

# Experimental Game Theory and Its Application in Sociology and Political Science

Guest Editors: Arthur Schram, Vincent Buskens, Klarita Gërxhani,  
and Jens Großer





---

# **Experimental Game Theory and Its Application in Sociology and Political Science**



Journal of Applied Mathematics

---

## **Experimental Game Theory and Its Application in Sociology and Political Science**

Guest Editors: Arthur Schram, Vincent Buskens,  
Klarita Gërxhani, and Jens Großer



Copyright © 2015 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in “Journal of Applied Mathematics.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Saeid Abbasbandy, Iran  
Mina B. Abd-El-Malek, Egypt  
Mohamed A. Abdou, Egypt  
Subhas Abel, India  
Janos Abonyi, Hungary  
Sergei Alexandrov, Russia  
M. Montaz Ali, South Africa  
Mohammad R. Aliha, Iran  
Carlos J. S. Alves, Portugal  
Mohamad Alwash, USA  
Gholam R. Amin, Oman  
Igor Andrianov, Germany  
Boris Andrievsky, Russia  
Whye-Teong Ang, Singapore  
Abul-Fazal M. Arif, KSA  
Sabri Arik, Turkey  
Ali R. Ashrafi, Iran  
Allaberen Ashyralyev, Turkey  
Francesco Aymerich, Italy  
Jalel Azaiez, Canada  
Seungik Baek, USA  
Hamid Bahai, UK  
Olivier Bahn, Canada  
Antonio Bandera, Spain  
Jean-Pierre Barbot, France  
Mostafa Barigou, UK  
Roberto Barrio, Spain  
Henri Baudrand, France  
Alfredo Bellen, Italy  
Vasile Berinde, Romania  
Jafar Biazar, Iran  
Anjan Biswas, KSA  
Abdellah Bnouhachem, Morocco  
Gabriele Bonanno, Italy  
Walter Briec, France  
James Robert Buchanan, USA  
Humberto Bustince, Spain  
Xiao Chuan Cai, USA  
Piermarco Cannarsa, Italy  
Yijia Cao, China  
Jinde Cao, China  
Zhenfu Cao, China  
Bruno Carpentieri, Netherlands  
Ana Carpio, Spain  
Alexandre Carvalho, Brazil

Song Cen, China  
Tai-Ping Chang, Taiwan  
Shih-sen Chang, China  
Wei-Der Chang, Taiwan  
Shuenn-Yih Chang, Taiwan  
Kripasindhu Chaudhuri, India  
Yuming Chen, Canada  
Jianbing Chen, China  
Xinkai Chen, Japan  
Rushan Chen, China  
Ke Chen, UK  
Zhang Chen, China  
Zhi-Zhong Chen, Japan  
Ru-Dong Chen, China  
Hui Cheng, China  
Chin-Hsiang Cheng, Taiwan  
Eric Cheng, Hong Kong  
Ching-Hsue Cheng, Taiwan  
Qi Cheng, USA  
Jin Cheng, China  
Francisco Chiclana, UK  
Jen-Tzung Chien, Taiwan  
Chongdu Cho, Korea  
Han H. Choi, Republic of Korea  
Md S. Chowdhury, Malaysia  
Yu-Ming Chu, China  
Hung-Yuan Chung, Taiwan  
Angelo Ciaramella, Italy  
Pedro J. Coelho, Portugal  
Carlos Conca, Chile  
Vitor Costa, Portugal  
Lirong Cui, China  
Jie Cui, USA  
Livija Cveticanin, Serbia  
Gaoliang Dai, Germany  
Hua Dai, China  
Binxiang Dai, China  
Michael Defoort, France  
Hilmi Demiray, Turkey  
Mingcong Deng, Japan  
Youjun Deng, France  
Orazio Descalzi, Chile  
Raffaele Di Gregorio, Italy  
Kai Diethelm, Germany  
Daniele Dini, UK

Urmila Diwekar, USA  
Vit Dolejsi, Czech Republic  
BoQing Dong, China  
Rodrigo W. dos Santos, Brazil  
Wenbin Dou, China  
Rafael Escarela-Perez, Mexico  
Magdy A. Ezzat, Egypt  
Meng Fan, China  
Ya Ping Fang, China  
István Faragó, Hungary  
Didier Felbacq, France  
Ricardo Femat, Mexico  
Antonio J. M. Ferreira, Portugal  
George Fikioris, Greece  
Michel Fliess, France  
Marco A. Fontelos, Spain  
Dimitris Fotakis, Serbia  
Tomonari Furukawa, USA  
Maria Gandarias, Spain  
Xiao-wei Gao, China  
Huijun Gao, China  
Xin-Lin Gao, USA  
Naiping Gao, China  
Laura Gardini, Italy  
Winston Garira, South Africa  
Leonidas N. Gergidis, Greece  
Bernard J. Geurts, Netherlands  
Sandip Ghosal, USA  
Dibakar Ghosh, India  
Pablo González-Vera, Spain  
Alexander N. Gorban, UK  
Keshlan S. Govinder, South Africa  
Said R. Grace, Egypt  
Jose L. Gracia, Spain  
Maurizio Grasselli, Italy  
Zhi-Hong Guan, China  
Nicola Guglielmi, Italy  
Frédéric Guichard, Canada  
Kerim Guney, Turkey  
Shu-Xiang Guo, China  
Vijay Gupta, India  
Saman K. Halgamuge, Australia  
Ridha Hambli, France  
Abdelmagid S. Hamouda, Qatar  
Bo Han, China

Maoan Han, China	Svein Kleiven, Sweden	Yansheng Liu, China
Pierre Hansen, Canada	Kazutake Komori, Japan	Zhijun Liu, China
Ferenc Hartung, Hungary	Vassilis Kostopoulos, Greece	Kang Liu, USA
Xiao-Qiao He, China	Jisheng Kou, China	Weiqing Liu, China
Yuqing He, China	Roberto A. Kraenkel, Brazil	Tao Liu, China
Nicolae Herisanu, Romania	Kannan Krithivasan, India	Shutian Liu, China
Onesimo Hernandez-Lerma, Mexico	Vadim A. Krysko, Russia	Fei Liu, China
Luis Javier Herrera, Spain	Jin L. Kuang, Singapore	Chongxin Liu, China
J. C. D. Hoenderkamp, Netherlands	Gennady M. Kulikov, Russia	Jose L. López, Spain
Thomas Höhne, Germany	Venugopal Kumaran, India	Shiping Lu, China
Wei-Chiang Hong, Taiwan	Mirosław Lachowicz, Poland	Hongbing Lu, China
Sun-Yuan Hsieh, Taiwan	Hak-Keung Lam, UK	Benzhuo Lu, China
Ying Hu, France	Tak-Wah Lam, Hong Kong	Hao Lu, USA
Ning Hu, Japan	Heung-Fai Lam, Hong Kong	Yuan Lu, China
Jianguo Huang, China	Luciano Lamberti, Italy	Henry Horng-Shing Lu, Taiwan
Dan Huang, China	Peter G L Leach, Cyprus	Gilles Lubineau, KSA
Ting-Zhu Huang, China	Usik Lee, Republic of Korea	Zhen Luo, China
Zhilong L. Huang, China	Jaehong Lee, Korea	Lifeng Ma, China
Zhenkun Huang, China	Wen-Chuan Lee, Taiwan	Changfeng Ma, China
Asier Ibeas, Spain	Myung-Gyu Lee, Korea	Li Ma, China
Mustafa Inc, Turkey	Jinsong Leng, China	Ruyun Ma, China
Gerardo Iovane, Italy	Xiang Li, China	Shuping Ma, China
Anuar Ishak, Malaysia	Qingdu Li, China	Krzysztof Magnucki, Poland
Takeshi Iwamoto, Japan	Wenlong Li, Hong Kong	Nazim I. Mahmudov, Turkey
George Jaiani, Georgia	Shuai Li, Hong Kong	Oluwole D. Makinde, South Africa
GunHee Jang, Korea	Lixiang Li, China	Flavio Manenti, Italy
Zhongxiao Jia, China	Wan-Tong Li, China	Francisco J. Marcellán, Spain
Daqing Jiang, China	Shiyong Li, China	Vasile Marinca, Romania
Haijun Jiang, China	Qin Li, China	Giuseppe Marino, Italy
Jianjun Jiao, China	Hua Li, Singapore	Guiomar Martín-Herrán, Spain
Zhen Jin, China	Yan Liang, China	Carlos Martín-Vide, Spain
Xing Jin, China	Jin Liang, China	Alessandro Marzani, Italy
Lucas Jódar, Spain	Dongfang Liang, UK	Nikos E. Mastorakis, Bulgaria
Zlatko Jovanoski, Australia	Ching-Jong Liao, Taiwan	Nicola Mastronardi, Italy
Tadeusz Kaczorek, Poland	Teh-Lu Liao, Taiwan	Panayotis Takis Mathiopoulos, Greece
Dongsheng Kang, China	Yong-Cheng Lin, China	Gianluca Mazzini, Italy
Sangmo Kang, USA	Gao Lin, China	Michael McAleer, Netherlands
Ido Kanter, Israel	Yiping Lin, China	Stephane Metens, France
Abdul Hamid Kara, South Africa	Chong Lin, China	Michael Meylan, Australia
Dimitrios A. Karras, Greece	Ignacio Lira, Chile	Efren Mezura-Montes, Mexico
Ihsan Kaya, Turkey	Sheng Liu, Australia	Fan Min, China
Dogan Kaya, Turkey	Zhengrong Liu, China	Alain Miranville, France
Chaudry M. Khaliq, South Africa	Zhuangjian Liu, Singapore	Ram N. Mohapatra, USA
Waqar A. Khan, Pakistan	Meiqin Liu, China	Syed A. Mohiuddine, KSA
Khalil Khanafer, USA	Peter Liu, Taiwan	Gisele Mophou, France
Hyunsung Kim, Korea	Peide Liu, China	José Morais, Portugal
Younjea Kim, Republic of Korea	Pengfei Liu, China	Cristinel Mortici, Romania
Jong Hae Kim, Republic of Korea	Chein-Shan Liu, Taiwan	Emmanuel Moulay, France

Jaime E. Munoz Rivera, Brazil	Samir H. Saker, Egypt	Charalampos Tsitouras, Greece
Javier Murillo, Spain	R. Sakthivel, Republic of Korea	Antonia Tulino, USA
Roberto Natalini, Italy	Juan J. Salazar González, Spain	Stefano Ubertini, Italy
Srinivasan Natesan, India	Miguel A. F. Sanjuan, Spain	Jeong S. Ume, Korea
Tatsushi Nishi, Japan	Bogdan Sasu, Romania	Alexandrina Untaroiu, USA
Mustapha Nourelfath, Canada	Richard Saurel, France	Sergey Utyuzhnikov, UK
Andreas "Ochsner, Australia	Wolfgang Schmidt, Germany	Kuppapalle Vajravelu, USA
Wlodzimierz Ogryczak, Poland	Mehmet Sezer, Turkey	Alvaro Valencia, Chile
Roger Ohayon, France	Naseer Shahzad, KSA	Olaf van der Sluis, Netherlands
Javier Oliver, Spain	Pengjian Shang, China	Erik Van Vleck, USA
Soontorn Oraintara, USA	M. Shariyat, Iran	Ezio Venturino, Italy
Donal O'Regan, Ireland	Hui-Shen Shen, China	Jesus Vigo-Aguiar, Spain
Martin Ostoja-Starzewski, USA	Jian Hua Shen, China	Michael N. Vrahatis, Greece
Turgut "Oziş, Turkey	Yong Shi, China	Michael Vynnycky, Sweden
Claudio Padra, Argentina	Yasuhide Shindo, Japan	Xiang Wang, China
Vincent Pagneux, France	Patrick Siarry, France	Yaonan Wang, China
Reinaldo Martinez Palhares, Brazil	Fernando Simões, Portugal	Shuming Wang, Singapore
Quanke Pan, China	Theodore E. Simos, Greece	Youqing Wang, China
Endre Pap, Serbia	Georgios Sirakoulis, Greece	Dongqing Wang, China
Sehie Park, Korea	Robert Smith?, Canada	Peiguang Wang, China
Manuel Pastor, Spain	Francesco Soldovieri, Italy	Yuh-Rau Wang, Taiwan
Giuseppe Pellicane, South Africa	Abdel-Maksoud A. Soliman, Egypt	Heng Wang, Singapore
Francesco Pellicano, Italy	Xinyu Song, China	Baolin Wang, China
Juan Manuel Peña, Spain	Qiankun Song, China	Mingxin Wang, China
Jian-Wen Peng, China	Yuri N. Sotskov, Belarus	Qing-Wen Wang, China
Ricardo Perera, Spain	Ivanka Stamova, USA	Guangchen Wang, China
Malgorzata Peszynska, USA	Niclas Strömberg, Sweden	Junzo Watada, Japan
Allan C. Peterson, USA	Ray K.L. Su, Hong Kong	Wei Wei, China
Andrew Pickering, Spain	Housheng Su, China	Junjie Wei, China
Somyot Plubtieng, Thailand	Chengjun Sun, Hong Kong	Jinjia Wei, China
Hector Pomares, Spain	Jitao Sun, China	Guoliang Wei, China
Roland Potthast, UK	Wenyu Sun, China	Li Weili, China
Kerehalli V. Prasad, India	W. Y. Szeto, Hong Kong	Martin Weiser, Germany
Radu-Emil Precup, Romania	Toshio Tagawa, Japan	Frank Werner, Germany
Mario Primicerio, Italy	Ying Tan, China	Man Leung Wong, Hong Kong
Morteza Rafei, Netherlands	San-Yi Tang, China	Xiangjun Wu, China
P. Rattanadecho, Thailand	XianHua Tang, China	Cheng Wu, China
Laura Rebollo-Neira, UK	Nasser-Eddine Tatar, KSA	Shi-Liang Wu, China
Mahmoud M. Reda Taha, USA	Zhidong Teng, China	Wei Wu, China
Roberto Renò, Italy	Engang Tian, China	Qingbiao Wu, China
Bruno G. M. Robert, France	Alexander Timokha, Norway	Yuqiang Wu, China
J. A. Rodriguez-Velázquez, Spain	Aydin Tiriyaki, Turkey	Min-Hsien Wu, Taiwan
Ignacio Rojas, Spain	Yiying Tong, USA	Xian Wu, China
Carla Roque, Portugal	Hossein Torkaman, Iran	Tiecheng Xia, China
Debasish Roy, India	Mariano Torrisi, Italy	Yonghui Xia, China
Imre J. Rudas, Hungary	Juan J. Trujillo, Spain	Xian Xian-Yong, China
Abbas Saadatmandi, Iran	Jung-Fa Tsai, Taiwan	Xuejun Xie, China
Kunihiko Sadakane, Japan	George Tsiatas, Greece	Gongnan Xie, China





Daoyi Xu, China  
Yuesheng Xu, USA  
Zhiqiang Xu, China  
Wei Xu, China  
Fuzhen Xuan, China  
Gregory S. Yablonsky, USA  
Chao Yan, USA  
Chao Yang, China  
Chunyu Yang, China  
Guowei Yang, China  
Bin Yang, China  
Xiao-Jun Yang, China  
Dar-Li Yang, Taiwan  
Na Yang, China  
Chao-Tung Yang, Taiwan  
Zhichun Yang, China  
Her-Terng Yau, Taiwan  
Wei-Chang Yeh, Taiwan  
Guan H. Yeoh, Australia

Chih-Wei Yi, Taiwan  
Simin Yu, China  
Bo Yu, China  
Xiaohui Yuan, China  
Gonglin Yuan, China  
Jinyun Yuan, Brazil  
Rafal Zdunek, Poland  
Ashraf Zenkour, Egypt  
Guisheng Zhai, Japan  
Jianming Zhan, China  
Meng Zhan, China  
Zhengqiu Zhang, China  
Jiangang Zhang, China  
Ke Zhang, China  
Long Zhang, China  
Sheng Zhang, China  
Jifeng Zhang, China  
Heping Zhang, China  
Henggui Zhang, UK

Liang Zhang, China  
Jingxin Zhang, Australia  
Zhihua Zhang, China  
Shan Zhao, USA  
Chongbin Zhao, Australia  
Baosheng Zhao, China  
Yun-Bo Zhao, China  
Renat Zhdanov, USA  
Dong Zheng, USA  
Huaichun Zhou, China  
Bin Zhou, China  
Shangbo Zhou, China  
Huamin Zhou, China  
Quanxin Zhu, China  
Xinqun Zhu, Australia  
William Zhu, China  
Goangseup Zi, Korea  
Zhiqiang Zuo, China

# Contents

---

**Experimental Game Theory and Its Application in Sociology and Political Science**, Arthur Schram, Vincent Buskens, Klarita Gërxhani, and Jens Großer  
Volume 2015, Article ID 280789, 2 pages

**The Strategic Role of Nonbinding Communication**, Luis A. Palacio, Alexandra Cortés-Aguilar, and Manuel Muñoz-Herrera  
Volume 2015, Article ID 910614, 11 pages

**Intermediaries in Trust: Indirect Reciprocity, Incentives, and Norms**, Giangiacomo Bravo, Flaminio Squazzoni, and Károly Takács  
Volume 2015, Article ID 234528, 12 pages

**Altruism, Noise, and the Paradox of Voter Turnout: An Experimental Study**, Sarah A. Tulman  
Volume 2015, Article ID 972753, 22 pages

**Preference for Efficiency or Confusion? A Note on a Boundedly Rational Equilibrium Approach to Individual Contributions in a Public Good Game**, Luca Corazzini and Marcelo Tyszler  
Volume 2015, Article ID 961930, 8 pages

**Competition, Income Distribution, and the Middle Class: An Experimental Study**, Bernhard Kittel, Fabian Paetzel, and Stefan Traub  
Volume 2015, Article ID 303912, 15 pages

## Editorial

# Experimental Game Theory and Its Application in Sociology and Political Science

**Arthur Schram,<sup>1,2</sup> Vincent Buskens,<sup>3</sup> Klarita Gërxhani,<sup>4</sup> and Jens Großer<sup>5</sup>**

<sup>1</sup>*CREED, University of Amsterdam, Plantage Muidergracht 14, 1018 TV Amsterdam, Netherlands*

<sup>2</sup>*Robert Schuman Centre for Advanced Studies, European University Institute, Fiesole, Italy*

<sup>3</sup>*Department of Sociology/ICS, Utrecht University, Padualaan 14, 3584 CH Utrecht, Netherlands*

<sup>4</sup>*Department of Political and Social Sciences, European University Institute, Via Roccettini 9, San Domenico di Fiesole, 50014 Fiesole, Italy*

<sup>5</sup>*Department of Political Science, Florida State University, Bellamy Hall 531, Tallahassee, FL 32306-2230, USA*

Correspondence should be addressed to Arthur Schram; [schram@uva.nl](mailto:schram@uva.nl)

Received 13 August 2015; Accepted 16 August 2015

Copyright © 2015 Arthur Schram et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Game theory, laboratory experiments, and field experiments are common and powerful tools in many social sciences [1]. However, applications in Sociology and Political Science remain scarce and scattered [2]. Yet, the combination of game theory with controlled experiments provides a powerful tool to better understand social and political processes, for example, [3–5]. The mathematical structure offered by game theory and the control offered by an experimental environment allow the researcher to isolate sociological and/or political phenomena to study their development and their effects. The relationship between game theory and experiments is twofold. On the one hand, game theory provides solid ground on which to design an experiment and a formal benchmark that serves as a measuring rod for a structured analysis of observed behavior. On the other hand, experiments can be used to test equilibrium predictions and to pinpoint shortcomings of theory as well as point to directions in which the theory can be adapted.

