video-communication without audio). In contrast to this, the approach applied in this paper is based upon new technological possibilities adjusted specifically for the underlying data (Othman et al., 2019)[1]. This however, limits causal reasoning within this research. The videos and the data set are obtained from the experiment conducted in Altemeyer-Bartscher et al. (2017).

The main economic question with respect to laboratory research on VCM is how to increase the contributions of the individuals. In this paper, we extend this line of research by another question arising from the perspective of a social planner – how to identify groups that are going to provide socially sub-optimal contribution rates to the public good prior to their actual contributions. The ultimate goal, therefore, could be to introduce additional interventions only when the prediction based on a priori available information concludes that the group needs another push (e.g. nudges, formal institution) towards the social optimum. Since it is favorable to limit this type of public interventions to a minimum it is beneficial to identify sub-optimally performing groups as precisely as possible. In the context of the underlying experiment the predictions can be based on a priori available information: general distribution of sub-optimally performing groups in the population, communication data (e.g. content or facial expressions) which precedes the contribution stage. Hereby, revealing otherwise private information on future free-riding through easily observable facial expressions would have severe implications.

The general research question of how to identify well-performing groups goes along with two questions. First, whether simply seeing and being seen by the other members of the group while discussing the problem increases the contributions and second, whether the efficiency gains yield from the context channel. FFC, being superior to other types of communication (Brosig et al., 2003; Bochet et al., 2006) not only enables coordination but also reduces the social

[1] This paper is submitted yet at the moment (05/2019) not publicly accessible. The authors would gladly send the manuscript if desired.

distance among the participants. More recent analysis appended another explanation being that communication enables type detection of the interacting subjects (He et al., 2017). However, the question remains whether these processes can be approached technically using facial action unit detection algorithms. The technical approach described in Othman et al. (2019) is extended by classifying simple contextual information transcribed from the original communication.

Given previous research, it is an expedient hypothesis that contributions depend on the context of the communication, which resolves the coordination problem, and facial expressions, which are a possible channel to reduce social distance. The results of the paper indicate that some contextual information indeed provides explanation to group contributions at the end of the experiment. The end-game phase in the context of the underlying research is the time when no contribution stages are left ahead and the incentives to deviate from the socially optimal contribution rates are the highest. In practice this addresses everyday situations such as idling in the final days of an employment contract or more general issues, e.g. paying taxes shortly before leaving the jurisdiction of the tax officers. The experimental results indicate that groups which specifically mentioned that they aim to remain at the full contribution strategy until the end had significantly higher contribution rates in the last periods of the experiment. The evaluation of facial expressions enhances the analysis of which groups provided full contribution rates. Although there was only little deviation in the contribution rates over time, optimizing the model using the random forest classifier led to increases in predictive power as compared to the trivial classifier. The following analysis of the obtained true negatives supports the findings from the contextual analysis. Groups that stopped active communication earlier, i.e. spoke fewer words or instead of communicating stared into the empty space at the end of the communication period, contributed less at the last stage of the VCM.

The remaining paper is organized as follows. Section 2 provides a literature review. Section 3 briefly explains the experimental design and focuses on the data obtained from the transcribed communications, sections 4 and 5 present and discuss the results, and section 6 concludes.

2. Literature

From early on, communication was subject to experimental research. From the perspective of economic research, it gained on importance with communication devices becoming more and more widespread (Bordia, 1997). The contributions by Dawes et al. (1977) and Isaac and Walker (1988) are hereby the first illustrating the unambiguous effect of communication in prisoner's dilemma problems with multiple players. Following the argumentation of Frank (1988) the clues for this beneficial behavior may occur due to different reasons such as facial or verbal expressions. The results were partially unexpected because the communication used in the experiments was de facto cheap talk and the theoretical effect of cheap talk in scenarios with strong incentives to lie is expected to be low (e.g. Farrell and Rabin, 1996; Crawford, 1998). Due to this observation, several studies were conducted in order to distinguish the transmission channels of this effect. Brosig et al. (2003) and Bochet et al. (2006) use separate treatments for each form of communication, e.g. face-to-face, audio-video, audio only, video only, chat. Both analyses confirm face-to-face types of communication (table conference and video conference) as the superior means of communication. Brosig (2006) provided an overview of different types of experiments which involved communication illustrating the effect of communication in several different experimental designs. One possible idea is that, besides assisting the coordination of group behavior, FFC reduces social distance. This can be interpreted as a degree of reciprocity individuals believe in within social interactions and which affects the individuals' behavior (Hoffman et al., 1996). However as Brosig et al. (2003) showed, simply reducing the social distance by providing short time (the individuals saw each other for 10 seconds) visual identification did not provide significant increments in contribution behavior. Nonetheless the hypothesis that facial expressions, such as seeing a happy face, affect human behavior in economic experiments found support (Eckel and Wilson, 2003). Further, Haruvy et al. (2017) illustrated the interaction between communication and visibility in the laboratory and in a virtual world. In order to focus on the aspect of identification Andreoni and Petrie (2004) explicitly excluded the effect of changes in facial expressions. Using photos they argue that identification alone reduces free-riding and when combined with information it increases contributions to a charity. While these studies focused on static facial expressions, more recent analyses, e.g. Belot et al. (2012), Konrad et al. (2014), Sparks et al. (2016), Belot and van de Ven (2017), investigate dynamic expressions in the context of trust and deception detection using human assessment methods. Further, general evidence on the advantages of using incentivized economic experiments to analyze deception or cooperation was discussed by ten Brinke et al. (2016) and more specifically for facial expressions by Bonnefon et al. (2017).

Although some technologically advanced methods have been utilized in the past, they were sometimes used for other purposes. Fiedler et al. (2013) provided a study that focuses on the way humans gather information in public goods experiments by tracking the eye-movements of the subjects. In a simple sender-receiver experiment with biased transmission by Wang et al. (2010), deceiving senders had dilated pupils and reduced information gathering of the payoffs of the deceived receivers. The authors illustrated how obtaining and applying these information would increase the predictions of the true state and change the payoff allocation between the players. Further, research focused on the measurement of specific physiological characteristics and their effect on decision making as it is outlined in e.g. Sanfey et al. (2003), Kenning and Plassmann (2005), Glimcher et al. (2009), Dimoka et al. (2012), Al Osman et al. (2014) for neuro-economics or different types of biofeedback. Yet, in contrast to the analysis of facial cues, this type of analysis relies on less easily obtainable data. Automatic analysis of social behavior using computer vision and machine learning algorithms is an emerging field of research (George and Leroux, 2002; Gatica-Perez, 2009; Aran and Gatica-Perez, 2011). The ultimate aim is to infer human behavior by observing and analyzing the interaction of the group conversation taken from the audio and video channels. Hopfensitz and Mantilla (2018) analyzed the performance of FIFA World Cup players based on their portraits. Jaques et al. (2016) trained deep neural networks using the facial expression of one-minute segments of the conversation to predict whether a participant will experience bonding up to twenty minutes later. In contrast to previous research we aim to predict contribution behavior in a financially incentivized public goods game after three-minute FFC applying automatic dynamic analysis of facial expressions. For the best of our knowledge this has not been done so far.

While the analysis of facial expressions in experimental economics is a young research branch which, due to technological improvements, is gaining on relevance, the context of communication is more researched. Brosig et al. (2003) used a simple yet effective classification of the content in a very similar experimental setup. The authors identified groups that discussed e.g. the optimal strategy, threats and end-game effect. Chaudhuri et al. (2006) implemented intergenerational advices in a public goods experiment and looked at the content of the messages and how much public these were. However, no statistical analysis was pursued in order to investigate the effect of the content on the contribution behavior. In a more recent analysis, e.g. Lopez and Villamayor-Tomas (2017), Palfrey et al. (2017), Arechar et al. (2017) look deeper into the content of communication focusing on the relevance of information, strategic decisions communicated in it, or the level of truthfulness respectively. Using communication restrictions Zultan (2012) differentiates the effects of social and strategic communication prior to the ultimatum game. In another setup Chen and Houser (2017) analyze the ability of individuals to detect deception with one major finding being that the number of words increases the trustworthiness. Further, the authors illustrate how mentioning specific content relevant words, e.g. money, influences the credibility. However, we are not aware of a content analysis of unrestricted and simultaneously happening face-to-face communication in a public goods game that is directly linked to the contribution behavior of the subjects.

3. Design and Data

The data was collected alongside a larger experimental setup described in Altemeyer-Bartscher et al. (2017). The total experiment consisted out of three blocks of which only those blocks are useful that involve pre-play communication. For instructions please see Appendix A. In total 384 students took part in the experiment. The duration of every complete session was on average around 80 minutes. The subjects were incentivized using real money with the conversion rate of 1 Laboratory Dollar = 4.5 Cents. Only one of the three blocks was actually paid out. However, since the subjects were informed that the decision which block is payoff relevant is conducted randomly after the end of the experiment, it is ensured that every block is correctly incentivized. After the experiment, the subjects filled out a questionnaire including some demographics. The experimental design was executed in z-Tree (Fischbacher, 2007). The experiment was organized and recruited with the software hroot (Bock et al., 2014).

The relevant parts of the experiment for the analysis in this paper are mainly block two (including the first time communication – FTC) and parts of block three (second time communication – STC) of the complete experiment[2], where the participants had the chance to communicate face-to-face via audio-video communication software prior to the VCM. During the communication period of three minutes, the participants were free to discuss anything. The duration of the group communication was determined by using pilot sessions. Since the discussions were non-binding the communication constituted cheap talk. The VCM is described by the following profit function for individual j in period k:

$$\pi_{jk}\left(g_{jk}\right) = z\text{-}g_{jk} + \frac{\alpha}{n}\sum_{j=1}^{n} g_{jk}\,, \quad j=1,\ldots 4$$

[2] By design of the complete experiment some participants were allowed to have the identical type of communication in the third block. However, due to randomization between the experimental blocks it was ensured that no individual in STC can meet anybody they have talked in FTC again.

with the initial endowment (z) = 20 Laboratory Dollar (LD), the efficiency multiplier (α) = 2, g_{jk} representing the amount of LD subject j invested in period k. The individuals repeat this decision 10 times in constant groups of four individuals.[3] After every of the 10 periods the participants receive anonymous information on the contributions of other subjects in their group. After the last period the individuals are informed about their total payoff for this block.

Table 1. Overview of content variables.

Parameter	Definition	Coding
Full Investment	The participant(s) mentioned to invest full contributions	"0" – no, "1" – yes
End-game awareness	The participant(s) mentioned that they should contribute fully until/in the end. (No explicit agreement required)	"0" – no, "1" – yes
Previous experiences	The participant(s) discussed experiences from previous block or prior experiments	"0" – no, "1" – yes
Threat and Consequences	The participant(s) "threatened" potential free-riders by explaining consequences, e.g. they will reduce their contributions.	"0" – no, "1" – yes
Disagree	How many players in the group (temporarily) disagreed with the solution	Numbers from 1 to 4
Information provider	Which player in the team explained the dilemma/solution of the dilemma (first).	Numbers 1 to 16 (linked to the specific individual in every session)

The face-to-face communication was recorded digitally. In order to analyze the context of the discussions, the communication was first transcribed and in a second step independently rechecked.[4] In order to assess several aspects of the discussion in a way that is less prone to subjective perspective of content, several meta parameters were chosen that are purely objective including the total word count of the conversations and the individual word counts of the respective subjects. Another set of variables was obtained using a classification scheme similar to the one conducted in Brosig et al. (2003). Table 1 summarizes these obtained variables[5] as

[3] Given this design, the dilemma occurs as the payoffs are encircled within the three simplified scenarios: (a) 40 LD if all subjects contribute fully, (b) 20 LD if no one contributes anything, (c) 50 LD for the individual that freerides while the all others contribute fully and receive 30 LD respectively.

[4] Due to technical constraints it was not possible to automatize voice-to-text-transcription. The audio channel was codified by the recording software which strongly impedes distinguishing the different speakers. The video recordings and the manually transcribed discussion protocols are available on request.

[5] The classification was conducted by one coder and later rechecked using a semi-automatic approach. Hereby, we distinguished words and phrases (e.g. "last period", "last stage", "until the end" and so on) that are likely to appear

well as how they are coded. Hereby it is important to differentiate between who these parameters address. The majority of the parameters focus explicitly on the group behavior. However, the individual word counts and the categorization on the main provider of information in the group refer to specific subjects and are analyzed on the individual level.

4. Results

Similar to previous research, face-to-face communication had a strong effect on contribution rates. This, however, caused a problem with respect to the evaluation of the possible explanations. Figure 1 depicts the problem showing the group contribution rates and the respective deviation at different periods using a box plot. Hereby it is only statistically plausible to focus the analysis on the last period of the experiment. Therefore, the result section focuses basically on the end-game behavior and the question what parameters are interrelated with making contributions more stable over time. In the first part of the results section, the predominant issue is the contextual analysis of the experiment. The second part of the section discusses the effect of facial expressions on the contribution behavior.

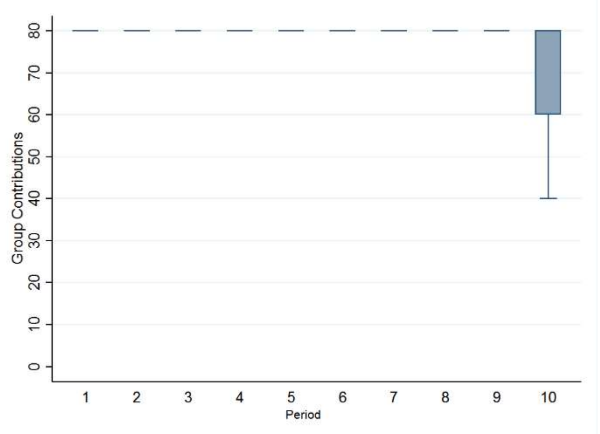


Figure 1. Boxplot of group contributions over ten periods (outliers not displayed).

It is further important to stress that we refrain from making causal claims. Despite the analysis being founded on a laboratory experiment, there are evidently no treatments in place to test for causality. In fact, the question whether the communication affects group contributions or the individual preferences towards contribution influence the communication cannot be answered based on this data set. Yet, it is plausible to assume an interaction between these factors.

4.1. Content Analysis

The analysis of the content is based on the transcribed communication protocols of the group discussions.[6] Since it is arguable whether combing first time communication and second time

when discussing the respective parameter (End-Game) and used a search function. In a second step the differently categorized results of the approaches were analyzed by the authors prior to and independent of linking the data to the experimental data. Although subjectively arguable, the categorization by the coder is more nuanced.

[6] Likewise to the subsequent analysis of facial expressions it is theoretically possible to automatize the process of content analysis. This was not done mainly for two reasons. First, in the course of the recording process the voices were saved using one channel though coming from four different sources. An automatic a posteriori mapping of voices to the individuals in the group was not possible with a sufficient reliability. Second, the language used was

communication is generally possible, the analysis is conducted separately for these cases. This reduces the observations to 96 for FTC and 31 for STC. The descriptive statistics of the variables, obtained from the communication protocols, are displayed in Table 2. Several findings can be obtained from this summary. Virtually every group was able to find and agree on the socially optimal solution of full contributions. The variables End-game, Experiences, and Threat vary among the groups. However, only the first one has a significant effect on the contribution behavior as it is depicted in Table 3 for both blocks. This is less surprising since the dependent variables provide variance only in the end-game part of the experiment and the parameter End-game specifically focuses on whether the group is aware of the problem. Groups that discussed the issue show significantly higher contribution rates in the last period. Discussing previous experiences or threatening co-players in case of free-riding did not have any significant effect. The combination of these observations yields the possible conclusion that the experimental setup might have been too easy for the communication. In fact, it appears to be enough to simply discuss the optimal strategy and recollect the end-game problem to achieve extraordinarily high and stable contribution rates. The benefits of this strategy are high. Furthermore, in the end, virtually all subjects agree on the strategy. The conjecture of the VCM being too easy for the communication goes along with previous research (Brosig, 2006) mentioning the role of the complexity of the problem.