The aim of the special issue is to encourage original research that seeks to study sociological or political phenomena using laboratory experiments that are based on game theoretical benchmarks and that seek mathematical modeling

of game theoretical arguments to inspire experiments in the fields of Sociology and Political Science, and vice versa.

In a research article of the special issue, G. Bravo et al. experimentally study whether intermediaries can positively influence cooperation between a trustor and a trustee in an investment or trust game. Another article by L. A. Palacio et al. develops a game theoretical foundation for experimental investigations of the strategic role in games with nonbinding communication. In another article, L. Corazzini and M. Tyszler employ quantal response equilibrium (QRE) to find out the extent of confusion and efficiency motives of laboratory participants in their decisions to contribute to public good. The article by S. A. Tulman utilizes QRE (i.e., noisy decision-making) and altruism-motivated players to investigate the “paradox of voter turnout” in a participation game experiment. Finally, in another article, B. Kittel et al. present a laboratory study in which they examine the role of the middle class on income distribution within the framework of a contest game.

We hope that the selection of articles in this special issue will help to inspire scholars in Sociology and Political Science to add mathematics to their tool box and adopt game theory and experimentation in their research methodology.

## Acknowledgment

We sincerely thank all the authors and reviewers who contributed greatly to the success of the special issue.

Arthur Schram  
Vincent Buskens  
Klarita Gërxhani  
Jens Großner

## References

- [1] A. Falk and J. J. Heckman, “Lab experiments are a major source of knowledge in the social sciences,” *Science*, vol. 326, no. 5952, pp. 535–538, 2009.
- [2] M. Jackson and D. R. Cox, “The principles of experimental design and their application in sociology,” *Annual Review of Sociology*, vol. 39, pp. 27–49, 2013.
- [3] R. Corten and V. Buskens, “Co-evolution of conventions and networks: an experimental study,” *Social Networks*, vol. 32, no. 1, pp. 4–15, 2010.
- [4] K. Gërxhani, J. Brandts, and A. Schram, “The emergence of employer information networks in an experimental labor market,” *Social Networks*, vol. 35, no. 4, pp. 541–560, 2013.
- [5] J. Großner and A. Schram, “Public opinion polls, voter turnout, and welfare: an experimental study,” *American Journal of Political Science*, vol. 54, no. 3, pp. 700–717, 2010.

## Research Article

# The Strategic Role of Nonbinding Communication

Luis A. Palacio,<sup>1</sup> Alexandra Cortés-Aguilar,<sup>1</sup> and Manuel Muñoz-Herrera<sup>2</sup>

<sup>1</sup>Escuela de Economía y Administración, Universidad Industrial de Santander, Calle 9 con 27, Bucaramanga, Colombia

<sup>2</sup>ICS, Faculty of Behavioural Social Sciences, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, Netherlands

Correspondence should be addressed to Manuel Muñoz-Herrera; [e.m.munoz.herrera@rug.nl](mailto:e.m.munoz.herrera@rug.nl)

Received 1 December 2014; Revised 27 March 2015; Accepted 28 March 2015

Academic Editor: Jens Großer

Copyright © 2015 Luis A. Palacio et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies the conditions that improve bargaining power using threats and promises. We develop a model of strategic communication, based on the *conflict game with perfect information*, in which a noisy *commitment message* is sent by a better-informed sender to a receiver who takes an action that determines the welfare of both. Our model captures different levels of aligned-preferences, for which classical games such as *stag hunt*, *hawk-dove*, and *prisoner's dilemma* are particular cases. We characterise the Bayesian perfect equilibrium with nonbinding messages under *truth-telling beliefs* and *sender's bargaining power* assumptions. Through our equilibrium selection we show that the less conflict the game has, the more informative the equilibrium signal is and less credibility is necessary to implement it.

## 1. Introduction

Bargaining power refers to the relative ability that a player has in order to exert influence upon others to improve her own wellbeing. It is related also to idiosyncratic characteristics such as patience, so that a player turns the final outcome into her favour if she has better outside options or if she is more patient [1]. In addition, Schelling [2] described bargaining power as the chance to cheat and bluff, the ability to set the best price for oneself. For instance, when the union says to the management in a firm, “we will go on strike if you do not meet our demands,” or when a nation announces that any military provocation will be responded with nuclear weapons, it is clear that communication has been used with a strategic purpose, to gain bargaining power.

In bargaining theory, *strategic moves* are actions taken prior to playing a subsequent game, with the aim of changing the available strategies, information structure, or payoff functions. The aim is to change the opponent's beliefs, making it credible that the position is unchangeable. Following Selten [3], the formal notion of credibility is subgame perfectness. (Schelling developed the notion of credibility as the outcome that survives iterated elimination of weakly dominated strategies. We know that, in the context of generic extensive-form games with complete and perfect information, this procedure

does indeed work (see [4]).) Nevertheless, we argue that if a message is subgame perfect, then it is neither a threat nor a promise. Consider the following example: a union says to management: “If you increase our salaries, we will be grateful.” In such case, credibility is not in doubt, but we could hardly call this a promise or a threat. Schelling [2] denotes fully credible messages as *warnings*; and we follow this differentiation to threats and promises.

Commitment theory was proposed by Schelling [2] (for a general revision of Schelling's contribution to economic theory, see Dixit [4] and Myerson [5]), who introduced a tactical approach for communication and credibility inside game theory. Hirshliefer [6, 7] and Klein and O'Flaherty [8] worked on the analysis and characterisation of strategic moves in the standard game theory framework. In the same way, Crawford and Sobel [9] formally showed that an informed agent could reveal his information in order to induce the uninformed agent to make a specific choice.

There are three principal reasons for modelling pre-play communication: information disclosure (signalling), coordination goals (cheap-talk), and strategic influence (in Schelling's sense). Following Farrell [10] and Farrell and Rabin [11], the main problem in modelling nonbinding messages is the “babbling equilibrium,” where statements mean nothing. However, they showed that cheap talk can convey

information in a general signalling environment, displaying a particular equilibrium in which statements are meaningful. In this line, Rabin [12] developed *credible message profiles*, looking for a meaningful communication equilibrium in cheap-talk games.

Our paper contributes to the strategic communication literature in three ways. First, we propose a particular characterisation of *warnings*, *threats*, and *promises* in the *conflict game with perfect information*, as mutually exclusive categories. For this aim, we first define a sequential protocol in the  $2 \times 2$  *conflict game* originally proposed by Baliga and Sjöström [13]. This benchmark game is useful because it is a stylised model that captures different levels of aligned-preferences, for which classical games such as *stag hunt*, *hawk-dove*, and *prisoner's dilemma* are particular cases.

Second, we model strategic moves with nonbinding messages, showing that choosing a particular message and its credibility are related to the level of conflict. In this way, the *conflict game with nonbinding messages* captures a bargaining situation where people talk about their intentions, by simply using cheap talk. More precisely, we analyse a game where a second player (the sender) can communicate her action plan to the first mover (the receiver). (To avoid confusion and gender bias, the sender will be denoted as “she,” and the receiver as “he.”) In fact, the sender must decide after she observes the receiver’s choice, but the commitment message is a preplay move.

Third, we introduce a simple parameterisation that can be used as a baseline for experimental research. By means of this model it is possible to study how, in a bargaining environment, information and communication influence the power one of the parts may have. In other words, this addresses the following: the logic supporting Nash equilibrium is that each player is thinking, given what the other does, what is the best he could do. Players, in a first stance, cannot influence others’ behaviour. On the contrary, Schelling [2] argues that players may consider what they can do, as a preplay move, to influence (i.e., manipulate) the behaviour of their counterpart and turn their payoffs in their favour. Therefore, our behavioural model provides a framework where it is possible to (experimentally) study the strategic use of communication in order to influence others, under different levels of conflict.

We analyse conceptually the importance of three essential elements of commitment theory: (i) the choice of a response rule, (ii) the announcement about future actions, and (iii) the credibility of messages. We answer the following questions: *what is the motivation behind threats and promises?* and *can binding messages improve the sender’s bargaining power?* In this paper, *threats* and *promises* are defined as a second mover self-serving announcement, committing in advance how she will play in all conceivable eventualities, as long as it specifies at least one action that is not her best response (see [4, 7]). With this definition, we argue that binding messages improve the sender’s bargaining power in the *perfect information conflict game*, even when it is clear that by assuming binding messages we avoid the problem of credibility.

The next step is to show that credibility is related to the probability that the sender fulfills the action specified in

TABLE 1: The  $2 \times 2$  conflict game.

		Player 2	
		$d$	$h$
Player 1	$d$	1, 1	$x, y$
	$h$	$y, x$	0.25, 0.25

the nonbinding message. For this, we highlight that players share a common language, and the literal meaning must be used to evaluate whether a message is credible or not. Hence, the receiver has to believe in the literal meaning of announcements if and only if it is highly probable to face the truth. Technically, we capture this intuition in two axioms: *truth-telling beliefs* and *the sender’s bargaining power*. We ask, *are nonbinding messages a mechanism to improve the sender’s bargaining power?* and *how much credibility is necessary for a strategic move to be successful?* In equilibrium, we can prove that nonbinding messages will convey private information when the conflict is low. On the other hand, if the conflict is high, there are too strong incentives to lie, and cheap talk becomes meaningless. However, even in the worse situation, the nonbinding messages can transmit some meaning in equilibrium if the players focus on the possibility of fulfilling threats and promises.

The paper is organised as follows. In Section 2, the *conflict game* is described. In Section 3 the conditioned messages will be analysed, and the definitions of *threats* and *promises* are presented. Section 4 presents the model with nonbinding messages, showing the importance of response rules, messages, and credibility to improve the sender’s bargaining power. Finally, Section 5 concludes.

## 2. The $2 \times 2$ Conflict Game

The  $2 \times 2$  *conflict game* is a noncooperative symmetric environment. There are two decision makers in the set of players,  $N = \{1, 2\}$ . (In this level of simplicity, players’ identity is not relevant, but since the purpose is to model Schelling’s strategic moves, in the following sections player 2 is going to be a sender of *commitment messages*.) Players must choose an action  $s_i \in S_i = \{d, h\}$ , where  $d$  represents being dove (peaceful negotiator) and  $h$  being hawk (aggressive negotiator). The utility function  $u_i(s_1, s_2)$  for player  $i$  is defined by the payoffs matrix in Table 1, where rows correspond to player 1 and columns correspond to player 2.

Note that both mutual cooperation and mutual defection lead to equal payoffs, and the combination of strategies  $(d, d)$  is always Pareto optimal. In the same way, the combination of strategies  $(h, h)$  is not optimal and can only be understood as the disagreement point. Assuming that  $y \geq x$ , payoffs are unequal when a player behaves aggressively and the other cooperates, given that the player who plays aggressively has an advantage over his/her opponent. In addition, we will assume that  $x \neq 0.25$  and  $y \neq 1$  to avoid the multiplicity of irrelevant equilibria. Therefore, it will always be preferred that the opponent chooses  $d$ . To have a parameterisation that serves as a baseline for experimental design, it is desirable to fix  $x \in [0, 0.5]$  and  $y \in [0.5, 1.5]$  within these intervals,



TABLE 2: Nash equilibria in the conflict game.

	$(s_1^*, s_2^*)$	$(U_1^*, U_2^*)$	Pareto optimal
C1	$(d, d)$	$(1, 1)$	Yes
C2	$(d, d)$ $(h, h)$	$(1, 1)$ $(0.25, 0.25)$	Yes No
C3	$(d, h)$ $(h, d)$	$(x, y)$ $(y, x)$	Yes Yes
C4	$(h, h)$	$(0.25, 0.25)$	No

because if they are modelled as random variables with uniform distribution we would have four games with the same probability of occurring.

Under these assumptions, the  $2 \times 2$  conflict game has four particular cases that, according to Hirshliefer [6], can be ordered by their level of conflict or affinity in preferences:

- (1) *Level of conflict 1 (C1)*: if  $y < 1$  and  $x > 0.25$ , there is no conflict in this game because cooperating is a dominant strategy.
- (2) *Level of conflict 2 (C2)*: if  $y < 1$  and  $x < 0.25$ , this is the so-called *stag hunt* game, which formalises the idea that lack of trust may lead to disagreements.
- (3) *Level of conflict 3 (C3)*: if  $y > 1$  and  $x > 0.25$ , depending on the history used to contextualise it, this game is known as either *hawk-dove* or *chicken* game. Both anticipation and dissuasion are modelled here, where fear of consequences makes one of the parts give up.
- (4) *Level of conflict 4 (C4)*: if  $y > 1$  and  $x < 0.25$ , this is the classic *prisoners dilemma*, where individual incentives lead to an inefficient allocation of resources.

Based on the system of incentives, it is possible to explain why these games are ordered according to their level of conflict, from lowest to highest (see Table 2). In the C1 game the players' preferences are well aligned and there is no coordination problem because the Nash equilibrium is unique in dominant strategies. Therefore, a rational player will always choose to cooperate  $d$ , which will lead to the outcome that is Pareto optimal. In the C2 game mutual cooperation  $(d, d)$  is a Nash equilibrium, but it is not unique in pure strategies. The problem lies in coordinating on either a Pareto dominant equilibrium  $(d, d)$  or a risk dominant equilibrium  $(h, h)$ . In other words, negotiating as a dove implies a higher risk and will only take place if a player believes that the adversary will do the same. This is the reason why it is possible to state that lack of trust between the parties may lead to the disagreement point.

The C3 game portrays an environment with higher levels of conflict, since there are two equilibria with unequal payoffs. In other words, players face two problems, a distributive and a coordination one. If only one of the players chooses to behave aggressively, this will turn the result in his/her favour, but it is impossible to predict who will be aggressive and who will cooperate. In this  $2 \times 2$  environment there is no clear criterion to predict the final outcome and therefore the behaviour.

TABLE 3: The conflict game: illustrative cases.

(a) C1: $y - x = 0$			
1	2		
	$d$	$h$	
	$d$	1, 1	0.5, 0.5
$h$	0.5, 0.5	0.25, 0.25	
(b) C2: $y - x = 0.5$			
1	2		
	$d$	$h$	
	$d$	1, 1	0, 0.5
$h$	0.5, 0	0.25, 0.25	
(c) C3: $y - x = 1$			
1	2		
	$d$	$h$	
	$d$	1, 1	0.5, 1.5
$h$	1.5, 0.5	0.25, 0.25	
(d) C4: $y - x = 1.5$			
1	2		
	$d$	$h$	
	$d$	1, 1	0, 1.5
$h$	1.5, 0	0.25, 0.25	

The last game is the classical social dilemma about the limitations of rational behaviour to allocate resources efficiently. The C4 game is classified as the most conflictive one because the players are faced with a context where the rational choice clearly predicts that the disagreement point will be reached. Additionally, we will argue along this document that changing incentives to achieve mutual cooperation is not a simple task in this bargaining environment.

Until this moment we have used equilibrium unicity and its optimality to argue that the games are ordered by their level of conflict. However, it is possible to understand the difference in payoffs ( $y - x$ ) as a proxy of the level of conflict. In other words, the difference in payoffs between the player who takes the advantage by playing aggressively and the player who is exploited for cooperating is large, we can state that the incentives lead players to a preference to behave aggressively (see the illustrative cases in Table 3).

### 3. Response Rules and Commitment Messages

We consider now the conflict game with a sequential decision making protocol. The idea is to capture a richer set of strategies that allows us to model threats and promises as self-serving messages. In addition, the set of conditioned strategies include the possibility of implementing ordinary commitment, because a simple unconditional message is always available for the sender.