Table 2. Descriptive statistics of content variables and meta parameters for FTC and STC.

Variable	Obs	Mean	Std. dev.	Min	Max	Obs	Mean	Std. dev.	Min	Max
	FTC					STC				
End Game	384	.302	.460	0	1	124	.323	.469	0	1
Invest All	384	.979	.143	0	1	124	1	0	1	1
Subjects Against	384	.083	.277	0	1	124	.097	.297	0	1
Prev. Expe-rience	384	.365	.482	0	1	124	.742	.439	0	1
Threats/ Conse-quences	384	.229	.421	0	1	124	.355	.480	0	1
Total word count	384	244.760	117.552	33	516	124	260.710	131.647	18	470
Ind. word count	382	61.521	55.108	0	306	123	65.715	62.004	1	237

often very colloquial German. Any other automatic or semi-automatic analysis would have required establishing or improving the word libraries as well as a certain classification for methods like e.g. sentiment analysis. This, in turn, should be similar to the method chosen in this paper, yet require a substantially higher effort.

Table 3. Effect of mentioning the End-Game and length of communication.

Group Contributions	End Game = 1	End Game = 0	Above-average word count	Below-average word count
FTC	75.862 (p=0.015)	64.328	71.392 (p = 0.083)	63.756
STC	78.2 (p=0.079)	65.714	69.1 (p=0.799)	70.909

Note: P-values of two-sided t-tests are depicted in brackets.

Furthermore, using simple metadata such as the total number of words spoken by the group enhances the findings. The mean of words spoken in the communication (248) is taken as a naive threshold to distinguish between two groups. Pursuing the tests on these binary variables yields the conclusion that the more words were spoken the better was the cooperation in the last periods. Surprisingly this holds only for the second but not for the third block. Here the group contributions are virtually identical as is displayed in Table 3.

In order to provide robustness with respect to the naively chosen threshold and ensure the effect was not driven by other variables, this paper presents regression results with different control variables. Due to dealing with censored data on group level the analysis utilizes Tobit regressions with boundaries at 0 and 80. To complete the analysis on group level several demographic values were aggregated to a group level, i.e. number of males or economists in the group and the sum of the ages of the individuals. The results are provided in Table 4 for the two blocks separately. Hereby the length of communication is significant in FTC and insignificant in STC. This holds despite adding the strongest content variable End-Game.

The table further shows that the End-Game variable, on contrary to the word count, is significant in FTC and weakly significant in STC. A possible explanation is that breaking the factually non-binding promise to keep high cooperation until the very end induces psychological costs to the individuals. Therefore, making such an agreement increases the cooperation. From the logical point of view, such a promise can only come up at the end of a specific train of thought. First, the group has to find the socially optimal strategy being fully cooperative. Second, the group members have to agree on their future contribution behavior. Only then it is reasonable to discuss the specifics towards the end of the game. This further provides a possible explanation why the lengths of communication matters only in FTC. The groups could have learned from their FTC and discussed the end-game earlier. Based on this idea it is plausible to assume that discussing the more negative experiences in FTC yields contributions to decrease while discussing the more positive experiences in STC yields contributions to increase.

Table 4. Tobit results for FTC and STC.

	FTC			STC		
Dependent variable: Group Contributions	(1)	(2)	(3)	(4)	(5)	(6)
Total word count	0.160**	0.167**	0.170**	0.025	0.028	-0.181
Number of economists		4.827	1.452		2.341	2.517
Number of males		-3.243	-3.533		-6.036	-7.456
Aggregate of age		-0.506	-0.911		3.289*	1.394
End-game			38.578**			95.108*
Invest all			64.731*			Omitted
Subjects against			-34.844			13.569
Previous Experience			-4.953			53.761*
Threats and consequences			-19.159			63.893
Constant	63.492***	107.305	85.059	95.985***	-202.146	-50.383
Adj. R-square	0.019	0.021	0.057	0.001	0.038	0.131
Number of Observations	96	96	96	31	31	31
LR-Chi2	7.32	8.43	22.59	0.09	4.58	15.81

Note: ***/**/* denote significance at 0.01/0.05/0.1 levels respectively. The variable Invest All was omitted due to lack of variation. See Table 2 for descriptive statistics.

On the individual level, the paper focuses on the question who explained the contribution strategy in the team. Besides analyzing the information providers it was possible to analyze the number of words communicated and relate both to basic demographic information, such as gender and field of study. The results of this analysis are depicted in Table 5. Despite the logical idea that economists being more familiar with the public goods dilemma due to their study or presumably higher experience in public goods experiments, the analysis does not indicate economists taking the lead neither in the qualitative variable (information provider) nor in the quantitative variable (talker – the person who talked the most in the group). This holds for the first communications (in block two) and the second communication period (selected groups in block three). However, things are different when considering the gender. Male subjects contributed significantly more to the communication. In block two (three) 32.5% (31.9%) of all males were coded as information providers while the share of information providers among females was 15.8% (16.4%)[7]. The results obtained for the variable talker are almost identical due to a high correlation between these variables, i.e. the individuals that explained the dilemma were also those who spoke the most in the group.

Table 5. Analysis of leadership.

		Male	Female	Econ	No-Econ
FTC	Talker	0.344	0.145 (p = 0.000)	0.269	0.243 (p = 0.551)
FTC	Information provider	0.325	0.158 (p = 0.000)	0.241	0.257 (p= 0.719)
STC	Talker	0.362	0.109 (p=0.001)	0.295	0.206 (p=0.258)
STC	Information provider	0.319	0.164 (p=0.048)	0.197	0.302 (p=0.180)

Note: P-values of two-sided t-tests are depicted in brackets.

The results presented so far represent standard analysis approaches with respect to the VCM. In order to achieve comparability with the subsequent analysis of facial expressions, which by its structure purely aims to predict group contribution behavior, we extend the section by providing a very simple prediction model based on all context variables that were obtained on the group level. Hereby the focus lies upon the binary distinction whether groups contributed the full amount or deviated from it. In line with the initial research question this aims to distinguish groups that do not need further institutionary support to achieve socially optimal contribution rates and those that do. The eventual prediction originates from a simple logit model with the following structure:

$$\widehat{FC} = \beta_0 + \beta_1 NW + \beta_2 EG + \beta_3 IA + \beta_4 SA + \beta_5 PE + \beta_6 TC + \varepsilon_i$$

with FC being the binary prediction whether the group provided full contributions, NW – number of words, EG – end-game, IA – invest all, SA – subjects against, PE – previous experience, TC – threats and consequences. Since the initial analysis of contents indicated some potential differences between FTC and STC the predictions are conducted for the two blocks separately. As the results of the logit model do not provide binary estimates but the probability for the value to be one (i.e. full-contribution) the accuracy of the predictions depends on a specific decision rule. This refers to the question from which probability onwards to assign the binary variable as a prediction for full-contribution. The naïve threshold of 50% yields results that outperform the average full-contribution rate. However, as there is no prior knowledge on the choice of threshold in such scenarios it is reasonable to investigate different thresholds and the respective effects on the accuracy. The subsequent Figures 2 and 3 present the results for the FTC and STC respectively. Several findings can be obtained from this analysis. Firstly, the predictions for FTC are on average worse than for STC. The most efficient threshold lies around

[7] Please note that by definition one out four individuals was coded as information provider. The fact that the respective numbers do not on average yield 0.25 is due to slightly higher share of males in the total sample.