Schelling [2] distinguishes between two different types of strategic moves: ordinary commitments and threats. An ordinary commitment is the possibility of playing first,

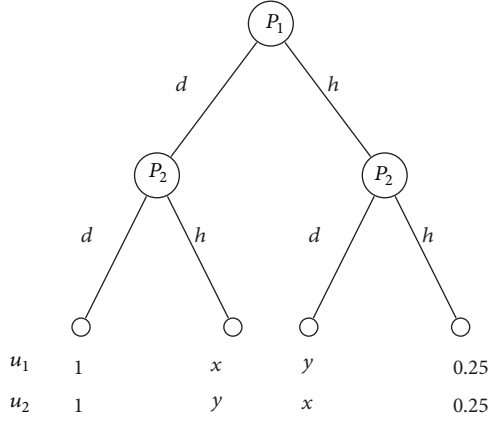


FIGURE 1: The conflict game with perfect information.

announcing that a decision has already been made and it is impossible to be changed, which forces the opponent to make the final choice. On the other hand, threats are second player moves, where she convincingly pledges to respond to the opponent's choice in a specified contingent way (see [7]).

**3.1. The Conflict Game with Perfect Information.** Suppose that player 1 moves first and player 2 observes the action made by player 1 and makes his choice. In theoretical terms, this is a switch from the  $2 \times 2$  strategic game to the extensive game with perfect information in Figure 1. A strategy for player 2 is a function that assigns an action  $s_2 \in \{d, h\}$  to each possible action of player 1,  $s_1 \in \{d, h\}$ . Thus, the set of strategies for player 2 is  $S_2 = \{dd, dh, hd, hh\}$ , where  $s_2 = s_{2d}s_{2h}$  represents a possible reaction rule, such that the first component  $s_{2d}$  denotes the action that will be carried out if player 1 plays  $d$ , and the second component  $s_{2h}$  is the action in case that 1 plays  $h$ . The set of strategies for player 1 is  $S_1 = \{d, h\}$ .

In this sequential game with perfect information a strategy profile is  $(s_1, s_2)$ . Therefore, the utility function  $u_i(s_1, s_2)$  is defined by  $u_i(d, s_{2d}s_{2h}) = u_i(d, s_{2d})$  and  $u_i(h, s_{2d}s_{2h}) = u_i(h, s_{2h})$ , based on the  $2 \times 2$  payoff matrix presented before. As the set of strategy profiles becomes wider, the predictions based on the Nash equilibrium are less relevant. Thus, in the conflict game with perfect information the applicable equilibrium concept is the subgame perfect Nash equilibrium (SPNE).

**Definition 1 (SPNE).** The strategy profile  $(s_1^*, s_2^*)$  is a SPNE in the conflict game with perfect information if and only if  $u_2(s_1, s_2^*) \geq u_2(s_1, s_2)$  for every  $s_2 \in S_2$  and for every  $s_1 \in S_1$ ; and  $u_1(s_1^*, s_2^*) \geq u_1(s_1, s_2^*)$  for every  $s_1 \in S_1$ .

The strategy  $s_2^* = s_{2d}^*s_{2h}^*$  represents the best response for player 2 in every subgame. In the same way, the strategy  $s_1^*$  is the best response for player 1 when player 2 chooses  $s_2^*$ . By definition and using the payoffs assumptions, it is clear that the strategy  $s_2^* = x_{2d}^*s_{2h}^*$  is the unique weakly dominant strategy for player 2 and, in consequence, the reason for player 1 to forecast his counterpart's behaviour based on the common knowledge of rationality. The forecast

TABLE 4: SPNE in the conflict game with perfect information.

	$(s_1^*, s_2^*)$	$(u_1^*, u_2^*)$	Pareto optimal
C1	$(d, dd)$	$(1, 1)$	Yes
C2	$(d, dh)$	$(1, 1)$	Yes
C3	$(h, hd)$	$(y, x)$	Yes
C4	$(h, hh)$	$(0.25, 0.25)$	No

possibility leads to a first mover advantage, as we can see in Proposition 2.

**Proposition 2 (first mover advantage).** If  $(s_1^*, s_2^*)$  is a SPNE in the conflict game with perfect information, then  $u_1(s_1^*, s_2^*) \neq x$  and  $u_2(s_1^*, s_2^*) \neq y$ .

The intuition behind Proposition 2 is that there is an advantage related to the opportunity of playing first, which is the idea behind the ordinary commitment. In consequence, the equilibrium that is reached is that in favour of Player 1, because he always obtains at least as much as his opponent. This is true except for the C4 game, because the level of conflict is so high that regardless of what player 1 chooses he cannot improve his position. The SPNE for each game is presented in Table 4.

We can see that the possibility to play a response rule is not enough to increase player 2's bargaining power. For this reason, we now consider the case where player 2 has the possibility to announce the reaction rule she is going to play, before player 1 makes his decision.

**3.2. Threats and Promises as Binding Messages.** Following Schelling [14], the sender's bargaining power increases if she is able to send a message about the action she is going to play, since with premeditation other alternatives have been rejected. For the receiver it must be clear that this is the unique relevant option. This strategic move can be implemented if it is possible to send binding messages about second mover's future actions. With this kind of communication we are going to show that there always exists a message that allows player 2 to reach an outcome at least as good as the outcome in the SPNE. By notation,  $m_2 \in S_2$  is a conditioned message, where  $m_2 = m_{2d}m_{2h}$ . From now on, player 2 represents the sender and player 1 the receiver.

**Definition 3 (commitment message).**  $m_2^* \in S_2$  is a commitment message if and only if  $u_2(s_{1m}^*, m_2^*) \geq u_2(s_1^*, s_2^*)$ , where  $u_1(s_{1m}^*, m_2^*) \geq u_1(s_1, m_2^*)$  for every  $s_1 \in S_1$ . It means  $s_{1m}^*$  is player 1 best response given  $m_2^*$ .

The idea behind *commitment messages* is that player 2 wants to achieve an outcome at least as good as the one without communication, given the receiver's best response. This condition only looks for compatibility of incentives, since the receiver also makes his decisions in a rational way. Following closely the formulations discussed in Schelling [14], Klein and O'Flaherty [8], and Hirshliefer [7], we classify the commitment messages in three mutually exclusive categories: *warnings*, *threats*, and *promises*.



TABLE 5: Commitment messages.

	Warning	$(u_1^*, u_2^*)$	Threat	$(u_1, u_2)$	Promise	$(u_1, u_2)$
C1	$(d, dd)$	$(1, 1)$	$(d, dh)$	$(1, 1)$		
C2	$(d, dh)$	$(1, 1)$	$(d, dd)$	$(1, 1)$		
C3	$(h, hd)$	$(y, x)$	$(d, hh)$	$(x, y)$	$(d, dh)$	$(1, 1)$
C4	$(h, hh)$	$(0.25, 0.25)$			$(d, dh)$	$(1, 1)$

**Definition 4** (warnings, threats, and promises). (1) The commitment message  $m_2^* \in S_2$  is a warning if and only if  $m_2^* = s_2^*$ .

(2) The commitment message  $m_2^* \in S_2$  is a threat if and only if  $u_2(d, m_2^*) = u_2(d, s_2^*)$  and  $u_2(h, m_2^*) < u_2(h, s_2^*)$ .

(3) The commitment message  $m_2^* \in S_2$  is a promise if and only if  $u_2(d, m_2^*) < u_2(d, s_2^*)$ .

The purpose of a *warning* commitment is to confirm that the sender will play her best response after every possible action of the receiver. Schelling does not consider *warnings* as strategic moves, but we prefer to use it in this way because the important characteristic of *warnings* is their full credibility condition. If agents want to avoid miscoordination related to the common knowledge of rationality, they could communicate it and believe it as well. On the contrary, credibility is an inherent problem in *threats* and *promises*. The second and third points in Definition 4 show that at least one action in the message is not the best response after observing the receiver's choice. In *threats*, the sender does not have any incentive to implement the punishment when the receiver plays hawk. In *promises*, the sender does not have any incentive to fulfill the agreement when the receiver plays dove.

The strategic goal in the conflict game is to deter the opponent of choosing hawk, because by assumption  $u_i(s_i, d) > u_i(s_i, h)$ . This is exactly the purpose of these binding messages, as shown in Proposition 5.

**Proposition 5** (second mover advantage). *If  $\hat{m}_2$  is a threat or a promise in the conflict game with perfect information, then  $s_{1\hat{m}}^* = d$ .*

The intuition behind Proposition 5 is that, in Schelling's terms, if a player has the possibility to announce her intentions, she will use threats or promises to gain an advantage over the first mover. That is, player 2 uses these messages because, if believed by player 1, she can make him cooperate.

Proposition 6 specifies for which cases player 2 influences player 1's choices by means of threats and promises. That is, in which cases, when player 1 has no incentives to cooperate, messages can prompt a change in his behaviour.

**Proposition 6** (message effectivity). *There exists a commitment message  $m_2^*$  such that  $u_2(s_{1m}^*, m_2^*) > u_2(s_1^*, s_2^*)$  if and only if  $y > 1$ .*

Therefore, threats and promises provide a material advantage upon the adversary only in cases with high conflict (e.g., C3 and C4). Thus, the condition  $y > 1$  is not satisfied in C1 and C2 cases, where the level of conflict is low.

The implication is that mutual cooperation is achieved in equilibrium and this outcome is the highest for both players. The use of messages under these incentives only needs to confirm the sender's rational choice. If player 2 plays  $m^* = s_2^*$ , receiver can anticipate this rational behaviour, which is completely credible. This is exactly the essence of the subgame perfect Nash equilibrium proposed by Selten [3].

An essential element of commitments is to determine under what conditions the receiver must take into account the content of a message, given that the communication purpose is to change the rival's expectations. The characteristic of a *warning* is to choose the weakly dominant strategy, but for *threats* or *promises* at least one action is not a best response. Proposition 6 shows that in the C3 and C4 cases the sender's outcome is strictly higher if she can announce that she does not follow the subgame perfect strategy. We summarise these findings in Table 5.

Up to this point we have considered the first two elements of commitment theory. We started by illustrating that the messages sent announce the intention the sender has to execute a plan of action (i.e., the choice of a response rule). Subsequently, we described for which cases messages are effective (i.e., self-serving announcements). Now we inquire about the credibility of these strategic moves, because if the sender is announcing that she is going to play in an opposite way to the game incentives, this message does not change the receiver's beliefs. The message is not enough to increase the bargaining power. It is necessary that the specified action is actually the one that will be played, or at least that the sender believes it. The objective in the next section is to stress the credibility condition. It is clear that binding messages imply a degree of commitment at a 100% level, but this condition is very restrictive, and it is not a useful way to analyse a real bargaining situation. We are going to prove that for a successful strategic move the degree of commitment must be high enough, although it is not necessary to tell the truth with a probability equal to 1.

#### 4. The Conflict Game with Nonbinding Messages

The credibility problem is related to how likely it is that the message sent coincides with the actions chosen. The sender announces her way of playing, but it could be a bluff. In other words, the receiver can believe in the message if it is highly probable that the sender is telling the truth. In order to model this problem the game now proceeds as follows. In the first stage *Nature* assigns a type to player 2 following a probability distribution. The sender's type is her action plan; her way of playing in case of observing each of the possible

receiver's action. In the second stage player 2 observes her *type* and sends a signal to player 1. The signal is the disclosure of her plan, and it can be seen as a noisy message, because it is nonbinding. In the last stage, player 1, after receiving the signal information, chooses an action. This choice determines the players' payoffs together with the actual type of player 2.

Following the intuition behind credible message profile in Rabin [12], a commitment announcement can be considered credible if it fulfills the following conditions. (i) When the receiver believes the literal meanings of the statements, the types sending the messages obtain their best possible payoff; hence those types will send these messages. (ii) The statements are truthful *enough*. The *enough* comes from the fact that some types might lie to player 1 by pooling with a commitment message and the receiver knows it. However, the probability of facing a lie is small enough that it does not affect player 1's optimal response.

The objective of this section is to formalise these ideas using our benchmark *conflict game*. The strategic credibility problem is intrinsically dynamic, and it makes sense if we consider threats and promises as nonbinding messages. Bearing these considerations in mind, from now on the messages are used to announce the sender's intentions, but they are *cheap talk*. Clearly, negotiators talk, and in most of the cases it is free, but we show that this fact does not imply that cheap talk is meaningless or irrelevant.

**4.1. The Signalling Conflict Game.** Consider a setup in which player 2 moves first; player 1 observes a message from player 2 but not her type. They choose as follows: In the first stage *Nature* assigns a type  $\theta_2$  to player 2 as a function that assigns an action  $s_2 \in \{d, h\}$  to each action  $s_1 \in \{d, h\}$ . Player 2's type set is  $\Theta_2 = S_2 = \{dd, dh, hd, hh\}$ , where  $\theta_2 = s_{2d}s_{2h}$ . *Nature* chooses the sender's type following a probability distribution, where  $p(\theta_2) > 0$  is the probability to choose the type  $\theta_2$ , and  $\sum_{\theta_2 \in \Theta_2} p(\theta_2) = 1$ . In the second stage, player 2 observes her own type and chooses a message  $m_2 \in \Theta_2$ . At the final stage, player 1 observes this message and chooses an action from his set of strategies  $S_1 = \{d, h\}$ . The most important characteristic of this *conflict game with nonbinding messages* is that communication cannot change the final outcome. Though strategies are more complex in this case, the  $2 \times 2$  payoff matrix in the *conflict game* is always the way to determine the final payoffs.

In order to characterise the utility function we need some notation. A message profile  $m_2 = (m_{dd}, m_{dh}, m_{hd}, m_{hh})$  is a function that assigns a message  $m_2 \in \Theta_2$  to each type  $\theta_2 \in \Theta_2$ . The first component  $m_{dd} \in S_2$  is the message chosen in case of observing the type  $\theta_2 = dd$ ; the second component  $m_{dh} \in S_2$  is the message chosen in case of observing the type  $\theta_2 = dh$ , and so on. By notation,  $m_{\theta_2} = s_{2d}s_{2h}$  is a specific message sent by a player with type  $\theta_2$ , and  $m_2 = (m_{\theta_2}, m_{-\theta_2})$  is a generic message profile with emphasis on the message sent by the player with type  $\theta_2$ .

There is imperfect information because the receiver can observe the message, but the sender's type is not observable. Thus, the receiver has four different information sets, depending on the message he faces. A receiver's strategy  $s_{1m} = (s_{1dd}, s_{1dh}, s_{1hd}, s_{1hh})$  is a function that assigns

an action  $s_1 \in S_1$  to each message  $m_2 \in \Theta_2$ , where  $s_{1dd}$  is the action chosen after observing the message  $m_2 = dd$ , and so on. In addition,  $s_{1m} = (s_{1m}, s_{1(-m)})$  is a receiver's generic strategy with emphasis on the message he faced. In this case, the subindex  $m$  is the way to highlight that the receiver's strategies are a profile of single actions. Therefore, in the *conflict game with nonbinding messages* the utility function is  $u_i(s_{1m_{\theta_2}}, s_{1(-m_{\theta_2})}, m_{\theta_2}, m_{-\theta_2}) = u_i(s_1, s_2)$  for  $s_{1m_{\theta_2}} = s_1$  and  $\theta_2 = s_2$ .

In this specification, messages are payoff irrelevant and what matters is the sender's type. For this reason, it is necessary to define the receiver's beliefs about who is the sender when he observes a specific message. The receiver's belief  $\alpha_{\theta_2|m_2} \geq 0$  is the conditional probability of obtaining the message from a sender of type  $\theta_2$ , given that he observed the message  $m_2$ . Naturally,  $\sum_{\theta_2 \in \Theta_2} \alpha_{\theta_2|m_2} = 1$ .

All the elements of the *conflict game with nonbinding messages* are summarised in Figure 2. The most salient characteristics are the four information sets in which the receiver must choose and that messages are independent of payoffs. For instance, the upper left path (blue) describes each possible decision for the sender of type  $dd$ . In the first place, *Nature* chooses the sender's type; in this case  $\theta_2 = dd$ . In the next node,  $dd$  must choose a message from the 4 possible reaction rules. We say that  $dd$  is telling the truth if she chooses  $m_{dd} = dd$ , leading to the information set at the top. We intentionally plot the game in a star shape in order to highlight the receiver's information sets. At the end, the receiver chooses between  $d$  and  $h$ , and *cheap talk* implies that there are 4 feasible payoffs.

The signalling conflict game has a great multiplicity of Nash equilibria. For this particular setting, a characterisation of this set is not our aim. Our interest lies on the characterisation of the communication equilibrium. For this reason the appropriate concept in this case is the perfect Bayesian equilibrium.

**Definition 7 (PBE).** A perfect Bayesian equilibrium is a sender's message profile  $m_2^* = (m_{dd}^*, m_{dh}^*, m_{hd}^*, m_{hh}^*)$ , a receiver's strategy profile  $s_{1m}^* = (s_{1dd}^*, s_{1dh}^*, s_{1hd}^*, s_{1hh}^*)$ , and a beliefs profile  $\alpha_{\theta_2|m_2}^*$  after observing each message  $m_2$ , if the following conditions are satisfied:

- (1)  $m_2^*$  is the  $\text{argmax}_{m_2 \in \Theta_2} u_{\theta_2}(s_{1m}^*, m_{\theta_2}, m_{-\theta_2})$ ,
- (2)  $s_{1m}^*$  is the  $\text{argmax}_{s_1 \in S_1} \sum_{\theta_2 \in \Theta_2} \alpha_{\theta_2|m_2} \cdot u_1(s_{1m_{\theta_2}}, s_{1(-m_{\theta_2})}, m_2^*)$ ,
- (3)  $\alpha_{\theta_2|m_2}^*$  must be calculated following Bayes' rule based on the message profile  $m_2^*$ . For all  $\theta_2$  who play the message  $m_2^*$ , the beliefs must be calculated as  $\alpha_{\theta_2|m_2}^* = p_{\theta_2} / \sum p_{m_2^*}$ .