60% and yields accuracy rates of 68.7% and 80.7% for the two blocks respectively. The combination of the two blocks yields a total accuracy rate of 70.1%. While it is arguable whether to analyze the two blocks separately, we can provide a possible explanation for the differences between the blocks. Investigating the simple descriptive statistics (Table 2) there is one major change to observe. There is a strong and significant (p-value of two sided t-test: 0.0002) increase in the number of groups that discussed previous experiences. While it is not possible to link the increment in accuracy to this single variable, there are reasonable differences with respect to what experiences the individuals had at the different points in time. In the FTC the only experience from the experiment was the communication-free experimental set up that yields lower payoffs. In the STC the experience combines the experiences from the communication-free experiment and the more successful pre-play communication setup of the experiment. As these past experiences influence the content, this justifies a distinct analysis of FTC and STC. The results are depicted in Figures 2 and 3.

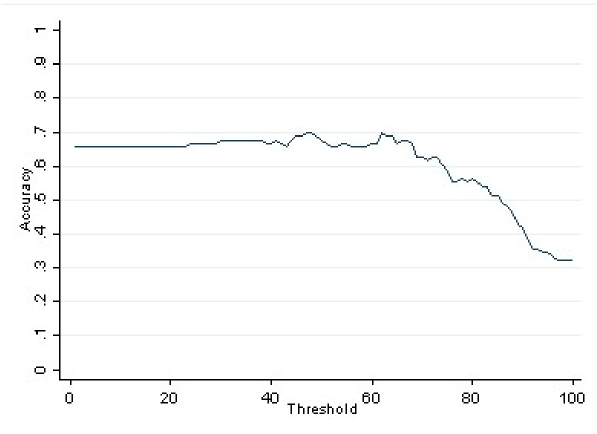


Figure 2. Accuracy of logit predictions depending on the classifying probability threshold in percent for FTC.

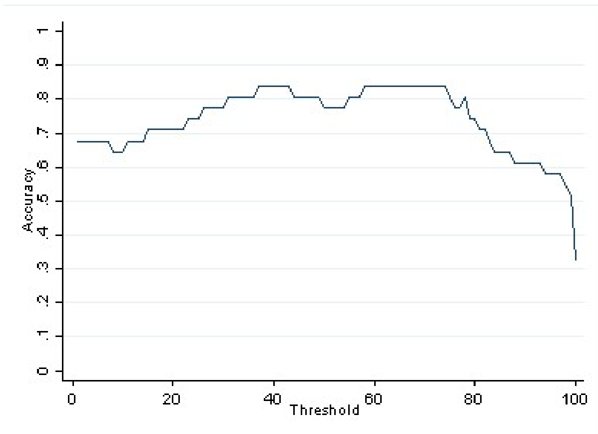
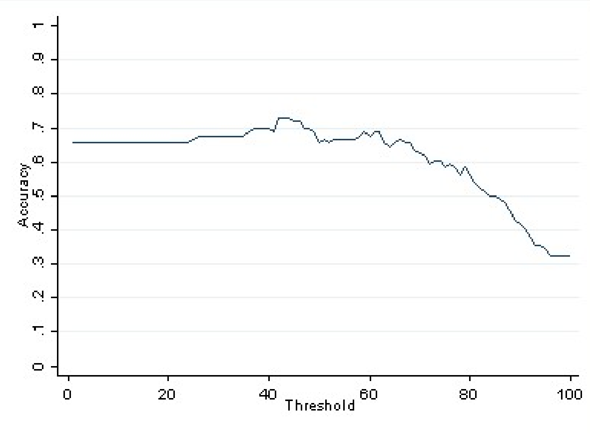


Figure 3. Accuracy rates of logit predictions depending on the classifying probability threshold in percent for STC.

Ultimately it is worth noting that the given predictions are based purely upon the content variables. On the contrary to the aforementioned Tobit-regressions the available demographics are not added into the model. This limitation is used as the demographics were obtained from a questionnaire following the entire experiment. Therefore, it is arguable whether this information is a priori obtainable and can be used to predict contribution before they take place. However, we analyzed the effect of including the demographics for one major reason that builds the bridge to the subsequent analysis of facial expressions. Research has shown that it is possible to predict gender and age with a substantial accuracy even based on random photos available on the internet (Levi and Hassner, 2015). By improvements this classification became better and can be applied on videos (Han et al., 2018). If the algorithms are to achieve the same quality as self-reported data in the questionnaire this would have an effect on the prediction accuracies in FTC (to 73%) and STC (to 90%) as is illustrated in Figures 4 and 5.



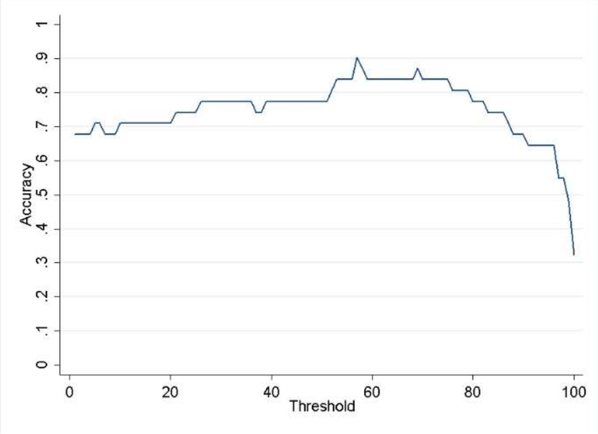


Figure 4. Accuracy of logit predictions depending on the classifying probability threshold in percent for FTC.

Figure 5. Accuracy rates of logit predictions depending on the classifying probability threshold in percent for STC.

4.2. Analysis of facial expressions

Subsequently, a video-based automatic facial expression analysis is presented examining 3 minutes of communication to predict the end-game behavior of the groups. To do so, a binary classifier was trained that predicts whether all 4 participants of a group will contribute fully in the very last period of the experiment or if anyone deviates, i.e., the binary classifier doesn't predict the contribution of each participant but likewise to the analysis in section 4.1 for the entire group. The dataset consists of 127 different groups divided into 24 sessions. The same subject might appear at most in two groups, but only within the same session. To train person-independent models the analysis uses leave-one-session-out cross-validation. This ensures that no subject appears in the training and test set simultaneously.

Each face-to-face communication (FFC) video has four participants. First, the facial activity descriptors are calculated for each individual face and frame. Then, the activity descriptors of each individual in all 4!=24 possible ways are concatenated to form the group activity descriptors. All group activity descriptors get the same label (see next paragraph FADs and GADs). This obviously increases the dataset, which is favorable in this case since the 127 FFC videos constitute a comparatively small dataset. Finally, all 3048 (= 24 * 127) group activity descriptors are classified individually and the classification outcome is averaged to obtain the prediction score.

4.2.1 Facial Activity Descriptors (FADs) and Group Activity Descriptors (GADs) :

Using OpenFace (Baltrusaitis et al., 2016) facial features from each individual face and frame are extracted. OpenFace first detects the face, facial landmarks, estimates eye-gaze, head pose, and extracts facial action units (AUs) (see Figure 6). The list of facial features used is depicted in Table 6. For the analysis, only those features were used that OpenFace can estimate robustly (see Baltrusaitis et al., 2016): 3 head pose features (yaw, pitch, and roll), 8 AU intensities features, and 10 AUs presence features. The FAD of each individual is extracted from the selective facial features for all frames in the FFC videos using the method of (Saxen, et al.

2017). After calculating the 4 FADs of the group, they are concatenated in all 24 possible ways (e.g. 1234, 1243, 1432, etc.) to form the GADs. Each GAD is given a group label.

Table 6. List of selected facial features that perform well based on OpenFace's own analysis (Baltrusaitis et al., 2016).

Head pose	**AU**	**AU Full name**	**Prediction**	**AU**	**AU Full name**	**Prediction**
Yaw	AU1	Inner brow raiser	P	AU15	Lip corner depressor	I
Pitch	AU2	Outer brow raiser	I&P	AU17	Chin raiser	I
Roll	AU4	Brow lowerer	P	AU20	Lip stretched	I&P
	AU5	Upper lid raiser	P	AU23	Lip tightener	I&P
	AU7	Lid tightener	I	AU26	Jaw drop	P
	AU9	Nose wrinkler	P	AU28	Lip suck	P
	AU14	Dimpler	I	AU45	Blink	I&P

Note: I – Action Unit intensity: 0 (absence of action unit), 1 (faint) to 5 (strong), P - presence (0 absent, 1 present).