The conditions in this definition are incentive compatibility for each player and Bayesian updating. The first condition requires message  $m_{\theta_2}^*$  to be optimal for type  $\theta_2$ . The second condition requires strategy  $s_{1m}^*$  to be optimal given the beliefs profile  $\alpha_{\theta_2|m_2}^*$ . For the last condition, Bayesian updating, the receiver's

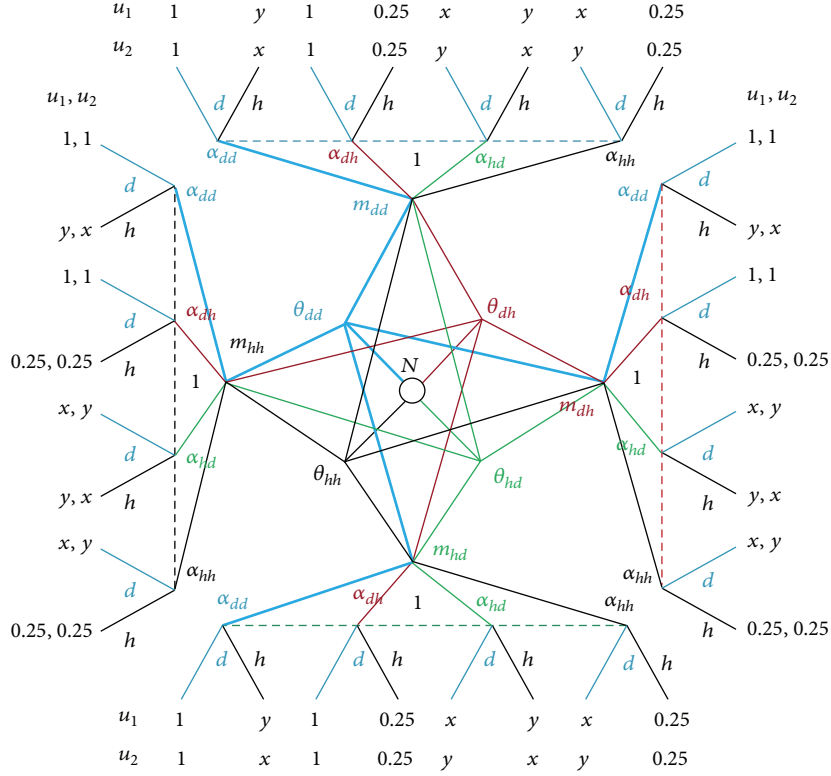


FIGURE 2: Conflict game with nonbinding messages.

beliefs must be derived via Bayes' rule for each observed message, given the equilibrium message profile  $m_2^*$ .

**4.2. The Commitment Equilibrium Properties.** There are, in general, several different equilibria in the *conflict game with nonbinding messages*. The objective of this section is to show that a particular equilibrium that satisfies the following properties leads to a coordination outcome, given it is both salient and in favour of the sender. In what follows we will present Axioms 1 and 2 which will be used to explain which is the particular equilibrium that can be used as a theoretical prediction in experimental games with different levels of conflict.

**Axiom 1** (truth-telling beliefs). If the receiver faces a message  $\bar{m}_2^* = \hat{\theta}_2$ , then  $\alpha_{\hat{\theta}_2|\bar{m}_2^*} > 0$ . If the message  $\bar{m}_2 = \bar{\theta}_2$  is not part of the messages profile  $m_2^*$ , then  $\alpha_{\bar{\theta}_2|\bar{m}_2} = 1$ .

Following Farrell and Rabin [11] we assume that people in real life do not seem to lie as much or question each other's statements as much, as the game theoretic predictions state. Axiom 1 captures the intuition that for people it is natural to take seriously the literal meaning of a message. This does not mean that they believe everything they hear. It rather states that they use the meaning as a starting point and then assess credibility, which involves questioning in the form of "why would she want me to think that? Does she have incentives to actually carry out what she says?"

More precisely, *truth-telling beliefs* emphasise that in equilibrium when the receiver faces a particular message, its literal meaning is that the sender has the intention of playing in this way. Thus, the probability of facing truth-telling messages must be greater than zero. In the same way, when the sender does not choose a particular message, she is signalling that there are no incentives to make the receiver believe this, given that the receiver's best response is  $h$ . Therefore, we can assume that the receiver must fully believe in the message, because both players understand that the purpose of the strategic move is to induce the receiver to play  $d$ . If the sender is signalling the opposite, she is showing her true type by mistake; then the receiver believes her with probability 1 (see the column "belief of truth-telling" in Table 6).

**Axiom 2** (senders' bargaining power). If  $m_{\theta_2}^*$  is part of the messages profile  $m_2^*$ , then  $s_{1m_{\theta_2}}^* = d$ .

Axiom 2 captures the use of communication as a means to influence the receiver to play dove. That is, there is an equilibrium where the only messages sent are those that induce the receiver to cooperate. In order to characterise a communication equilibrium such as the one described above, we first focus on the completely separating message profile, when the sender is telling the truth. Naturally,  $m_{\theta_2}$  is a truth-telling message if and only if  $m_{\hat{\theta}_2} = \hat{\theta}_2$  (see column "message by type" in Table 6), and given the message the receiver's best

TABLE 6: Perfect Bayesian equilibria that satisfy Axioms 1 and 2.

	Message by type ( $m_{dd}^*, m_{dh}^*, m_{hd}^*, m_{hh}^*$ )	Player 1's best resp. ( $s_{1dd}^*, s_{1dh}^*, s_{1hd}^*, s_{1hh}^*$ )	Belief of truth-telling ( $\alpha_{dd dd}^*, \alpha_{dh dh}^*, \alpha_{hd hd}^*, \alpha_{hh hh}^*$ )
C1	( $dd, dh, dd, hh$ )	( $d, d, h, d$ )	$\left( \frac{p_{dd}}{(p_{dd} + p_{hd})}, 1, 1, 1 \right)$
C2	( $dd, dh, dd, dh$ )	( $d, d, h, h$ )	$\left( \frac{p_{dd}}{(p_{dd} + p_{hd})}, \frac{p_{dh}}{(p_{dh} + p_{hh})}, 1, 1 \right)$
C3	( $dh, dh, hh, hh$ )	( $h, d, h, d$ )	$\left( 1, \frac{p_{dh}}{(p_{dd} + p_{dh})}, 1, \frac{p_{hh}}{(p_{hd} + p_{hh})} \right)$
C4	( $dh, dh, dh, dh$ )	( $h, d, h, h$ )	$\left( 1, \frac{p_{dh}}{(p_{dd} + p_{dh} + p_{hd} + p_{hh})}, 1, 1 \right)$

response will be to cooperate (see column “player 1’s best response” in Table 6).

With this in mind, it is possible to stress that a contribution of our behavioural model is to develop experimental designs that aim to unravel the strategic use of communication to influence (i.e., manipulate) others’ behaviour. That is, the Nash equilibrium implies that players must take the other players’ strategies as given and then they look for their best response. However, commitment theory, in Schelling’s sense, implies an additional step, where players recognise that opponents are fully rational. Based on this fact, they evaluate different techniques for turning the other’s behaviour into their favour. In our case, the sender asks herself, “This is the outcome I would like from this game; is there anything I can do to bring it about?”

**Proposition 8** (there is always a liar). *The completely truth-telling messages profile  $m_2 = (dd, dh, hd, hh)$  cannot be part of any PBE of the conflict game with nonbinding messages.*

Proposition 8 shows that the completely truth telling message profile is not an equilibrium in the conflict game. The problem lies in the sender type  $hd$ , because revealing her actual type is not incentive compatible and there always exists at least one successful message to induce the counterpart to play dove. For this reason, we can ask whether there exists some message that induces the sender to reveal her actual type but at the same time leads to a successful strategic move. Definition 9 is the bridge between *nonbinding messages* and *commitment messages* presented in the previous section.

**Definition 9** (self-committing message). Let  $\widehat{m}_{\widehat{\theta}_2}^*$  be a truth-telling message and  $\alpha_{\widehat{\theta}_2|\widehat{m}_{\widehat{\theta}_2}^*}(\widehat{\theta}_2) = 1$ .  $\widehat{m}_{\widehat{\theta}_2}^*$  is a self-committing message if and only if  $u_{\widehat{\theta}_2}(s_{1m}^*, \widehat{m}_{\widehat{\theta}_2}^*, \widehat{m}_{-\widehat{\theta}_2}^*) \geq u_{\widehat{\theta}_2}(s_{1m}^*, m_{\widehat{\theta}_2}, \widehat{m}_{-\widehat{\theta}_2}^*)$ , for every  $m_{\widehat{\theta}_2} \in \Theta_2$ .

We introduce the *self-committing* message property because we want to stress that a strategic move is a two-stage process. Not only is communication useful in revealing information, but also it can be used to manipulate others’ behaviour. The sender of a message must consider how the receiver would react if he believes it and if that behaviour works in her favour she will not have incentives to lie. A message is *self-committing* and, if believed, it creates

incentives for the sender to fulfill it [12]. The idea behind a *threat* or a *promise* is to implement some risk for the opponent in order to influence him, but this implies a risk for the sender too. This fact has led to associating strategic moves with slightly rational behaviours, when actually, in order to be executed, a very detailed evaluation of the consequences is needed. Proposition 10 and its corollary explain the relation between the conditioned messages and the incentives to tell the truth.

**Proposition 10** (incentives to commit). *Let  $\widehat{m}_2 = \widehat{m}_{\widehat{\theta}_2}^*$  be a commitment message in the conflict game with perfect information. If  $s_{1m}^*(\widehat{\theta}_2) = d$ , then  $\widehat{m}_{\widehat{\theta}_2}^*$  is a self-committing message.*

**Corollary to Proposition 10.** *If  $\widehat{m}_2$  is a threat or a promise in the conflict game with perfect information, then  $\widehat{m}_{\widehat{\theta}_2}^* = \widehat{m}_2$  is a self-committing message.*

The intuition behind Proposition 10 and its corollary is that if a message induces the other to cooperate, then the sender has incentives to tell the truth. Moreover, as illustrated in Proposition 5, threats and promises always induce the counterpart to cooperate; therefore, they endogenously give the sender incentives to comply with what is announced.

As we can see in the *conflict game with perfect information* (for an illustration see Table 5), in the C1 and C2 cases the *warning* is the way to reach the best outcome. If we consider the possibility to send nonbinding messages when the sender’s type is equal to a *warning* strategy, then revealing her type is self-committing. The problem in the C3 and C4 cases is more complex given the *warning* message is not self-committing and the way to improve the bargaining power is using a *threat* or a *promise*. This fact leads to a trade-off between choosing a weakly dominant strategy that fails to induce the opponent to play dove and a strategy that improves her bargaining power but implies a higher risk for both of them.

The required elements for a perfect Bayesian equilibrium at each game are shown in Tables 6 and 7. It is important to bear in mind that the beliefs that appear in Table 7 are necessary conditions for implementing the PBE presented in Table 6, given that they satisfy *truth-telling beliefs* and *sender’s bargaining power*.



TABLE 7: Beliefs that support the perfect Bayesian equilibrium.

	Warning	Threat	Promise
C1	$\alpha_{dd dd}^* \geq \frac{\alpha_{hd dd}^* (y - x)}{(1 - y)}$	Truth	
C2	$\alpha_{dh dh}^* \geq \frac{\alpha_{hh dh}^* (0.25 - x)}{0.75}$	$\alpha_{dd dd}^* \geq \frac{\alpha_{hd dd}^* (y - x)}{(1 - y)}$	
C3	Lie	$\alpha_{hh hh}^* \geq \frac{\alpha_{hd hh}^* (y - x)}{(x - 0.25)}$	$\alpha_{dh dh}^* \geq \frac{\alpha_{dd dh}^* (y - 1)}{0.75}$
C4	Lie		$\alpha_{dh dh}^* \geq \frac{\alpha_{dd dh}^* (y - 1) + \alpha_{hd dh}^* (y - x) + \alpha_{hh dh}^* (0.25 - x)}{0.75}$

The problem of which message must be chosen is as simple as follows in the next algorithm: first, the sender tells the truth. If the truth-telling message leads the receiver to play dove, then she does not have any incentive to lie. In the other case, she must find another message to induce the receiver to play dove. If no message leads the receiver to play dove, messages will lack any purpose, and she will be indifferent between them.

Table 6 shows the messages; the receivers' strategies and their belief profiles in a particular equilibrium we argue is the most salient. As we showed above, in the conflict game (see Table 5) the sender is always in favour of those messages where the receiver's best response is dove. In the C1 case there are three different messages, in the C2 and C3 cases there are two messages, and the worst situation is the C4 case, where every type of player sends the same message. This fact leads to a first result: if the conflict is high, there are very strong incentives to lie and communication leads to a pooling equilibrium.

In addition, notice that Table 5 specifies which messages will be used as commitment messages in the conflict game with binding communication illustrated in Figure 1. That is, if credibility is exogenous the theoretical prediction would be that such messages are sent. This means that messages are not randomly sent, but there is a clear intention behind them, to induce the receiver to choose the action most favourable for the sender. Now, Table 7 presents the minimum probability threshold that can make the strategic move successful. That is, if credibility is sufficiently high the message works and achieves its purpose, in the conflict game with nonbinding communication illustrated in Figure 2.

In Section 3 we assumed that the sender could communicate a completely credible message in order to influence her counterpart. The question is, how robust is this equilibrium if we reduce the level of commitment? Proposition 11 summarises the condition for the receiver to choose dove as the optimal strategy. It is the way for calculating the beliefs that are shown in Table 7.

**Proposition 11** (incentives to cooperate).  $s_{1m(\theta_2)}^* = d$  if and only if  $(1 - y)\alpha_{dd|m(\theta_2)} + (0.75)\alpha_{dh|m(\theta_2)} + (x - y)\alpha_{hd|m(\theta_2)} + (x - 0.25)\alpha_{hh|m(\theta_2)} \geq 0$ .

Based on Proposition 11, the second result is that cheap talk always has meaning in equilibrium. We consider that this equilibrium selection is relevant because the sender

focuses on the communication in the literal meanings of the statements but understands that some level of credibility is necessary to improve her bargaining power. Table 7 summarises the *true enough* property of the statements. Here, the receiver updates his beliefs in a rational way and he chooses to play dove if and only if it is his expected best response. We can interpret the beliefs in Table 7 as a threshold, because if this condition is satisfied, the sender is successful in her intention of manipulating the receiver's behaviour. Thus, some level of credibility is necessary, but not at a 100% level.

It is clear that if the conflict is high, the commitment threshold is also higher. In C1 and C2 cases the sender must commit herself to implement the *warning* strategy, which is a weakly dominant strategy. In the C3 case the strategic movement implies a *threat* or a *promise*, formulating an aggressive statement in order to deter the receiver from behaving aggressively. The worst situation is the C4 case, where there is only one way to avoid the disagreement point, to implement a *promise*. The promise in this game is a commitment that avoids the possibility of exploiting the opponent, because fear can destroy the agreement of mutual cooperation.

In the scope of this paper, threats are not only punishments and promises are not only rewards. There is a credibility problem because these strategic moves imply a lack of freedom in order to avoid the rational self-serving behaviour in a simple one step of thinking. The paradox is that this decision is rational if the sender understands that her move can influence other players' choices, because communication is the way to increase her bargaining power. This implies a second level of thinking, such as a forward induction reasoning.

## 5. Conclusions

In this paper we propose a behavioural model following Schelling's tactical approach for the analysis of bargaining. In his Essay on Bargaining 1956, Schelling analyses situations where subjects watch and interpret each other's behaviour, each one better acting taking into account the expectations that he creates. This analysis shows that an opponent with rational beliefs expects the other to try to disorient him and he will ignore the movements he perceives as stagings especially played to win the game.

The model presented here captures different levels of conflict by means of a simple parameterisation. In

a bilateral bargaining environment it analyses the strategic use of binding and nonbinding communication. Our findings show that when messages are binding, there is a first mover advantage. This situation can be changed in favour of the second mover, if the latter sends threats or promises in a preplay move. On the other hand, when players have the possibility to send nonbinding messages, their incentives to lie depend on the level of conflict. When conflict is low, the sender has strong incentives to tell the truth and cheap talk will almost fully transmit private information. When conflict is high, the sender has strong incentives to bluff and lie. Therefore, in order to persuade the receiver to cooperate with her nonbinding messages, the sender is required to provide a minimum level of credibility (not necessarily a 100%).

In summary, the equilibrium that satisfies *truth-telling beliefs* and *sender's bargaining power* allows us to show that the less conflict the game has, the more informative the equilibrium signal is, and the less stronger the commitment needed to implement it is. Our equilibrium selection is based on the assumption that in reality people do not seem to lie as much, or question each other's statements as much, as rational choice theory predicts. For this reason, the *conflict game with nonbinding messages* is a good environment to test different game theoretical hypotheses, because it is simple enough to be implemented in the lab.

With this in mind, the strategic use of communication in a conflict game, as illustrated in our model, is the right way to build a bridge between two research programs: the theory on bargaining and that on social dilemmas. As Bolton [15] suggested, bargaining and dilemma games have been developed in experimental research as fairly separate literatures. For bargaining, the debate has been centred on the role of fairness and the nature of strategic reasoning. For dilemma games, the debate has involved the relative weights that should be given to strategic reputation building, altruism, and reciprocity. The benefit of the structure and payoff scheme we propose is to study all these elements at the same time. Our model provides a simple framework to gather and interpret empirical information. In this way, experiments could indicate which parts of the theory are most useful to predict subjects' behaviour and at the same time we can identify behavioural parameters that the theory does not reliably determine.

Moreover, the game presented here can be a very useful tool to design economic experiments that can lead to new evidence about bilateral bargaining and, furthermore, about human behaviour in a wider sense. On the one hand, it can contribute to a better understanding of altruism, selfishness, and positive and negative reciprocity. A model that only captures one of these elements will necessarily portray an incomplete image. On the other hand, bargaining and communication are fundamental elements to understand the power that one of the parts can have.

In further research, we are interested in exploring the emotional effects of cheating or being cheated on, particularly by considering the dilemma that takes place when these emotional effects are compared to the possibility of obtaining material advantages. To do so, it is possible to even consider a simpler version of our model using a coarser type space

(e.g., only hawk and dove). This could illustrate the existing relationship between the level of conflict and the incentives to lie. As the model predicts, the higher the level of conflict the more incentives players have to not cooperate, but they are better off if the counterpart does cooperate. Therefore, players with type *hawk* would be more inclined to lie and disguise themselves as cooperators. By measuring the emotional component of lying and being lied to, we will be able to show that people do not only value the material outcomes of bargaining but that the means used to achieve those ends are also important to them.

## Appendix

*Proof of Proposition 2.* Suppose that  $u_1(s_1^*, s_2^*) = x$  and  $u_2(s_1^*, s_2^*) = y$ ; then  $y \geq 1$ . If  $s_2^* = hd$ , then  $u_1(d, hd) \geq u_1(h, hd)$  and  $x \geq y$ , but by assumption  $y > x$ . If  $s_2^* = hh$ , then  $u_1(d, hh) \geq u_1(h, hh)$  and  $x \geq 0.25$ , and at the same time  $u_2(h, hh) \geq u_2(h, hd)$ . The only compatible case is  $0.25 = x$ , but by assumption  $0.25 \neq x$ . Therefore,  $u_1(s_1^*, s_2^*) \neq x$  and  $u_2(s_1^*, s_2^*) \neq y$ .  $\square$

*Proof of Proposition 5.* Let  $\hat{m}_2$  be a threat or a promise. Following Definitions 3 and 4,  $u_2(s_{1\hat{m}}^*, \hat{m}_2^*) \geq u_2(s_1^*, s_2^*)$ . Suppose that  $s_{1\hat{m}}^* = h$ ; then there are two possibilities,  $\hat{m}_2^* = s_2^*$  or  $u_2(h, \hat{m}_2^*) \geq u_2(s_1^*, s_2^*)$ . If  $\hat{m}_2^* = s_2^*$ , then by definition  $\hat{m}_2^*$  is neither a threat nor a promise. If  $u_2(h, \hat{m}_2^*) \geq u_2(s_1^*, s_2^*)$ , then  $s_1^* = d$  or  $s_1^* = h$ . If  $s_1^* = d$ , by assumption  $u_2(h, \hat{m}_2^*) < u_2(d, s_2^*)$ . If  $s_1^* = h$  and  $\hat{m}_2^*$  is a threat, then  $u_2(h, \hat{m}_2^*) < u_2(s_1^*, s_2^*)$ . If  $s_1^* = h$  and  $\hat{m}_2^*$  is a promise, it must fulfill  $u_2(h, \hat{m}_2^*) \geq u_2(h, s_2^*)$  and  $u_2(d, \hat{m}_2^*) < u_2(d, s_2^*)$ . The C1 and C2 games are not under consideration because  $s_1^* = d$  and for C3 y C4 cases there are no messages for which these conditions are true at the same time. Therefore,  $s_{1\hat{m}}^* = d$ .  $\square$

*Proof of Proposition 6.* Let us consider the message  $m_2 = dh$ . By Proposition 2 we know that  $u_2(s_1^*, s_2^*) \neq y$ , and by assumption  $u_1(d, dh) > u_1(h, dh)$ , then  $m_2 = dh$  is a commitment message, because  $u_2(d, dh) = 1 \geq u_2(s_1^*, s_2^*)$ . If  $u_2(d, dh) > u_2(s_1^*, s_2^*)$ , then  $1 > u_2(s_1^*, s_2^*)$ , to satisfy this condition and using Proposition 2 again; we conclude that  $(s_1^*, s_2^*) = (h, hs_{2h}^*)$ . As  $s_2^* = hs_{2h}^*$  and it is part of the SPNE, then  $u_2(d, hs_{2h}^*) > u_2(d, ds_{2h}^*)$ , and therefore  $y > 1$ .

The proof in the other direction is as follows. Let  $y > 1$ ; then  $s_2^* = hs_{2h}^*$ . Using Proposition 2 we know that  $u_1(s_1^*, s_2^*) \neq x$ ; therefore  $s_1^* = h$ . Now  $u_2(s_1^*, s_2^*) < 1$ . As we show in the first part,  $m_2 = dh$  is a commitment message such that  $u_2(s_{1m}^*, dh) = 1$ . Therefore, there exists a commitment message such that  $u_2(s_{1m}^*, m_2^*) > u_2(s_1^*, s_2^*)$ .  $\square$

*Proof of Proposition 8.* Consider the senders' types  $\theta_{dh} = dh$  and  $\theta_{hd} = hd$ . If  $m_2^*$  is a completely truth-telling message, then  $\alpha_{dh|dh}^* = 1$  and  $\alpha_{hd|hd}^* = 1$ . By assumptions  $u_1(d, s_{1(-dh)}, dh, m_{-dh}) = 1$  and  $u_1(h, s_{1(-dh)}, dh, m_{-dh}) = 0.25$ , then  $s_{1dh}^* = d$ . In the same way,  $u_1(d, s_{1(-hd)}, hd, m_{-hd}) = x$  and  $u_1(h, s_{1(-hd)}, hd, m_{-hd}) = y$ ; then  $s_{1hd}^* = h$ . Therefore, the utility for the sender is  $u_{dh}(d, s_{1(-dh)}, dh, m_{-dh}) = 1$  and  $u_{hd}(d, s_{1(-hd)}, hd, m_{-hd}) = x$ . These conditions imply that the sender type *hd* has incentives to deviate and  $m_2 = (dd, dh, hd, hh)$  cannot be part of any PBE.  $\square$

*Proof of Proposition 10.* Let  $\widehat{m}_2 = \widehat{m}_{\widehat{\theta}_2}^*$  be a commitment message in the conflict game with perfect information and  $s_{1m^*(\widehat{\theta}_2)}^* = d$ . If  $\widehat{m}_{\widehat{\theta}_2}^* = \widehat{m}_2$  is not a self-committing message, then another message  $\overline{m}_{\widehat{\theta}_2}$  must exist such that  $u_{\widehat{\theta}_2}(d, s_{1(-m^*(\widehat{\theta}_2))}, \widehat{m}_{\widehat{\theta}_2}^*, m_{-\widehat{\theta}_2}^*) < u_{\widehat{\theta}_2}(s_{1m}, \overline{m}_{\widehat{\theta}_2}, m_{-\widehat{\theta}_2}^*)$ . Given the payoff assumptions,  $u_{\widehat{\theta}_2}(d, s_{1(-m^*(\widehat{\theta}_2))}, \widehat{m}_{\widehat{\theta}_2}^*, m_{-\widehat{\theta}_2}^*) \geq u_{\widehat{\theta}_2}(s_{1m}, m_{\widehat{\theta}_2}^*, m_{-\widehat{\theta}_2}^*)$  for every  $m_{\widehat{\theta}_2}^* \in \Theta_2$ . Therefore,  $\widehat{m}_{\widehat{\theta}_2}^* = \widehat{m}_2$  is a self-committing message.  $\square$

*Proof of Corollary to Proposition 10.* The proof to the corollary follows from Propositions 5 and 10, and thus it is omitted.  $\square$

*Proof of Proposition 11.* The expected utility for each receiver's strategy is as follows:

$$u_1(d, m_{\theta_2}) = 1\alpha_{dd|m(\theta_2)} + 1\alpha_{dh|m(\theta_2)} + x\alpha_{hd|m(\theta_2)} + x\alpha_{hh|m(\theta_2)},$$

$$u_1(h, m_{\theta_2}) = y\alpha_{dd|m(\theta_2)} + 0.25\alpha_{dh|m(\theta_2)} + y\alpha_{hd|m(\theta_2)} + 0.25\alpha_{hh|m(\theta_2)},$$

$$\text{therefore, } u_1(d, m_{\theta_2}) \geq u_1(h, m_{\theta_2}) \text{ if and only if } (1 - y)\alpha_{dd|m(\theta_2)} + (0.75)\alpha_{dh|m(\theta_2)} + (x - y)\alpha_{hd|m(\theta_2)} + (x - 0.25)\alpha_{hh|m(\theta_2)} \geq 0.$$

$\square$

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This paper was elaborated during the authors' stay at the University of Granada, Spain. The authors are grateful for the help and valuable comments of Juan Lacomba, Francisco Lagos, Fernanda Rivas, Aurora García, Erik Kimbrough, Sharlane Scheepers, and the seminar participants at the VI International Meeting of Experimental and Behavioural Economics (IMEBE) and the Economic Science Association World Meeting 2010 (ESA). Financial support from the Spanish Ministry of Education and Science (Grant code SEJ2009-11117/ECON), the Proyecto de Excelencia (Junta de Andalucía, P07-SEJ-3261), and the Project VIE-1375 from the Universidad Industrial de Santander (UIS) in Bucaramanga, Colombia, is also gratefully acknowledged.

## References

- [1] A. Rubinstein, "Perfect equilibrium in a bargaining model," *Econometrica*, vol. 50, no. 1, pp. 97–109, 1982.
- [2] T. Schelling, *The Strategy of Conflict*, Harvard University Press, Cambridge, Mass, USA, 1960.
- [3] R. Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, vol. 4, no. 1, pp. 25–55, 1975.
- [4] A. Dixit, "Thomas Schelling's contributions to game theory," *Scandinavian Journal of Economics*, vol. 108, no. 2, pp. 213–229, 2006.
- [5] R. B. Myerson, "Learning from schelling's strategy of conflict," *Journal of Economic Literature*, vol. 47, no. 4, pp. 1109–1125, 2009.
- [6] J. Hirshliefer, *On the Emotions as Guarantors of Threats and Promises*, MIT Press, Cambridge, Mass, USA, 1987.
- [7] J. Hirshliefer, "Game-theoretic interpretations of commitment," in *Evolution and the Capacity for Commitment*, R. M. Nesse, Ed., Russell Sage Foundation, New York, NY, USA, 2001.
- [8] D. B. Klein and B. O'Flaherty, "A game-theoretic rendering of promises and threats," *Journal of Economic Behavior and Organization*, vol. 21, no. 3, pp. 295–314, 1993.
- [9] V. P. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, no. 6, pp. 1431–1451, 1982.
- [10] J. Farrell, "Meaning and credibility in cheap-talk games," *Games and Economic Behavior*, vol. 5, no. 4, pp. 514–531, 1993.
- [11] J. Farrell and M. Rabin, "Cheap talk," *Journal of Economic Perspectives*, vol. 10, no. 3, pp. 103–118, 1996.
- [12] M. Rabin, "Communication between rational agents," *Journal of Economic Theory*, vol. 51, no. 1, pp. 144–170, 1990.
- [13] S. Baliga and T. Sjöström, "Arms races and negotiations," *Review of Economic Studies*, vol. 71, no. 2, pp. 351–369, 2004.
- [14] T. Schelling, "An essay on bargaining," *The American Economic Review*, vol. 46, pp. 281–306, 1956.
- [15] G. E. Bolton, "Bargaining and dilemma games: from laboratory data towards theoretical synthesis," *Experimental Economics*, vol. 1, no. 3, pp. 257–281, 1998.

## Research Article

# Intermediaries in Trust: Indirect Reciprocity, Incentives, and Norms

Giangiacomo Bravo,<sup>1</sup> Flaminio Squazzoni,<sup>2</sup> and Károly Takács<sup>3</sup>

<sup>1</sup>Department of Social Studies, Linnaeus University, Universitetsplatsen 1, 35195 Växjö, Sweden

<sup>2</sup>Department of Economics and Management, University of Brescia, Via San Faustino 74B, 25122 Brescia, Italy

<sup>3</sup>MTA TK Lendület Research Center for Educational and Network Studies (RECENS), Hungarian Academy of Sciences, Országház Utca 30, Budapest 1014, Hungary

Correspondence should be addressed to Giangiacomo Bravo; [giangiacomo.bravo@lnu.se](mailto:giangiacomo.bravo@lnu.se)

Received 2 September 2014; Accepted 4 December 2014

Academic Editor: Vincent Buskens

Copyright © 2015 Giangiacomo Bravo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Any trust situation involves a certain amount of risk for trustors that trustees could abuse. In some cases, intermediaries exist who play a crucial role in the exchange by providing reputational information. To examine under what conditions intermediary opinion could have a positive impact on cooperation, we designed two experiments based on a modified version of the investment game where intermediaries rated the behaviour of trustees under various incentive schemes and different role structures. We found that intermediaries can increase trust if there is room for indirect reciprocity between the involved parties. We also found that the effect of monetary incentives and social norms cannot be clearly separable in these situations. If properly designed, monetary incentives for intermediaries can have a positive effect. On the one hand, when intermediary rewards are aligned with the trustor's interest, investments and returns tend to increase. On the other hand, fixed monetary incentives perform less than any other incentive schemes and endogenous social norms in ensuring trust and fairness. These findings should make us reconsider the mantra of incentivization of social and public conventional policy.

## 1. Introduction

A trust relationship is an exchange where at least two parties interact, that is, a trustor and a trustee, and in which there is a certain amount of risk for the former. If the trustor decides to place trust, the trustee can honour or abuse it. If honouring trust is costly, as what happened in one-shot interactions and sometimes even in repeated exchanges, the trustee will have no rational incentive to be trustworthy. Knowing this, the trustor is hardly likely to make the first move [1].

Understanding how trust can be established in such hostile situations is of paramount importance. One of the most interesting sociological explanations suggests that social and economic exchanges are embedded in social contexts where certain norms and roles have evolved to mediate between individual interests. For instance, intermediaries might act as advisories and mediators between the parties involved and

reputation or gossip can help to spread information about unknown partners that helps trustors to take the risk of interaction [1–3].

Recent experimental results have shown that individuals can overcome distrust and cooperate more frequently if behaviour in the exchange is observed by a third party [4–8]. This happens even when the opinion of a third party has no consequence on the payoffs of the individuals and reputational building strategies are ruled out [9]. This indicates that, in many real situations, third parties can reduce information asymmetry and temptations of free riding, induce mutual trust, and ensure collective benefit. This requires understanding why and under which conditions information from third parties should be trusted by trustors and what type of incentives can make judgements or recommendations by third parties credible to the eyes of trustors. Indeed, first, intermediaries' opinion is often subjective. Secondly,



it is formulated on trustee's past behaviour in situations where the trustees could have strategically mimicked trust-worthy signals in view of benefits from future trustors. This transforms the problem of trust in a "secondary problem of trust" [10]. Here, the challenge is to understand under what conditions, in anonymous exchanges with unknown partners, intermediaries could be considered reliable and how they should be motivated to provide information that increases trust.

For instance, let us consider certain important social and economic interactions, such as the trust relationships between employees and managers in big companies or between two potential partners for a relationship. In these cases, intermediaries are called on to express their opinion on certain attributes or behaviour, which are crucial to create trust. Sometimes they do so voluntarily, without receiving any material payoffs, such as someone recommending a friend as a partner to another friend. In other cases, intermediaries are financially motivated professionals, such as an HR manager recommending an employee to be upgraded to his/her company manager. Therefore, while in certain spheres of social life, the social function of intermediaries has been institutionalized through material incentives and roles, in other situations informal or voluntary norms have developed.

The aim of our paper was to examine these trust problems in triadic relations (i.e., between trustors, intermediaries, and trustees) to better understand conditions that could transform the intermediary opinion in a trust carrier. We conducted two experiments where subjects played a modified version of the repeated investment game with intermediaries added to the typical trustor/trustee dyadic relation. We manipulated incentives and roles of intermediaries to test their impact on cooperation in particularly hostile conditions. For this, we meant a situation where (i) intermediaries formulated their opinion on the trustee behaviour on a limited set of information and (ii) their opinion was not public and (iii) did not have long-term consequences on the material payoffs of the trustees. In this situation, intermediaries had only a limited set of information (i) and bad standing was not so risky for trustees (ii-iii). Hence, our experimental situation was intentionally designed so that intermediary opinion was poorly credible for trustors. We examined the importance of indirect reciprocity considerations and their interplay with material incentives and social norms. Furthermore, we tested different ways to align intermediary's incentives, for example, respectively, with the trustors' or the trustees' interests, and the impact of roles' rotation.

We compared results from two laboratory experiments Experiment 1, first reported in [11], and Experiment 2, previously unpublished. The latter was designed to carefully examine the effect of indirect reciprocity on the behaviour of intermediaries—and, more generally, on trust at the system level—since this mechanism was crucial in the first experiment. The rest of the paper introduces our research background and hypotheses (Section 2), presents the design of the two experiments (Section 3) and their outcomes (Section 4), and finally discusses certain social and policy implications of the results (Section 5).