4.2.2 Training

Our model uses leave-one-session-out cross-validation using Random Forest classifiers[8], i.e. the GADs of one session are held out for testing while the rest form the training set. In total, 24 sessions provide 24 results. The classification outputs for the 24 GADs per FFC videos are simply averaged and thresholded, whereas an optimal threshold is calculated based on the training set. The performance measure used is accuracy. For reference, an informed guess (trivial classifier) was calculated in each fold, which always votes for the majority class (usually full contribution). It provides 24 different results, depending on the distributions of the test sets. Based on the entire dataset, the average precision of the trivial classifier was 64.47%.

Since it is unprobable that the whole FFC is equally likely to predict future behavior, in the best case scenario the analysis could be limited to the facial expressions occurring in the aftermath of specific contentwise relevant statements. However, due to the aforementioned technical difficulties, it was not possible to synchronize the content with the facial expressions. This yields another approach. To investigate which part of the FFC video is more important for predicting contribution behavior, we process the FADs from different parts of the FFC video. First, we divide each FFC video into 3, 4, and 5 equal long videos (splits). Second, we extract the FADs and GADs of these 12 different splits (3 + 4 + 5 = 12) and train a model for each split and multiple combinations (33 in total, see the details in Othman et al., 2019). We introduce three different categories containing models from different split combinations (beginning models, end models, and a combination of beginning and end models). Each category includes 11 splits or combinations that belong to the particular part (beginning, end or both parts) of the

[8] For more detailed information on the utilized parameters consult Othman et al. 2019

FFC video. Each category provides 264 (11 splits * 24 results / split) different results. Table 7 shows an overview of different parameters for the automatic analysis of facial expression.

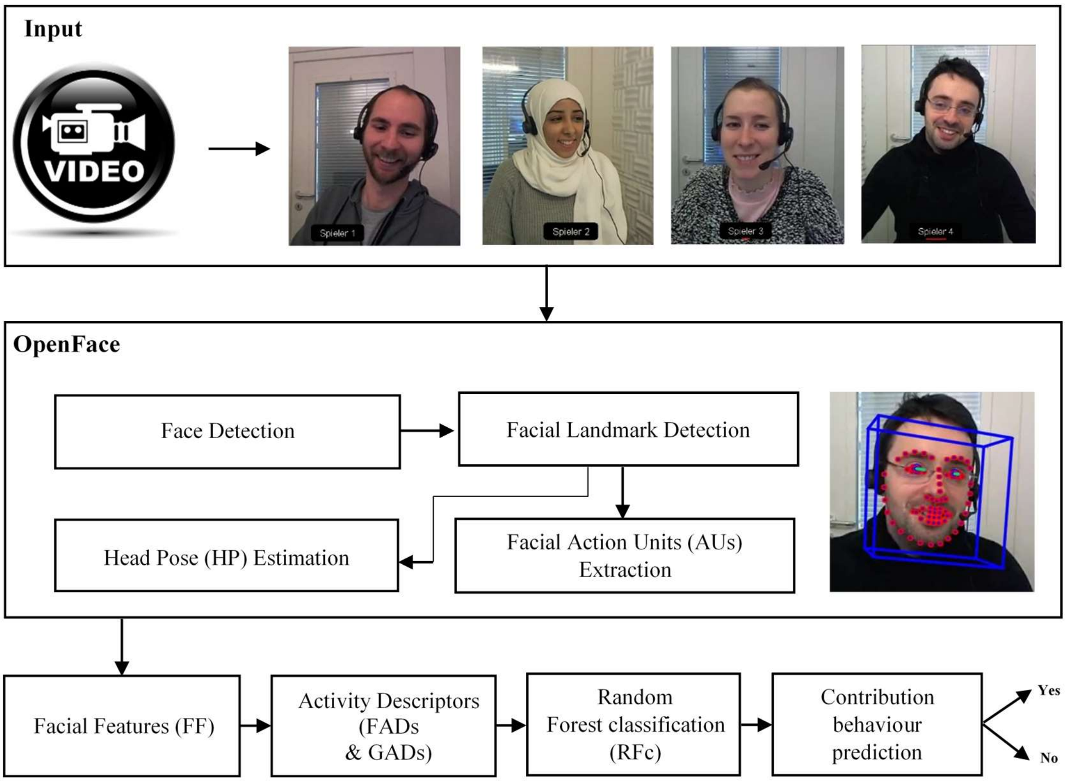


Figure 6. Automatic analysis of facial expressions from FFC video (4 players per video) using OpenFace, FADs, GADs and RFc to predict the contribution behavior of groups in the last game. OpenFace includes: face detection, facial landmark detection, head pose tracking, and facial Action Unit estimation. Pictures were obtained from a replication of the original setup.

Table 7. An overview of different parameters for the automatic analysis of facial expression.

<table>
<tr><th colspan="2">Entire dataset</th><th colspan="3">Each FFC video</th><th colspan="2">Each training set</th></tr>
<tr><td rowspan="2">Videos</td><td rowspan="2">127</td><td colspan="2" rowspan="2">Splits
(division of the video into 3, 4, 5 splits)</td><td rowspan="2">12</td><td>categories</td><td>3</td></tr>
<tr><td>Models</td><td>33</td></tr>
<tr><td>Sessions</td><td>24</td><td rowspan="3">Categories
(multiple combinations of 12 splits)</td><td>Combined splits</td><td>11</td><td colspan="2">each model</td></tr>
<tr><td colspan="2">Leave-one-session-out cross-validation</td><td>Beginning splits</td><td>11</td><td>Frame level features</td><td>21</td></tr>
<tr><td>Training set</td><td>24</td><td>End splits</td><td>11</td><td colspan="2">Each test set</td></tr>
<tr><td>Test set</td><td>24</td><td colspan="3">Each split</td><td rowspan="2">Results of each category</td><td rowspan="2">11</td></tr>
<tr><td colspan="2">Final result: average accuracy of the test set</td><td colspan="2">GADs</td><td>24</td></tr>
</table>

4.2.3 Results

The results show that predicting group behavior based on facial visual cues from the FFC video is complex and only slightly better than guessing (with the knowledge of the distribution – see trivial classifier). This task was expected to be especially difficult, since the decisions attempted to forecast are subject to much more hidden influences. Nevertheless, on average, end models predict about 70% of the decisions, which is significantly more than guessing (see Figure 7 and Table 8). Moreover, looking at the correctly and wrongly classified FFC videos, little difference in the behavior between groups was found at the beginning of the FFC videos. However, groups that do not contribute fully show less engagement later on in the FFC videos. The current results seem promising, and our hypotheses to explain these results is that the last part of FFC videos can be used to predict the contribution behavior of groups since it is easier to tell if the group is communicating well when the introductory phase already ended. Furthermore, participants might control their facial expressions more at the beginning of communication, while their spontaneous facial expressions expose at the end of the communication. Although the results indicate that the last part of the FFC video is much more informative for predicting the group behavior, we have no proof that last part of the FFC video is better than the beginning due to lack of data. Our findings suggest further research is needed.

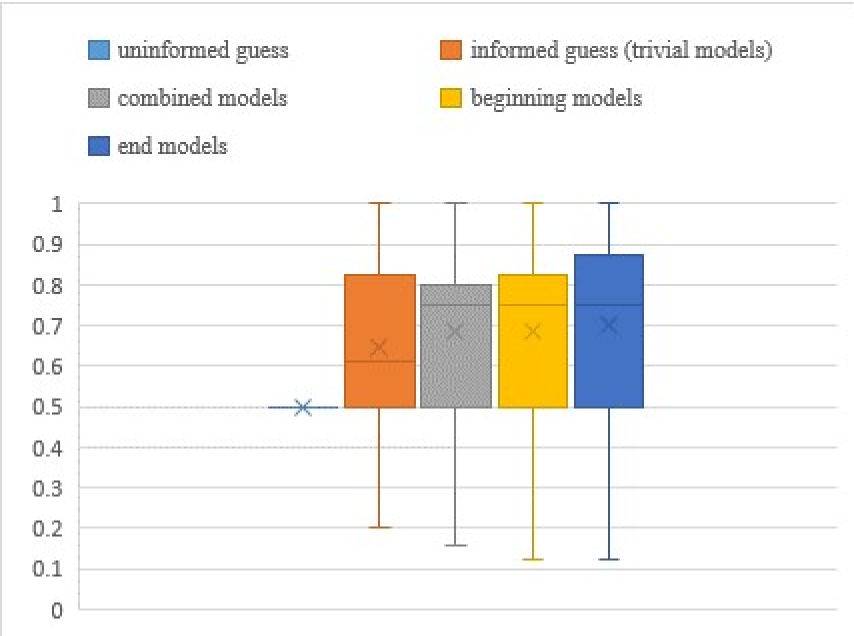


Figure 7. Boxplot of comparing the accuracy of uninformed guess and trivial models with three different RFc models (combined, beginning and end models). The accuracy of RFc models is better than uninformed guess and trivial models. End models have an average accuracy of about 70%. Beginning and combined models get similar accuracy of about 68% (details are given in Othman et al. 2019). Crosses represent mean values, boxes show 25% and 75% quantiles and median, whiskers show minimum and maximum values

Table 8. U-Test for comparing the results of uniformed guess, trivial models, combined models, beginning models and end models.

Models	p-value
uninformed guess & trivial models	0.0000
uninformed guess & combined models	0.0000
uninformed guess & beginning models	0.0000
uninformed guess & end models	0.0000
trivial models & combined models	0.0530
trivial models & beginning models	0.0519
trivial models & end models	0.0082
combined models & beginning models	0.9872
combined models & end models	0.4435
beginning models & end models	0.5147

Note: The rank of these models from the best to worst based on p-value is the end models, the beginning models, combined models, trivial models, and uninformed guess respectively.

4. Discussion

This paper provides further evidence on a strong positive effect of FFC on contribution levels in a public goods experiment. Applying a detailed analysis of content and facial expressions during the three minute long communication period, the paper illustrates several results. In conformance with previous literature both elements of communication have an effect on the contributions in the end-game phase. The strongest effect has the discussion or simple mentioning of the end-game phenomenon. Even though the discussions constituted cheap talk, an informal agreement builds a certain protection against free-riding at the end of the experiment. Therefore, it is plausible to assume an interaction with the visual identification of the co-players. Breaking an agreement after reducing the social distance by visual identification is hereby less likely. Another measure of mutual cooperation in the end-game is the length of discussions. The content analysis showed that groups with longer discussions provided on average higher contributions as compared to those with shorter ones. Combining all content variables to provide a forecast of free-riding in the last period of the experiment increases the precision as compared to naive distribution based guesses. At the same time, the analysis of facial expressions yields comparable results in terms of precision. Further results illustrate that having economic education does not affect group contributions in the experiment. Despite higher experience in this type of problems, economists do not actively lead the way out of the dilemma by narrating the solution. However, it is apparent that, as compared to females, male subjects more actively take lead in the discussion and propose the commonly agreed strategy.

With respect to the dyadic analysis approach, it has to be mentioned that the dilemma in the underlying experiment was too easy for the subjects. This yielded the variance of the dependent variable to be very low. Despite this challenging initial position, the results obtained are conclusive with respect to previous literature and the dyadic approach of analysis. They further allow predictions on the experiment-specific contribution behavior with accuracies between 70% and 80%. We are aware that besides the imbalanced data the biggest shortfall is the absence of a combined analysis. While, in the framework of this paper, this was not possible for technical reasons, the basic results hint at such codependencies, e.g. discussing the end-game behavior occurs at the end of a logical train of thoughts, the length of the discussion matters, and the end models of the facial expressions analysis perform better than the beginning models.

Based on the obtained results several improvements can be proposed on how to advance this type of research. First, it is essential to provide a larger variance of the dependent variable. With respect to the applied experimental setup, possible changes would refer to decreasing the efficiency multiplier in the experiment or limiting the information on the contributions which both should diminish cooperation rates. Second, it is crucial to reduce the noise in the data. Therefore several solutions can be thought of. One that also tackles the issue of simultaneously reducing the cooperation rates would be to decrease the duration of the FFC. Another possible solution might be the aforementioned combination of the two approaches by focusing on facial expression at specific points in time after a content relevant statement was made. This would make the analysis of large parts of the video section unnecessary and therefore decrease the noise. However, this requires more research since it is unclear how long these time frames shall be. Another way to combine the different strands of research is to focus on the prediction of specific characteristics that are known to correlate with the dependent variable. With respect to facial expression analysis, this refers to estimating the demographics of the individuals in the group. This information can then be applied to the content analysis. The combination of content variables and demographics lead to prediction rates between 74% and 90%. Third, it is necessary to increase the total number of observations in order to enhance the analysis of the data. This could enable deep learning approaches for the facial expression analysis as well as the application of machine learning on the content analysis. An experimental setup respecting these conditions shall yield deeper insights about the causes for the superiority of FFC as compared to other forms of communication, as well as help analyzing and predicting economic behavior in payoff-relevant settings.

5. Conclusion

In a nutshell, the goal of this paper was to investigate whether it is possible to predict socially sub-optimal contribution behavior in a laboratory public goods game focusing on facial expressions and simply classified contents. This would assist identifying groups in need of further interventions. The main finding is that both approaches provide improvements, yet are currently limited in their scale, especially given a very noisy environment and an imbalanced sample. Further, the paper illustrates how the amalgamation of the two applied techniques can be used to achieve better prediction accuracies. While using content measures to predict results in a simple coordination experiment can be interpreted as an extension of the existing literature, obtaining similar results with simple machine learning algorithms based on facial expressions has strong novelty character. Assuming ongoing improvements in facial recognition algorithms, the research field of human-machine interaction – and therefore the economic importance of the technology – is expected to grow. Hence, it is crucial to understand whether and how human behavior can be predicted. This paper takes a stance on the issue from the perspective of economically meaningful interactions in a stylized setting of a public goods experiment.