## 2. Hypotheses

While dyadic embeddedness is important to explain trust and cooperation in situations involving stable relationships between two actors, in modern society trust is often mediated by agents who facilitate the exchange between trustors and trustees when they cannot rely on past experience [1, 12, 13]. Important examples of this have been empirically found in the development of trust between suppliers and customers in a variety of situations, such as in electronic markets on the web [14, 15] and in the US venture capital market [16]. In these cases, any chance for trust is undermined by information asymmetry between the involved parties and does not simply imply all-none choices, such as trusting or not, but requires a pondered rational calculus of trust investment.

Let us consider the *investment game*, a typical framework to model trust problems [17]. First, Player A (the trustor) receives an initial endowment of  $d_A$  points, with a fixed exchange rate in real money. Then, A is called to decide the amount  $I$  between 0 and  $d_A$  to send to Player B (the trustee), keeping the part  $(d_A - I)$ . The amount sent by A is multiplied by  $m > 1$  and added to B's own endowment  $d_B$ . (Note that in some investments games, including [17], B players had no endowment; that is,  $d_B = 0$ . Trustor investments tend to be lower in experiments where  $d_B > 0$  [18].) Then, B is called to decide the share of the amount received (plus his/her endowment) to return to A. As before, the amount  $R$  returned by B can be between 0 and  $(d_B + mI)$ . The amount not returned represents B's profit, while  $R$  is summed to the part kept by A to form his/her final profit. Therefore, the player's profit is as follows:

$$\begin{aligned} P_A &= d_A - I + R, \\ P_B &= d_B + mI - R. \end{aligned} \tag{1}$$

The structure of the game implies that the trustor can have an interest in investing on condition that the expected returns are higher than his/her own investments, that is, if  $R = q(d_B + mI) > I$ , where  $q$  is the proportion returned by the trustee. However, as the trustor can rationally presume that the trustee has no interest in returning anything, it is rationally expected that the trustor will not invest and trust will not be placed, giving rise to a suboptimal collective outcome of  $(d_A + d_B)$  while the social optimum  $(md_A + d_B)$  could have been reached with sufficient high levels of trust. However, empirical evidence contradicts this prediction. A recent review of 162 experimental replications of the investment game indicated that, on average, trustors invested about 50% (range 22–89%) of their endowment and trustors returned 37% (range 11–81%) of their amount [18].

In this type of games, the crucial challenge is to understand the mechanisms through which the trustor estimates the trustworthiness of the trustee by using available information. Following Coleman [1], we can identify two possible information sources for trustors: (i) direct knowledge of past behaviour of the trustee and (ii) knowledge of trustee behaviour obtained by a third party with positions and interest differently aligned to those of the other parties involved. In both cases, information on past behaviour of

the trustee may help the trustor in predicting the trustworthiness of the trustee, creating in turn reputational incentives for the trustee to overcome any temptation of cheating in view of future benefits.

While it is widely acknowledged that knowing the past behaviour of a trustee can increase a trustor's investment and cooperation in dyadically embedded interaction [19–23], the case of triadic relations is more interesting as these relationships can compensate for the lack of direct knowledge and contacts between actors, so enlarging the social circles and the extent of the exchange [24]. This requires us to understand the complex triangulation of the exchange and especially the role of trust intermediaries, who might have either analogous or different positions and interests to the trustors. When positions and interests between the trustor and the intermediary are aligned, the intermediary can act in the trustor's interest and the latter seriously considers the intermediary's opinion so that his/her decision will reflect the available reputational information. When the intermediary and the trustor do not have aligned interests, the outcome of the exchange is heavily dependent on the motivations behind the intermediary's action [1].

In order to represent triadic relations in trust situations, we modified the standard investment game framework by adding to the trustor (called Player A) and the trustee (called Player B) a third player, that is, the intermediary (called Player C). We assumed that the intermediary could observe B past behaviour (i.e., the amount of returns sent to A compared to the A investment) and was called to decide whether to provide A with honest and accurate information or not. If credible to the eye of A, information by C is expected to help A regulate his/her investment. If this is so, B Players have a rational interest in building a good standing at the eye of C and so behave more fairly with A. Therefore, if intermediaries are trusted both by the trustors and by the trustees it is expected to generate higher levels of cooperation and fairness.

When modelling intermediaries' behaviour, for the sake of simplicity we assumed that C could only choose between two levels of fairness: providing A with an accurate evaluation of B (=high fairness) or sharing inaccurate or deceptive information (=low fairness). Adapting Frey and Oberholzer-Gee's "motivation crowding-out" formalization [25], we also assumed that each intermediary chooses a level of fairness ( $f$ ) that maximizes his/her expected net benefit as follows:

$$\max [p(f)(b + e) + d(e, f) - k]. \quad (2)$$

Since  $m > 1$ , any increase of investments leads to higher stakes to share and a higher level of fairness is beneficial to everybody. In particular,  $b$  indicates the expected benefit for C due to the aggregate fairness, while  $e$  indicates his/her expected private earnings. Given that all these figures are *expected*, the intermediaries' benefits are weighted by the probability  $p$  to reach a given level of fairness  $f$ . The more C plays fair the more he/she contributes to providing a context for fairness ( $p'_f > 0$ ).

It is worth noting that a crucial component of the model is the intrinsic motivation of subjects ( $d$ ). This is expected

to increase with the overall level of fairness ( $d'_f > 0$ ) and to decline with private material earnings due to crowding-out effects ( $d'_e < 0$ ). Finally, we assumed that a fixed small but strictly positive cost  $k$  of fair behaviour existed due the cognitive requirements by C (e.g., information search, memory, and time) to perform a thoughtful evaluation of B.

Assuming that intermediaries choose the level of fairness  $f^*$  maximising their expected benefit, we derived the first order condition as follows:

$$p'_f(b + e) + d'_f = 0. \quad (3)$$

It is important to note that intermediaries could see a benefit  $b$  from higher levels of fairness only if they are expected to play as A or B in the future, that is, if roles are rotating. If roles are fixed,  $b = 0$  and the intermediary's decision depends only on the personal earnings  $e$  and the intrinsic motivation  $d$ . Given (3), *ceteris paribus* a situation where  $b = 0$  is expected to lead to a lower  $f^*$  than where intermediaries can derive benefits from the aggregate level of fairness by playing other roles in the future.

It is worth noting that this can be framed in terms of indirect reciprocity [23, 26, 27]. That is the idea of benefiting unknown trustors by punishing self-interested trustees to keep the fairness standards high and benefit from the reciprocity of other reliable intermediaries when cast as trustors (i.e.,  $b > 0$ ). This can induce intermediaries to provide reliable information and investors to trust reputational information. This concatenation of strategies cannot work in a system where interaction roles are fixed, given that intermediaries cannot expect benefits from their evaluation as future trustors (i.e.,  $b = 0$ ). In this case, trustors cannot expect that intermediaries provide a careful evaluation.

This led us to formulate our first hypothesis as follows.

**Hypothesis 1.** If roles of the interaction are fixed, indirect reciprocity motives cannot motivate positive behaviour by the parties involved, who will not see future benefits in keeping the levels of fairness high. This implies that *ceteris paribus* the fixed role condition will decrease fairness and cooperation.

The situation is different when the interests of intermediaries and trustors are aligned. In this case, intermediaries receive a direct payoff  $e > 0$  by behaving fair. Indeed, this interest alignment transforms the trust relationship in a typical principal-agent model, where the intermediary (the agent) can behave in the interest of the trustor (the principal). In this case, the rational choice theory predicts that monetary incentives are crucial to motivate the intermediary to act on the trustor's behalf, by guaranteeing that the self-interest of the former coincides with the objectives of the latter [28]. The fact that the negative effect of monetary incentives on intrinsic motivations ( $d'_e < 0$ ) could be compensated by higher levels of fairness ( $d'_f > 0$ ) will lead to higher levels of fairness and cooperation compared to the  $e = 0$  case.

This led us to formulate our second hypothesis as follows.

**Hypothesis 2.** If the intermediary responds to monetary incentives that are aligned with the trustor's interests, higher levels of fairness will lead to higher cooperation in the system.

In the symmetric case, where monetary incentives are aligned with the trustee's interests, intermediaries will receive a personal payoff from being unfair. Note also that, in this case, not only do the incentives crowd out intrinsic motivations but also  $d$  declines due to expected lower levels of fairness. This is why the intermediary is in a potential conflict of interest as he/she may be tempted to cheat the trustor by providing opinions that benefit the trustee. Knowing this, the trustor could be induced to question the reliability of the intermediary's opinion and decide not to enter into the exchange or reduce his/her investment to minimal levels. This is expected to erode the basis of cooperation.

This led us to formulate our third hypothesis as follows.

**Hypothesis 3.** When the intermediary's incentives are aligned with the trustee's interests the levels of fairness will decline, leading to lower cooperation in the system in comparison with the situation where no monetary incentives exist.

A particular case is when intermediaries receive monetary incentives that are independent of their actions and the level of fairness in the system, as in the case of fixed rewards. In this case,  $e$  will not enter the benefit calculus as expressed by the first term of (2) as it will be earned in all cases, while the incentive will decrease the intermediaries' intrinsic motivation because of  $d'_e < 0$ , leading to low levels of fairness. This will make the intermediary's opinion poorly credible for both the trustor and the trustee.

This led us to formulate our fourth hypothesis as follows.

**Hypothesis 4.** With intermediary's incentives that are fixed and independent from the interests of both the trustor and the trustee, the levels of fairness and cooperation in the system will be lower than in case of no monetary incentives.

Without any material interest in the exchange ( $e = 0$ ), intermediaries intrinsic motivations are expected to increase with  $f$ . As long as roles change over time, being  $b > 0$  and  $p'_f > 0$ , subjects are expected to personally benefit from higher levels of fairness in the system. Furthermore, the absence of monetary incentives can transform the interaction into a moral problem, with intermediaries induced to punish misbehaviour by trustees even more than in other incentive schemes [29, 30].

This led us to formulate our last hypothesis as follows.

**Hypothesis 5.** With alternating roles and without monetary incentives for intermediaries, fairness will increase leading to high levels of cooperation in the system.

### 3. Methods

To test our hypotheses, we built two experiments based on a modified version of the repeated investment game described above with  $d_A = d_B = 10$  MU and  $m = 3$  and with the restriction of choices to integer amounts (see the Appendix for a detailed description of the experiments). We extended the original dyadic game towards a third-party game where we introduced intermediaries (Players C) not

directly involved in the transaction but asked to rate trustees' behaviour (Players B) for the benefit of the trustors (Players A). The opinion of Players C was formulated individually and was shared with both Players A and B involved in the exchange. When selected as a C Player, the subject was matched with one Player A and one Player B and privately informed of the amount received and returned by the latter in the previous period. Then, he/she was asked to rate Player B's behaviour as "negative," "neutral," or "positive." His/her opinion was privately displayed to Player A before his/her investment decision. The rest of the game worked as the standard dyadic version described above. Note that any communication between subjects was forbidden and subjects played anonymously with possible partners from a large pool of subjects.

In the first experiment, game roles (i.e., trustor, trustee, and intermediary) alternated regularly throughout the game, with the same subject playing the same number of times in the three roles with randomly determined partners [11]. The second experiment followed the same design except that roles were fixed throughout the game. The aim of this second experiment was to rule out any (indirect) reciprocity motive from the intermediary behaviour, as the intermediary now could not provide reliable information to expect good future information in turn (in the trustor's role).

More specifically, in both experiments, monetary incentives for intermediaries systematically varied across treatments according to our hypotheses. In the *No incentive* treatment, intermediaries did not receive any rewards for their task, also losing potential earnings as trustors or trustees when selected to play in this role. In the *Fixed incentive* treatment, intermediaries received a fixed payoff of 10 MU, equal to the trustor and trustee endowments. In the *A incentive* treatment, intermediaries earnings were equal to the payoff obtained by the trustors they advised. In the *B incentive* treatment, intermediaries' earnings were equal to the payoff obtained by the trustees they rated.

## 4. Results

**4.1. Experiment 1 (Alternating Roles).** Our experiment produced investments and returns comparable to previous experiments in the *Baseline* [18] and, consistent with previous studies which introduced reputational motives [9, 22], showed that the presence of the intermediaries dramatically improved cooperation.

A total of 136 subjects (50% female) participated in the experiment, held in late 2010. Both investments and returns were higher when intermediaries were introduced, with investments increasing from an average of 3.22 MU in the *Baseline* up to 5.21 MU in *A incentive* and returns rising from 2.00 in the *Baseline* to 6.87 in *No incentive* (Figure 1). (Our dataset may be accessed upon request to the corresponding author. All statistical analyses were performed using the R 2.15.1 platform [31]. Please, note that the amounts exchanged in the first three periods of the game, when intermediaries had no previous information to evaluate, and in the last three periods, when trustees knew that no further rating would take place, were not included in the analysis.)

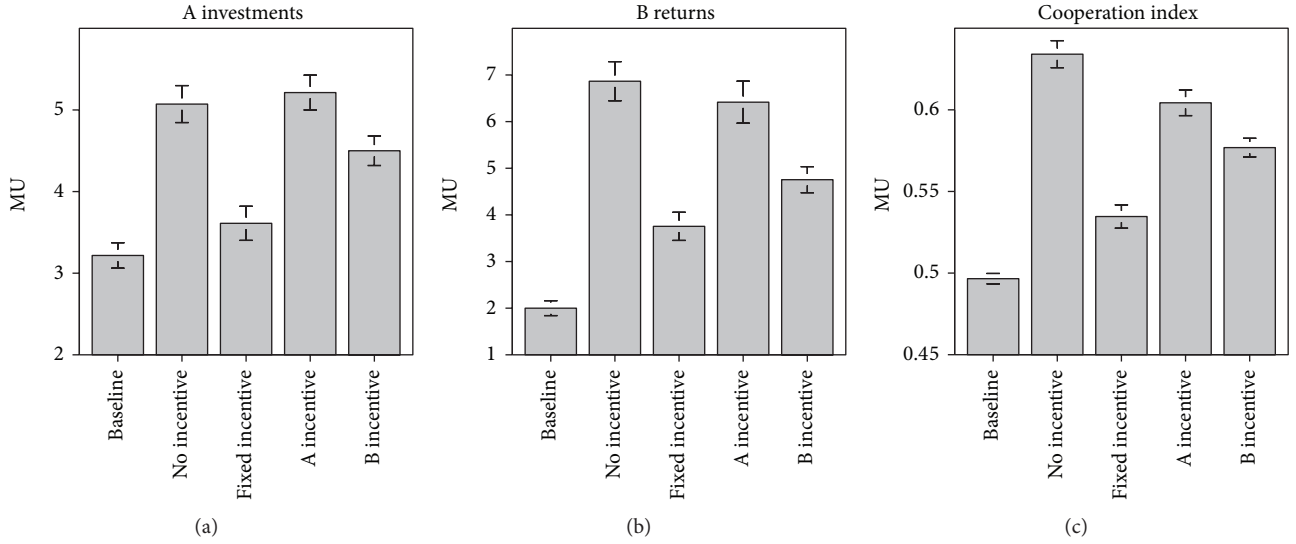


FIGURE 1: Experiment 2 results. Average investments, returns, and CI by treatment. Bars show the standard error of the means.

Differences with the *Baseline* for both investments and returns were significant at the 5% level for all treatments except *Fixed incentive*, where the difference was significant only for returns. Significant differences also existed for Player B returns. Both *No incentive* and *A incentive* led to higher returns than *Fixed incentive* (Wilcoxon rank sum tests on individual averages,  $W = 531.0$ ,  $P = 0.002$  one tailed, and  $W = 199.0$ ,  $P = 0.002$  one tailed, resp.). There were no significant differences between *No incentive* and *A incentive* ( $W = 385.0$ ,  $P = 0.365$ ). Differences in investments were not considerable but still statistically significant at 5% between *No incentive* and *Fixed incentive* ( $W = 508.0$ ,  $P = 0.006$  one tailed) and between *A incentive* and *Fixed incentive* ( $W = 176.5$ ,  $P = 0.001$  one tailed).

To compare the outcome of the different treatments better, we built an indicator that considered both social welfare and fairness, which are two aspects strictly linked with cooperation due to the social dilemma nature of the investment game [32]. Indeed, in any investment game, trust is a double-edged sword. On the one hand, it includes trustors who are better-off by placing trust and so can increase the welfare of the system by taking the risk of interaction. On the other hand, it also includes trustees who can act in a more or less trustworthy way, so contributing to giving rise to a fair exchange. Following [11], in order to look at these two sides of the this process, we built an index that combined attention to the social welfare, which only depended on A investments, and fairness, which depended on B returns, which we called *cooperation index* (CI). The level of social welfare is an important indicator of the system efficiency in the different treatments. This is indicated by  $E = I/d_A$ , where  $I$  was A investment and  $d_A$  was the endowment. This indicator took 0 when A invested nothing and 1 when A invested the whole endowment. Following previous research [33–36], *fairness* was calculated by comparing the difference in the payoffs to the sum of monetary gains:  $F = 1 - [(P_A - P_B)/(P_A + P_B)]$ , where  $P_A$  and  $P_B$  were the payoffs earned by A

and B Players, respectively. This was 0 when one of the players obtained the whole amount at stake and the other received nothing, while it was 1 when both players obtained the same payoff. Note that no trade-off exists between maximising  $E$  and  $F$ , given that, for any level of investment (including zero), only the amount returned determined the fairness payoff. The cooperation index was defined as  $CI = (E + F)/2$ . This was 0 when A Players invested nothing and B Players returned all their endowments, grew with the growth of A investments and a fairer distribution of final payoffs, and became 1 when A Players invested  $d_A$  and B Players returned half of their total endowment, that is,  $(d_B + mI)/2$ .