Al Osman, H., Eid, M., El Saddik, A., 2014. U-biofeedback: a multimedia-based reference model for ubiquitous biofeedback systems. Multimed. Tools Appl. 72, 3143–3168. https://doi.org/10.1007/s11042-013-1590-x
Altemeyer-Bartscher, M., Bershadskyy, D., Schreck, P., Timme, F., 2017. Endogenous institution formation in public good games: The effect of economic education.
Andreoni, J., Petrie, R., 2004. Public goods experiments without confidentiality: a glimpse into fund-raising. J. Public Econ. 88, 1605–1623. https://doi.org/10.1016/S0047-2727(03)00040-9
Aran, O., Gatica-Perez, D., 2011. Analysis of Group Conversations: Modeling Social Verticality, in: Computer Analysis of Human Behavior. Springer London, London, pp. 293–322. https://doi.org/10.1007/978-0-85729-994-9_11
Arechar, A.A., Dreber, A., Fudenberg, D., Rand, D.G., 2017. "I'm just a soul whose intentions are good": The role of communication in noisy repeated games. Games Econ. Behav. 104, 726–743. https://doi.org/10.1016/J.GEB.2017.06.013
Baltrusaitis, T., Robinson, P., Morency, L.P., 2016. OpenFace: An open source facial behavior analysis toolkit, in: 2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016. IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA. https://doi.org/10.1109/WACV.2016.7477553
Belot, M., Bhaskar, V., van de Ven, J., 2012. Can Observers Predict Trustworthiness? Rev. Econ. Stat. 94, 246–259. https://doi.org/10.1162/REST_a_00146
Belot, M., van de Ven, J., 2017. How private is private information? The ability to spot deception in an economic game. Exp. Econ. 20, 19–43. https://doi.org/10.1007/s10683-015-9474-8
Bochet, O., Page, T., Putterman, L., 2006. Communication and punishment in voluntary contribution experiments. J. Econ. Behav. Organ. 60, 11–26. https://doi.org/10.1016/J.JEBO.2003.06.006
Bock, O., Baetge, I., Nicklisch, A., 2014. hroot: Hamburg Registration and Organization Online Tool. Eur. Econ. Rev. 71, 117–120. https://doi.org/10.1016/J.EUROECOREV.2014.07.003
Bonnefon, J.-F., Hopfensitz, A., De Neys, W., 2017. Can We Detect Cooperators by Looking at Their Face? Curr. Dir. Psychol. Sci. 26, 276–281. https://doi.org/10.1177/0963721417693352
Bordia, P., 1997. Face-to-Face Versus Computer-Mediated Communication: A Synthesis of the Experimental Literature. J. Bus. Commun. 34, 99–118. https://doi.org/10.1177/002194369703400106
Brosig, J., 2006. Communication channels and induced behavior.
Brosig, J., Weimann, J., Ockenfels, A., 2003. The Effect of Communication Media on Cooperation. Ger. Econ. Rev. 4, 217–241. https://doi.org/10.1111/1468-0475.00080
Cason, T.N., Khan, F.U., 1999. A laboratory study of voluntary public goods provision with imperfect monitoring and communication. J. Dev. Econ. 58, 533–552. https://doi.org/10.1016/S0304-3878(98)00124-2
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature. Exp. Econ. 14, 47–83. https://doi.org/10.1007/s10683-010-9257-1
Chaudhuri, A., Graziano, S., Maitra, P., 2006. Social Learning and Norms in a Public Goods Experiment with Inter-Generational Advice1. Rev. Econ. Stud. 73, 357–380. https://doi.org/10.1111/j.1467-937X.2006.0379.x
Chen, J., Houser, D., 2017. Promises and lies: can observers detect deception in written messages. Exp. Econ. 20, 396–419. https://doi.org/10.1007/s10683-016-9488-x
Crawford, V., 1998. A Survey of Experiments on Communication via Cheap Talk. J. Econ. Theory 78, 286–298. https://doi.org/10.1006/JETH.1997.2359

Dawes, R.M., McTavish, J., Shaklee, H., 1977. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. J. Pers. Soc. Psychol. 35, 1–11. https://doi.org/10.1037/0022-3514.35.1.1
Dimoka, A., Banker, R.D., Benbasat, I., Davis, F.D., Dennis, A.R., Gefen, D., Gupta, A., Ischebeck, A., Kenning, P., Pavlou, P.A., Müller-Putz, G., Riedl, R., Brocke, J. vom, Weber, B., 2012. On the Use of Neurophysiological Tools in IS Research: Developing a Research Agenda for NeuroIS. MIS Q. 36, 679–702.
Eckel, C.C., Wilson, R.K., 2003. The Human Face of Game Theory: Trust and Reciprocity in Sequential Games.
Farrell, J., Rabin, M., 1996. Cheap Talk. J. Econ. Perspect. 10, 103–118. https://doi.org/10.1257/jep.10.3.103
Fiedler, S., Glöckner, A., Nicklisch, A., Dickert, S., 2013. Social Value Orientation and information search in social dilemmas: An eye-tracking analysis. Organ. Behav. Hum. Decis. Process. 120, 272–284. https://doi.org/10.1016/J.OBHDP.2012.07.002
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178. https://doi.org/10.1007/s10683-006-9159-4
Frank, R.H., 1988. Passions within reason: The strategic role of the emotions., Passions within reason: The strategic role of the emotions. W W Norton & Co, New York, NY, US.
Gatica-Perez, D., 2009. Automatic nonverbal analysis of social interaction in small groups: A review. Image Vis. Comput. 27, 1775–1787. https://doi.org/10.1016/J.IMAVIS.2009.01.004
George, S., Leroux, P., 2002. An Approach to Automatic Analysis of Learners' Social Behavior During Computer-Mediated Synchronous Conversations. Springer, Berlin, Heidelberg, pp. 630–640. https://doi.org/10.1007/3-540-47987-2_64
Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., 2009. Introduction, in: Neuroeconomics. Elsevier, pp. 1–12. https://doi.org/10.1016/B978-0-12-374176-9.00001-4
Han, H., Jain, A.K., Wang, F., Shan, S., Chen, X., 2018. Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach. IEEE Trans. Pattern Anal. Mach. Intell. 40, 2597–2609. https://doi.org/10.1109/TPAMI.2017.2738004
Haruvy, E., Li, S.X., McCabe, K., Twieg, P., 2017. Communication and visibility in public goods provision. Games Econ. Behav. 105, 276–296. https://doi.org/10.1016/J.GEB.2017.08.002
He, S., Offerman, T., van de Ven, J., 2017. The Sources of the Communication Gap. Manage. Sci. 63, 2832–2846. https://doi.org/10.1287/mnsc.2016.2518
Hoffman, E., McCabe, K., Smith, V.L., Hoffman, E., McCabe, K., Smith, V., 1996. Social Distance and Other-Regarding Behavior in Dictator Games. Am. Econ. Rev. 86, 653–60.
Hopfensitz, A., Mantilla, C., 2018. Emotional expressions by sports teams: An analysis of World Cup soccer player portraits. J. Econ. Psychol. https://doi.org/10.1016/J.JOEP.2018.04.008
Isaac, R.M., Walker, J.M., 1988. Communication and Free-riding Behavior: The Voluntary Contribution Mechanism. Econ. Inq. 26, 585–608. https://doi.org/10.1111/j.1465-7295.1988.tb01519.x
Jaques, N., McDuff, D., Kim, Y.L., Picard, R., 2016. Understanding and Predicting Bonding in Conversations Using Thin Slices of Facial Expressions and Body Language. Springer, Cham, pp. 64–74. https://doi.org/10.1007/978-3-319-47665-0_6
Kenning, P., Plassmann, H., 2005. NeuroEconomics: An overview from an economic perspective. Brain Res. Bull. 67, 343–354. https://doi.org/10.1016/j.brainresbull.2005.07.006
Konrad, K.A., Lohse, T., Qari, S., 2014. Deception choice and self-selection – The

importance of being earnest. J. Econ. Behav. Organ. 107, 25–39. https://doi.org/10.1016/J.JEBO.2014.07.012
Ledyard, J.O., 1995. Public Goods: A Survey of Experimental Research, in: Kagel, J.H., Roth, A. (Eds.), The Handbook of Experimental Economics. Princeton Univ. Press.
Levi, G., Hassner, T., 2015. Age and Gender Classification Using Convolutional Neural Networks.
Lopez, M.C., Villamayor-Tomas, S., 2017. Understanding the black box of communication in a common-pool resource field experiment. Environ. Sci. Policy 68, 69–79. https://doi.org/10.1016/J.ENVSCI.2016.12.002
Othman, E., Saxen, F., Bershadskyy, D., Werner, P., Al-Hamadi, A., Weimann, J., 2019. Predicting the human contribution behaviour in a public goods game from Face-to-Face Communication video using facial expressions and content analysis.
Palfrey, T., Rosenthal, H., Roy, N., 2017. How cheap talk enhances efficiency in threshold public goods games. Games Econ. Behav. 101, 234–259. https://doi.org/10.1016/J.GEB.2015.10.004
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the Ultimatum Game. Science 300, 1755–8. https://doi.org/10.1126/science.1082976
Saxen, F., Werner, P., Al-Hamadi, A., 2017. Real vs. Fake Emotion Challenge: Learning to Rank Authenticity from Facial Activity Descriptors, in: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017. pp. 3073–3078. https://doi.org/10.1109/ICCVW.2017.363
Sparks, A., Burleigh, T., Barclay, P., 2016. We can see inside: Accurate prediction of Prisoner's Dilemma decisions in announced games following a face-to-face interaction. Evol. Hum. Behav. 37, 210–216. https://doi.org/10.1016/J.EVOLHUMBEHAV.2015.11.003
ten Brinke, L., Vohs, K.D., Carney, D.R., 2016. Can Ordinary People Detect Deception After All? Trends Cogn. Sci. 20, 579–588. https://doi.org/10.1016/J.TICS.2016.05.012
Wang, J.T., Spezio, M., Camerer, C.F., 2010. Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. Am. Econ. Rev. 100, 984–1007. https://doi.org/10.1257/aer.100.3.984
Zultan, R., 2012. Strategic and social pre-play communication in the ultimatum game. J. Econ. Psychol. 33, 425–434. https://doi.org/10.1016/J.JOEP.2011.12.009