The treatment with the highest CI was *No incentive*, which led to more fairness than any other treatment (Figure 1). Differences in the CI were statistically significant at 10% with *A incentive* and at 5% with all other treatments. The high CI value in *No incentive* was especially important as in this case, unlike *A incentive*, intermediaries had no monetary incentive to cooperate with trustors. This would confirm that the lack of monetary incentives for intermediaries implied higher normative standards of behaviour for the other actors involved.

**4.2. Experiment 2 (Fixed Roles).** A total of 244 subjects (55% females) participated in the second experiment, which was organized in 2011. Results showed that trust and trustworthiness were generally lower than in the first experiment. Average investments ranged from 2.36 MU in *Fixed incentive* to 3.96 MU in *A incentive*. Returns ranged from 2.47 MU in *Fixed incentive* to 5.27 MU in *A incentive*. The cooperation index ranged from 0.507 MU in the *Baseline* to 0.557 MU in *A incentive* (Figure 2).

The only treatment leading to investments significantly higher than the *Baseline* was *A incentive* ( $W = 82$ ,  $P = 0.049$  one tailed), while returns were significantly higher (at the 10% level) in both *A incentive* and *B incentive* ( $W = 88$ ,  $P = 0.071$



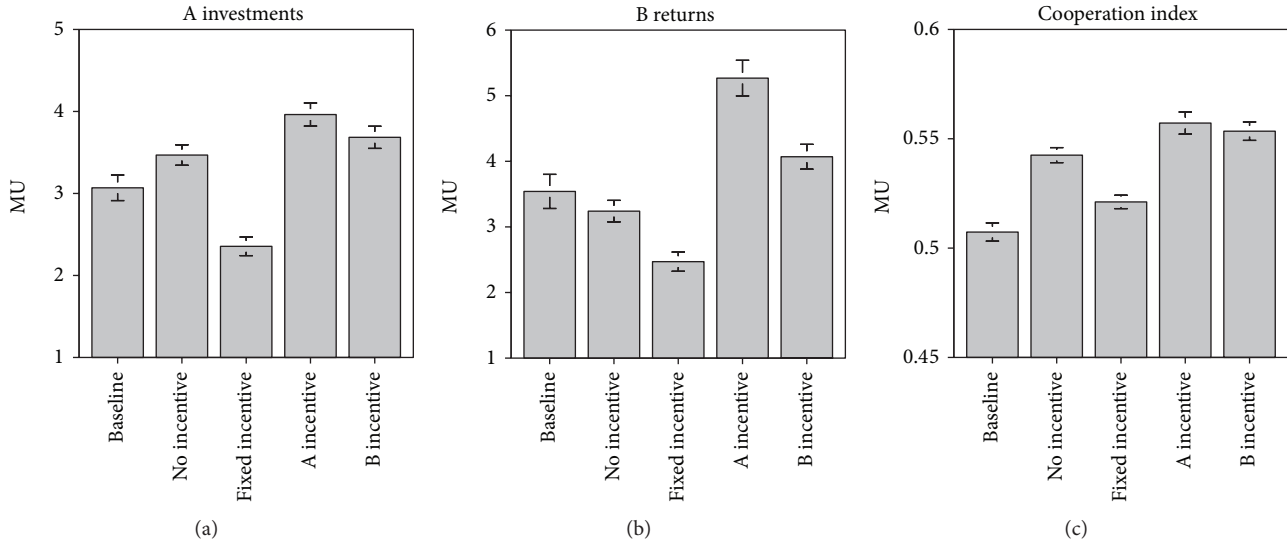


FIGURE 2: Experiment 2 results. Average investments, returns, and CI by treatment. Bars show the standard error of the means.

one tailed for both treatments). Finally, the cooperation index was significantly higher than the *Baseline* in *No incentive* ( $W = 351$ ,  $P = 0.019$  one tailed), *A incentive* ( $W = 299$ ,  $P = 0.003$  one tailed) and *B incentive* ( $W = 312$ ,  $P = 0.004$ ), but not in *Fixed incentive* ( $W = 462$ ,  $P = 0.288$  one tailed).

**4.3. Comparing the Two Experiments.** A comparison between alternating versus fixed role treatments highlighted the effects of indirect reciprocity (Table 1). Besides lower returns in the *Baseline*—a result consistent with the existing literature [18]—investments and returns were generally higher in the alternating roles experiment. This difference was especially relevant in *No incentive*. This was one of the best performing treatments in the first experiment, while, in fixed role treatments, investments and returns were similar to the *Baseline* (i.e., without intermediaries). On the other hand, the two treatments with sound monetary incentives led to a more modest decrease in investment and especially, in returns. Note also that *B incentive* performed similarly in the two experiments.

In order to examine the effect of the different factors involved in the exchange, we estimated a random effects model using dummies indicating the fixed role experiment, each treatment, the first and the last period, and the second half of the game as regressors (Table 2). Due to the considerable interdependence between the observed decisions, random effects (multilevel) regression analysis was performed that considered the nestedness of observations at the individual level. Results showed that fixed roles caused an overall decrease of trust but had less effect on trustworthiness. *A incentive* had a stronger impact on trustee behaviour, while *No incentive* had a stronger impact on CI, that is, on the fairness of the exchange. All conditions except *Fixed incentives* led to higher trust and trustworthiness than the *Baseline*. Finally, all cooperation indicators declined during the game, confirming the typical “end effect” found in previous studies [9, 22, 37].

If we consider the behaviour of the intermediaries, we can see that they generally played fairly, asking B Players to return significantly more to award a more positive rating (Table 3(a)). They were more demanding where trust was higher, namely, in *No incentive* and *A incentive* of the alternating roles experiment. In both cases, to award a positive rating to trustees, intermediaries asked trustees to return about one-third of the trustees’ endowments. On the other hand, in less cooperative treatments such as *No incentive* or *Fixed incentive* in the fixed roles experiment, they were less demanding (i.e., to one-sixth and even less). Moreover, negative ratings were more frequent in the most cooperative treatments. Although only descriptive, these results suggest that a more rigorous reputation process took place in these conditions, despite the overall higher trustworthiness of trustees achieved in these treatments (Table 3(b)).

It is also important to note that trustor investments reflected the behaviour of the intermediary. Generally, trustors invested more when their opponents received positive ratings, less in cases of neutral ratings and even less in cases of negative ratings (Table 4). It is worth noting that also in this case there were no important differences across treatments. Results suggest that trustors trusted intermediaries more when the latter gave positive ratings in *No incentives* and *A incentive* in the alternating roles experiment. In these cases, they invested on average 6.25 MU and 7.04 MU, respectively. They invested less in treatments where there was low trust, such as *Fixed incentive* in the fixed role condition, where average investment of trustors was 2.43 MU even when trustees had a positive rating.

## 5. Discussion

Our experiments have highlighted the crucial role of independent judges in trust situations, even when their opinions could be viewed as subjective and/or without serious long-term future consequences for the trustee. We found that,

TABLE 1: Overview of Experiments 1 and 2 results. The AR/FR column presents the ratio between the corresponding results in the alternating and the fixed role experiment. The last two columns present Wilcoxon rank sum tests of the null hypothesis that the result distribution is the same in the two experiments.

Treatment		Altern. roles		Fixed roles		AR/FR	Wilcoxon	
		Mean	SE	Mean	SE		<i>W</i>	<i>P</i>
Baseline	A investments	3.22	0.16	3.07	0.16	1.05	207.5	0.385
	B returns	2.00	0.16	3.54	0.26	0.56	142.5	0.079
	B ret. (prop.)	0.09	0.01	0.16	0.01	0.56	137.0	0.059
	CI	0.50	0.00	0.51	0.00	0.98	319.0	0.118
No incentive	A investments	5.07	0.23	3.47	0.12	1.46	174.0	0.029
	B returns	6.87	0.42	3.24	0.16	2.12	196.0	0.003
	B ret. (prop.)	0.24	0.01	0.14	0.01	1.71	189.0	0.006
	CI	0.63	0.01	0.54	0.00	1.17	396.0	0.000
Fixed inc.	A investments	3.61	0.21	2.36	0.11	1.53	170.0	0.040
	B returns	3.75	0.30	2.47	0.15	1.52	142.0	0.232
	B ret. (prop.)	0.17	0.01	0.10	0.01	1.70	150.0	0.153
	CI	0.54	0.01	0.52	0.00	1.04	264.0	0.319
A incentive	A investments	5.21	0.21	3.96	0.14	1.32	174.5	0.028
	B returns	6.42	0.45	5.27	0.27	1.22	147.0	0.180
	B ret. (prop.)	0.23	0.01	0.21	0.01	1.10	134.0	0.333
	CI	0.60	0.01	0.56	0.01	1.07	365.0	0.002
B incentive	A investments	4.50	0.18	3.68	0.14	1.22	154.0	0.121
	B returns	4.75	0.28	4.07	0.19	1.17	141.5	0.238
	B ret. (prop.)	0.19	0.01	0.18	0.01	1.06	130.0	0.387
	CI	0.58	0.01	0.55	0.00	1.05	307.0	0.071

TABLE 2: Multilevel regression analysis with individual-level random effects. Significance codes: \*\*\* $P < 0.001$ , \*\* $P < 0.01$ , \* $P < 0.05$ ,  $^{\dagger}P < 0.1$ .

Dependent	Investments	Returns	CI
(Intercept)	3.607***	0.154	0.516***
Fixed roles	-0.962***	-0.133	-0.030***
No incentive	1.339**	1.509**	0.090***
Fixed incentive	0.040	0.684	0.030*
A incentive	1.538***	2.004***	0.083***
B incentive	1.075**	1.012 $^{\dagger}$	0.066***
First period	0.066	0.954***	0.024*
Last period	-0.562**	-0.848**	-0.031**
Periods 16–30	-0.313***	-0.420***	-0.010*
A investment		0.853***	
<i>Random effects</i>			
$\sigma$ (id)	1.827	2.553	0.054
$\sigma$ (residual)	2.314	2.971	0.118
Number of interactions	4070	4070	4070
Number of individuals	222	222	222
<i>F</i>	21.5***	320.7***	121.1***

somewhat counterintuitively, the intermediary’s opinions had a stronger effect when the intermediary had no material interest from the exchange. If there was room for indirect reciprocity between the parties involved, triadic relationships could provide a moral base to overcome the negative traps

of self-interest. We also found that monetary incentives and moral action were not clearly separable or substitutes [38]. If properly designed, monetary incentives for intermediaries had a positive effect, especially on trustor’s investments, but were less effective than social norms and reciprocity in ensuring fair exchanges in terms of more equal distribution of payoffs between trustors and trustees.

First, the “shadow of reciprocity” implied that intermediaries kept the standards of evaluation high to benefit good information in turn by other intermediaries, when acting as trustors. This in turn induced trustees to be more reliable. Under evaluation, trustees overestimated the future impact of their good standing and behaved more fairly. On the other hand, having information about trustee behaviour induced trustors to predict higher cooperative responses by trustees and increase their investment. It is worth noting that recent work showed that this mechanism is relatively independent of the quality of information. For instance, in an experiment on financial decisions in situations of uncertainty and information asymmetry similar to our experiment, [39] showed that the availability of information from other subjects increased risky trust investment decisions of subjects independent of the quality of the information.

It is worth noting that the motivation crowding-out model (2) was able to match our empirical data qualitatively. Coherently with H1, the comparison of the two experiments confirmed the paramount importance of indirect reciprocity motives in trust situations. The idea that indirect reciprocity is fundamental in human societies has been suggested by [40],

TABLE 3: Summary statistics by rating. Average return proportion by rating (a) and rating distribution per treatment (b). AR stands for “alternating roles” and FR for “fixed roles.”

(a) Return proportion			
Treatment	Negative	Neutral	Positive
No incentive AR	0.14	0.22	0.34
No incentive FR	0.13	0.15	0.14
Fixed incentive AR	0.08	0.16	0.24
Fixed incentive FR	0.11	0.09	0.17
A incentive AR	0.15	0.22	0.33
A incentive FR	0.18	0.16	0.28
B incentive AR	0.14	0.20	0.22
B incentive FR	0.19	0.15	0.18
All treatments	0.15	0.16	0.22

(b) Rating distribution			
Treatment	Negative	Neutral	Positive
No incentive AR	0.42	0.22	0.36
No incentive FR	0.37	0.26	0.37
Fixed incentive AR	0.34	0.31	0.35
Fixed incentive FR	0.36	0.30	0.34
A incentive AR	0.41	0.25	0.35
A incentive FR	0.37	0.23	0.40
B incentive AR	0.33	0.29	0.38
B incentive FR	0.33	0.22	0.44
All treatments	0.36	0.26	0.38

TABLE 4: Average investments by treatment and rating.

Treatment	Negative	Neutral	Positive
No incentive AR	3.72	5.76	6.25
No incentive FR	2.39	3.24	4.62
Fixed incentive AR	1.84	4.05	4.81
Fixed incentive FR	2.18	2.44	2.43
A incentive AR	3.43	5.02	7.04
A incentive FR	2.75	4.46	4.71
B incentive AR	3.39	4.52	5.27
B incentive FR	2.79	4.00	4.10
All treatments	2.72	3.87	4.56

who argued that this is one of the most crucial forces in human evolution. This has been found in different experimental settings [21, 23]. Our experiments suggest that, *ceteris paribus*, indirect reciprocity explains a significant part of cooperation also in triadic relations (Tables 1 and 2). This indicates that evaluation systems where the roles of trustors, trustees, and evaluators rotate could work better and more efficiently than those in which roles are fixed.

H2 was also confirmed by our data. In both experiments A incentive guaranteed high levels of cooperation as intermediary evaluations fostered both investments by trustors and returns by trustees. In this treatment, trustees followed reputation building strategies and so returned more. Furthermore, trustors considered the opinions of intermediaries

credible and used them to discriminate between “good” and “bad” opponents. Therefore, intermediaries were functional to encapsulate mutual trust and cooperation.

H3 was the only hypothesis not fully supported by our data. On the one hand, B incentive led to less cooperation than A incentive. On the other hand, B incentive unexpectedly gave better results than the Baseline, as intermediaries were less demanding than in A incentive to award positive ratings but still discriminated between trustworthy and untrustworthy B Players (Table 3). This induced trustors to consider the opinion of reliable intermediaries and followed their ratings independent of the misalignment of mutual interests caused by the monetary incentives (Table 4). This would again confirm that, in condition of information asymmetry, any information on trustee is better than none, as in the Baseline.

The mismatch between monetary incentives and intermediary’s behaviour is interesting. Results indicated that intermediaries did not predictably respond to incentives as they played fairly in each treatment (Table 3). This can be explained in terms of intrinsic motivations and the sense of responsibility that are typically associated with such “neutral” positions. More specifically, it is worth noting that willingness to provide pertinent judgement by intermediaries was not sensitive to any variations in the incentive scheme. This would testify to the inherent moral dimension of this role. On the other hand, even the perception that their task was indirectly judged by the trustors, who were informed of their ratings, could have motivated the intermediaries to take their role seriously, independent of the incentives. More importantly, the alternating roles protocol allowed intermediaries to follow indirect reciprocity strategies and so they kept the credibility and quality of their ratings high in order to benefit from reciprocity by other intermediaries when cast as trustors. This explains the difference between the two experiments.

Consistent with H4, Fixed incentive led to less cooperation than other schemes. A comparison between No incentive and Fixed incentive allows us to understand this point better. In the alternating roles experiment, while No incentive led to more trust and cooperation, Fixed incentive barely improved the Baseline, that is, when intermediaries were not present. This does not make sense in a rational choice perspective, as in both cases the incentives of intermediaries were ambiguous. On the other hand, while intrinsic motivations were crucial to induce intermediaries to formulate reliable ratings in No incentive, these aspects were crowded out by monetary incentives in Fixed incentive. This is consistent with the *motivation crowding theory* [25, 41] and with many empirical studies that showed that incentivization policies which targeted self-interested individuals actually backfire by undermining individual “moral sentiments” in a variety of social and economic situations [29, 38]. In this sense, some possible extensions of our work could examine the relationship between crowding out effects on social norms and the presence of different incentive schemes in more detail. For instance, it would be interesting to understand if different incentives for the intermediaries could influence

future interactions between trustors and trustees without intermediaries. This could also help us to look at self-reinforcing effects of incentives and social norms on trust.

It is worth considering that more consistent monetary incentives for intermediaries could increase their credibility for the other parties involved, so motivating more cooperation. Indeed, a possible explanation of the low cooperation in *Fixed incentive* is that the magnitude of our monetary incentives was sufficient to crowd out intrinsic motives of subjects without promoting reciprocal and self-interested behaviour, as what happened in typical monetary markets [29]. Further work is necessary to examine this hypothesis by testing fixed incentives of different magnitudes, although it is worth considering that there are constraints in terms of magnitude of incentives that can be implemented also in real situations.

H5 was also confirmed by our results, at least in the alternating roles experiment. In this case, *No incentive* was the best treatment for cooperation. On the other hand, it did not promote trust and especially, trustworthiness when roles were fixed. As argued above, by fixing the roles, there was no room for indirect reciprocity strategies. The fact that intermediaries could expect future benefits from their roles and were subsequently cast as both trustors and trustees induced the parties involved to believe more in the credibility of the intermediaries' opinion. Also in this case, further empirical investigation is needed to compare social situations where intermediaries have a specialised role with situations where there is voluntarism and mixture of roles.

These findings imply that the "mantra" of incentivization popularized by most economists as a means to solve trust and cooperation problems, especially in economic and public policy, should be seriously reconsidered [42]. Not only should incentives be properly designed to produce predictable outcomes and this is often difficult, but also incentive-response behaviour of individuals is more heterogeneous and unpredictable than expected [43]. If this occurred in a simple laboratory game, where individuals had perfect information and the rules of the game were fully intelligible, one should expect even more heterogeneity of individual behaviour and unpredictability of social outcomes in real situations.