Appendix A Instructions

Instructions Experiment “Yellow“

Please read the instructions diligently. If questions arise, open the door to your cabin and remain seated. The experiment “Yellow” is carried out at the computer.
Your fellow participants will only play with you within the experiment “Yellow“.

After reading the instructions you will receive four control questions. The control questions are not considered for your final payment. As soon as you have answered the control questions the part of the experiment relevant for your final payment will start. Please be aware that either experiment „Yellow“, „Blue“ or „Red“ will be paid out. Which experiment will be eventually relevant is decided by chance.

Within the experiment, we will use laboratory dollars as the used currency. The underlying exchange rate is the following: 100 laboratory dollars = 4.5 EUR.

You and three other participants receive each **20 laboratory dollars** per round of contribution. You can contribute these laboratory dollars either to a private or a group account.

Private Account (P): The deposited laboratory dollars are being kept.

Group Account (G): Each of the four players can deposit money in this account. The sum of the deposits is doubled by the experimenter and redistributed equally to the four players. Hence, each player receives 0,5 laboratory dollars per contributed laboratory dollar.

The laboratory dollars can be split up in between the two accounts. You take your decision anonymously. None of the other players will learn how you split your laboratory dollars up. Profit of player i is calculated accordingly:

$$\text{Profit}=(20\text{-}G)+0{,}5\cdot\sum_{1}^{4}G_i$$

Please turn!

Instructions Experiment „Red"

Please read the instructions diligently. If questions arise, open the door to your cabin and remain seated. The experiment "Red" is carried out at the computer.
Your fellow participants will only play with you within the experiment „Red".

After reading the instructions you will receive four control questions. The control questions are not considered for your final payment. As soon as you have answered the control questions the part of the experiment relevant for your final payment will start. Please be aware that either experiment „Yellow", „Blue" or „Red" will be paid out. Which experiment will be eventually relevant is decided by chance.

Within the experiment, we will use laboratory dollars as the used currency. The underlying exchange rate is the following: 100 laboratory dollars = 4.5 EUR.

You and three other participants receive each **20 laboratory dollars** per round of contribution. You can contribute these laboratory dollars either to a private or a group account.

Private Account (P): The deposited laboratory dollars are being kept.

Group Account (G): Each of the four players can deposit money in this account. The sum of the deposits is doubled by the experimenter and redistributed equally to the four players. Hence, each player receives 0,5 laboratory dollars per contributed laboratory dollar.

The laboratory dollars can be split up in between the two accounts. You take your decision anonymously. None of the other players will learn how you split your laboratory dollars up. Profit of player i is calculated accordingly:

$$\text{Profit}=(20-G)+0{,}5\cdot\sum_{1}^{4} G_i$$

Video conference: Before you take your decision on how to split the laboratory dollars you will be talking to the three other players in a video conference for three minutes. During this time, you can see and talk to each other. The duration of the call can neither be reduced nor prolonged.
Subsequently to the video conference, each player makes the above described decision.

Please turn!

Instructions Experiment „Blue" *(Committing Choice)*

Please read the instructions diligently. If questions arise, open the door to your cabin and remain seated. The experiment "Blue" is carried out at the computer.
Your fellow participants will only play with you within the experiment „Blue".

After reading the instructions you will receive six control questions. The control questions are not considered for your final payment. As soon as you have answered the control questions the part of the experiment relevant for your final payment will start. Please be aware that either experiment „Yellow", „Blue" or „Red" will be paid out. Which experiment will be eventually relevant is decided by chance.

Within the experiment, we will use laboratory dollars as the used currency. The underlying exchange rate is the following: 100 laboratory dollars = 4.5 EUR.

You and three other participants receive each **20 laboratory dollars** per round of contribution. You can contribute these laboratory dollars either to a private or a group account.

Private Account (P): The deposited laboratory dollars are being kept.

Group Account (G): Each of the four players can deposit money in this account. The sum of the deposits is doubled by the experimenter and redistributed equally to the four players. Hence, each player receives 0,5 laboratory dollars per contributed laboratory dollar.

The laboratory dollars can be split up in between the two accounts. You take your decision anonymously. None of the other players will learn how you split your laboratory dollars up. Profit of player i is calculated accordingly:

$$\text{Profit}=(20\text{-}G)+0{,}5\cdot\sum_{1}^{4}G_i$$

Set-up of the video conference: At the beginning of the experiment you are asked whether you want to make the experiment this time with or without communication. Communication will be subject to a fee. To make the experiment with communication you must raise a required amount jointly as a group. This amount will pop up on your screen at the beginning of the experiment. The decision on how much you contribute will then be again taken anonymously. The deposited money for setting up communication is being deducted from your profit in the experiment „blue" at the end of it – whether communication is successfully set up or not. If the group raises the required amount, a three-minute video conference is being set up, see previous round. Otherwise, all group members have to wait for three minutes until other groups have finished their communication period, respectively. Subsequently, the decision on how to split up the laboratory dollars between private and group account are being made.

Please turn!

Instructions Experiment „Blue" (Non-Committing Choice)

Please read the instructions diligently. If questions arise, open the door to your cabin and remain seated. The experiment "Blue" is carried out at the computer.
Your fellow participants will only play with you within the experiment „Blue".

After reading the instructions you will receive six control questions. The control questions are not considered for your final payment. As soon as you have answered the control questions the part of the experiment relevant for your final payment will start. Please be aware that either experiment „Yellow", „Blue" or „Red" will be paid out. Which experiment will be eventually relevant is decided by chance.

Within the experiment, we will use laboratory dollars as the used currency. The underlying exchange rate is the following: 100 laboratory dollars = 4.5 EUR.

You and three other participants receive each **20 laboratory dollars** per round of contribution. You can contribute these laboratory dollars either to a private or a group account.

Private Account (P): The deposited laboratory dollars are being kept.

Group Account (G): Each of the four players can deposit money in this account. The sum of the deposits is doubled by the experimenter and redistributed equally to the four players. Hence, each player receives 0,5 laboratory dollars per contributed laboratory dollar.

The laboratory dollars can be split up in between the two accounts. You take your decision anonymously. None of the other players will learn how you split your laboratory dollars up. Profit of player i is calculated accordingly:

$$\text{Profit}=(20\text{-}G)+ 0{,}5\cdot \sum_{1}^{4} G_i$$

Set-up of the video conference: At the beginning of the experiment you are asked whether you want to make the experiment this time with or without communication. Communication will be subject to a fee. To make the experiment with communication you must raise a required amount jointly as a group. This amount will pop up on your screen at the beginning of the experiment. The decision on how much you contribute will then be again taken anonymously. The deposited money for setting up communication is being deducted from your profit in the experiment „blue" at the end of it – only if communication is successfully set up. If the group raises the required amount, a three-minute video conference is being set up, see previous round. Otherwise, all group members have to wait for three minutes until other groups have finished their communication period, respectively. Subsequently, the decision on how to split up the laboratory dollars between private and group account are being made.

Please turn!

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

Postal Adress: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Leibniz
Association