Our results finally suggest that insisting on incentivization potentially crowds out other social norms-friendly, endogenous mechanisms such as reputation, which could ensure socially and economically consistent results. While incentives might induce higher investment risks (but only if properly designed), social norms can also help to achieve a fairer distribution, nurturing good behaviour which can be even more endogenously sustainable in the long run [44].

## Appendices

### A. Details of the Experiment Organization

All participants were recruited using the online system ORSEE [45]. They were fully informed and gave their consent

when they voluntarily registered on ORSEE. Data collection fully complied with Italian law on personal data protection (D.L. 30/6/2003, n. 196). Under the applicable legal principles on healthy volunteers' registries, the study did not require ethical committee approval. All interactions were anonymous and took place through a computer network equipped with the experimental software z-Tree [46].

*Experiment 1 (Alternating Roles).* A total of 136 subjects (50% female) participated in the experiment held at the GECS experimental lab (see <http://www.eco.unibs.it/gecs/>) in the late 2010. Each experimental session took less than one hour and participants earned, on average, 14.82 Euro, including a 5-Euro show-up fee.

Twenty-eight subjects participated in a *Baseline* repeated investment game (hereafter *Baseline*) set using the following parameters:  $d_A = d_B = 10$  monetary units (MU),  $m = 3$ . Each MU was worth 2.5 Euro Cents and subjects were paid in cash immediately at the end of the experiment. The game was repeated 30 times with couples who were randomly reshuffled in each period. Players' roles regularly alternated throughout the game. This meant that each subject played exactly 15 times as A and 15 times as B.

All the other treatments, each played by 27 subjects, introduced a third player into the game (Player C) in the role of intermediary. Once C Players were introduced, we varied the monetary incentive schemes offered to them. In the *No incentive* treatment, intermediaries did not receive any rewards for their task, also losing potential earnings as trustors or trustees when selected to play in this role. In the *Fixed incentive* treatment, intermediaries received a fixed payoff of 10 MU, equal to the trustor and trustee endowments. In the *A incentive* treatment, intermediaries earnings were equal to the payoff obtained by the trustors they advised. In the *B incentive* treatment, intermediaries' earnings were equal to the payoff obtained by the trustees they rated.

*Experiment 2 (Fixed Roles).* A total of 244 subjects (55% female) participated in the second experiment, which was organized at the GECS experimental lab in the late 2011. Participants were recruited and played as in the first experiment. Each experimental session took less than one hour and participants earned, on average, 14.78 Euro, including a 5-Euro show-up fee.

Any overlap with the first experiment participants was avoided, so that the two experiments could in principle be viewed as a single experiment with a between-subject design. The treatments were as before, with the only difference that roles remained fixed throughout the game. Note that the fact that roles no longer alternated actually reduced the sample of observations per role, with a considerable consequence especially on the treatments involving three parties (i.e., all but the *Baseline*). To overcome this problem, we doubled the number of subjects participating in *No incentive*, *Fixed incentive*, *A incentive*, and *B incentive*, which were played by 54 subjects each, organized in two sessions, each one involving 27 participants.



## B. Instructions

Before the beginning of the game, participants read the game instruction on their computer screen. Subsequently, they filled a short test designed to check their understanding of the game. Participants could also ask the experimenters any further questions or issues. Here is the English translation of the original Italian instructions. Sentences in italics are treatment or experiment specific, whereas a normal font is used for instructions common to all treatments/experiments.

### Screen 1: Overall Information on the Experiment

- (i) All these instructions contain true information and are the same for all participants.
- (ii) Please, read them very carefully. At the end, some questions will be asked by the system to test your understanding of the experiment.
- (iii) The experiment concerns economic problems.
- (iv) During the experiment, you will be asked to take decisions, upon which your final earnings will depend. Earnings will be paid in cash at the end of the experiment.
- (v) Each decision will take place through your computer screen.
- (vi) During the experiment, it is prohibited to talk with anyone. If you do so, you will be excluded from the experiment and you will lose your earnings. Please, turn your mobile phones off.
- (vii) For any information and question, put your hands up and wait until an experimenter comes to your position.
- (viii) During the experiment, virtual monetary units (MU) are used that have a fixed exchange rate with real Euros. For each MU earned in the experiment, you will receive 2.5 Euro Cents. For example, if at the end of the experiment, your earning is 500 MU, this means that you will receive 12.50 Euros, plus a fixed show-up fee of 5 Euros.

### Screen 2: Interaction Rules

- (i) The experiment consists of a sequence of interaction rounds between pairs of players (*Baseline*).
- (ii) The experiment consists of a sequence of interaction rounds between groups of three players (*all three person treatments*).
- (iii) Pairs are randomly matched and change each round; therefore, they are made up of different individuals each round (*Baseline*).
- (iv) Groups are randomly matched and change each round; therefore, they are made up of different individuals each round (*all three person treatments*).
- (v) There is no way to know who you are playing with, nor is it possible to communicate with her/him.

- (vi) In each pair, each participant will perform a different role; roles are called "Player A" and "Player B" (*Baseline*).
- (vii) In each group, each participant will perform a different role; roles are called "Player A," "Player B," and "Player C" (*all three person treatments*).
- (viii) Roles are randomly assigned at the beginning of the experiment and then regularly changed for the rest of it. For example, one participant in the pair will play a sequence of rounds such as Player A, Player B, Player A, Player B, Player A, and Player B, and the other one will play a sequence such as Player B, Player A, Player B, Player A, Player B, and Player A. Therefore, over the experiment, all participants will play both roles the same number of rounds (*alternating roles experiment, Baseline*).
- (ix) Roles are randomly assigned at the beginning of the experiment and then regularly changed for the rest of it. For example, one participant in the group will play a sequence of rounds such as Player A, Player B, Player C, Player A, Player B, and Player C, and another one will play a sequence such as Player B, Player C, Player A, Player B, Player C, and Player A. Therefore, over the experiment, all participants will play all roles the same number of rounds (*alternating roles experiment, all three person treatments*).
- (x) Roles are randomly assigned at the beginning of the experiment. This means that each participant will play in the same role throughout the whole experiment (*fixed roles experiment, all treatments*).
- (xi) Each participant should make one decision each round.
- (xii) The experiment lasts 30 rounds.

### Screen 3: Task Structure

- (i) Player A plays first and receives an endowment of 10 MU. She/he has to decide how much of it to keep for her/himself and how much to send to Player B. Player A can send to Player B any amount, from 0 MU to the whole endowment (=10 MU).
- (ii) The amount of MU kept by Player A is part of her/his earning. The amount of MU sent to Player B is tripled and assigned to Player B.
- (iii) Player B also receives an endowment of 10 MU, as Player A did.
- (iv) Example 1: if Player A sends 2 MU, Player B receives  $2 \times 3 + 10 = 16$  MU.
- (v) Example 2: if Player A sends 5 MU, Player B receives  $5 \times 3 + 10 = 25$  MU.
- (vi) Example 3: if Player A sends 8 MU, Player B receives  $8 \times 3 + 10 = 34$  MU.
- (vii) Player B should decide how much of the whole amount received to send to Player A; Player B can send any amount to Player A from 0 MU to the entire sum.

- (viii) The amount of MU kept by Player B represents her/his earning; the amount of MU sent to Player A will add up to Player A earning.
- (ix) Before A and B players, respectively, take their decisions, Player C is asked to rate the decision Player B took the previous round (since no decision has been taken by B in the first round yet, C is not asked to rate in the first round) (*all three person treatments*).
- (x) Player C can rate Player B's decision as "negative," "neutral," or "positive" and the rating is communicated to both A and B Players with whom he/she is grouped (in the first rounds, since there is no decision undertaken by Player B yet, the rating of Player B is assigned by the system as "unknown") (*all three person treatments*).
- (xi) Player C does not receive any earnings for her/his action (*No incentive treatments*).
- (xii) Player C earning is fixed and is equal to 10 MU (*Fixed incentive treatments*).
- (xiii) Player C earning is equal to what Player A will earn at the end of the round (*A incentive treatments*).
- (xiv) Player C earning is equal to what Player B will earn at the end of the round (*B incentive treatments*).
- (xv) Earnings will be added up each round to give the final reward for each participant, which will be paid at the end of the experiment.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors would like to gratefully acknowledge comments on a preliminary version of this paper by Karl-Dieter Opp, Gianluca Carnabuci, Michael Maes, and the audience at the INAS Conference on Analytical Sociology, Columbia University, New York, 8-9 June, 2012, and at the ASA 2012 Annual Meeting, Denver, 17-20 August, 2012. They also would like to thank Marco Castellani and Niccolò Casnici for help during the experiment and Fabio Recinella for technical assistance in the lab. This work was supported by the Collegio Carlo Alberto and the University of Brescia. The authors also gratefully acknowledge support from the Intra-European Fellowship Program of the European Union (Grant no. GA-2009-236953).

## References

- [1] J. S. Coleman, *Foundations of Social Theory*, Harvard University Press, Harvard, Mass, USA, 1990.
- [2] R. S. Burt and M. Knez, "Kinds of third-party effects on trust," *Rationality and Society*, vol. 7, no. 3, pp. 255-292, 1995.
- [3] K. S. Cook and R. Hardin, "Norms of cooperativeness and networks of trust," in *Social Norms*, M. Hechter and K.-D. Opp, Eds., pp. 327-347, Russel Sage Foundation, New York, NY, USA, 2001.
- [4] M. Bateson, D. Nettle, and G. Roberts, "Cues of being watched enhance cooperation in a real-world setting," *Biology Letters*, vol. 2, no. 3, pp. 412-414, 2006.
- [5] I. Bohnet and B. S. Frey, "The sound of silence in prisoner's dilemma and dictator games," *Journal of Economic Behavior and Organization*, vol. 38, no. 1, pp. 43-57, 1999.
- [6] T. N. Cason and V.-L. Mui, "A laboratory study of group polarisation in the team dictator game," *Economic Journal*, vol. 107, no. 444, pp. 1465-1483, 1997.
- [7] K. J. Haley and D. M. T. Fessler, "Nobody's watching? Subtle cues affect generosity in an anonymous economic game," *Evolution and Human Behavior*, vol. 26, no. 3, pp. 245-256, 2005.
- [8] M. Rigdon, K. Ishii, M. Watabe, and S. Kitayama, "Minimal social cues in the dictator game," *Journal of Economic Psychology*, vol. 30, no. 3, pp. 358-367, 2009.
- [9] R. Boero, G. Bravo, M. Castellani, and F. Squazzoni, "Reputational cues in repeated trust games," *Journal of Socio-Economics*, vol. 38, no. 6, pp. 871-877, 2009.
- [10] M. Bacharach and D. Gambetta, "Trust in signs," in *Trust and Society*, K. S. Cook, Ed., vol. 2, pp. 148-184, Russell Sage, New York, NY, USA, 2001.
- [11] F. Squazzoni, G. Bravo, and K. Takács, "Does incentive provision increase the quality of peer review? An experimental study," *Research Policy*, vol. 42, no. 1, pp. 287-294, 2013.
- [12] K. Cook, R. Hardin, and M. Levi, *Cooperation without Trust*, Russell Sage, New York, NY, USA, 2005.
- [13] R. Hardin, *Trust and Trustworthiness*, Russell Sage, New York, NY, USA, 2004.
- [14] J. P. Bailey and Y. Bakos, "An exploratory study of the emerging role of electronic intermediaries," *International Journal of Electronic Commerce*, vol. 1, no. 3, pp. 7-20, 1997.
- [15] J. W. Palmer, J. P. Bailey, and S. Faraj, "The role of intermediaries in the development of trust on the WWW: the use and prominence of trusted third parties and privacy statements," *Journal of Computer-Mediated Communication*, vol. 5, no. 3, 2000.
- [16] O. Sorenson and T. E. Stuart, "Syndication networks and the spatial distribution of venture capital investments," *American Journal of Sociology*, vol. 106, no. 6, pp. 1546-1588, 2001.
- [17] J. Berg, J. Dickhaut, and K. McCabe, "Trust, reciprocity and social history," *Games and Economic Behavior*, vol. 10, no. 1, pp. 122-142, 1995.
- [18] N. D. Johnson and A. A. Mislin, "Trust games: a meta-analysis," *Journal of Economic Psychology*, vol. 32, no. 5, pp. 865-889, 2011.
- [19] V. Buskens and W. Raub, "Rational choice research on social dilemmas: embeddedness effects on trust," in *Handbook of Rational Choice Social Research*, R. Wittek, T. A. B. Snijders, and V. Nee, Eds., Russell Sage, New York, NY, USA, 2008.
- [20] V. Buskens, W. Raub, and J. van der Veer, "Trust in triads: an experimental study," *Social Networks*, vol. 32, no. 4, pp. 301-312, 2010.
- [21] E. Fehr and U. Fischbacher, "The nature of human altruism," *Nature*, vol. 425, no. 6960, pp. 785-791, 2003.
- [22] C. Keser, "Experimental games for the design of reputation management systems," *IBM Systems Journal*, vol. 42, no. 3, pp. 498-506, 2003.
- [23] I. Seinen and A. Schram, "Social status and group norms: indirect reciprocity in a repeated helping experiment," *European Economic Review*, vol. 50, no. 3, pp. 581-602, 2006.

- [24] D. Barrera and G. G. Van De Bunt, "Learning to trust: networks effects through time," *European Sociological Review*, vol. 25, no. 6, pp. 709–721, 2009.
- [25] B. S. Frey and F. Oberholzer-Gee, "The cost of price incentives: an empirical analysis of motivation crowding-out," *American Economic Review*, vol. 87, no. 4, pp. 746–755, 1997.
- [26] R. D. Alexander, *The Biology of Moral Systems*, Basic Books, New York, NY, USA, 1987.
- [27] M. A. Nowak and K. Sigmund, "The dynamics of indirect reciprocity," *Journal of Theoretical Biology*, vol. 194, no. 4, pp. 561–574, 1998.
- [28] J.-J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, Princeton, NJ, USA, 2002.
- [29] D. Ariely, A. Bracha, and S. Meier, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially," *American Economic Review*, vol. 99, no. 1, pp. 544–555, 2009.
- [30] J. Heyman and D. Ariely, "Effort for payment: a tale of two markets," *Psychological Science*, vol. 15, no. 11, pp. 787–793, 2004.
- [31] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2012.
- [32] P. Kollock, "Social dilemmas: the anatomy of cooperation," *Annual Review of Sociology*, vol. 24, pp. 182–214, 1998.
- [33] I. Almås, A. W. Cappelen, E. Ø. Sørensen, and B. Tungodden, "Fairness and the development of inequality acceptance," *Science*, vol. 328, no. 5982, pp. 1176–1178, 2010.
- [34] E. Fehr and K. M. Schmidt, "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.
- [35] M. A. Nowak, K. M. Page, and K. Sigmund, "Fairness versus reason in the ultimatum game," *Science*, vol. 289, no. 5485, pp. 1773–1775, 2000.
- [36] M. Rabin, "Incorporating fairness into game theory and economics," *American Economic Review*, vol. 83, pp. 1281–1302, 1993.
- [37] F. Cochard, P. Nguyen Van, and M. Willinger, "Trusting behavior in a repeated investment game," *Journal of Economic Behavior and Organization*, vol. 55, no. 1, pp. 31–44, 2004.
- [38] S. Bowles, "Policies designed for self-interested citizens may undermine 'the moral sentiments': evidence from economic experiments," *Science*, vol. 320, no. 5883, pp. 1605–1609, 2008.
- [39] R. Boero, G. Bravo, M. Castellani, and F. Squazzoni, "Why bother with what others tell you? An experimental data-driven agent-based model," *Journal of Artificial Societies and Social Simulation*, vol. 13, no. 3, article 6, 2010.
- [40] M. A. Nowak and R. Highfield, *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*, Free Press, New York, NY, USA, 2011.
- [41] B. S. Frey and R. Jegen, "Motivation crowding theory," *Journal of Economic Surveys*, vol. 15, no. 5, pp. 589–611, 2001.
- [42] F. Squazzoni, "A social science-inspired complexity policy: beyond the mantra of incentivization," *Complexity*, vol. 19, pp. 5–13, 2014.
- [43] E. Shafir, *The Behavioral Foundations of Public Policy*, Princeton University Press, Princeton, NJ, USA, 2013.
- [44] S. Bowles and S. Polanía-Reyes, "Economic incentives and social preferences: substitutes or complements?" *Journal of Economic Literature*, vol. 50, no. 2, pp. 368–425, 2012.
- [45] B. Greiner, "An online recruitment system for economic experiments," in *Forschung und Wissenschaftliches Rechnen 2003*, K. Kremer and V. Macho, Eds., pp. 79–93, Gesellschaft für wissenschaftliche Datenverarbeitung, Göttingen, Germany, 2004.
- [46] U. Fischbacher, "Z-Tree: zurich toolbox for ready-made economic experiments," *Experimental Economics*, vol. 10, no. 2, pp. 171–178, 2007.

## Research Article

# Altruism, Noise, and the Paradox of Voter Turnout: An Experimental Study

**Sarah A. Tulman**

*United States Department of Agriculture, Economic Research Service, 1400 Independence Avenue SW, Mail Stop 1800, Washington, DC 20250-0002, USA*

Correspondence should be addressed to Sarah A. Tulman; [sarah.tulman@ers.usda.gov](mailto:sarah.tulman@ers.usda.gov)

Received 4 December 2014; Revised 17 March 2015; Accepted 20 March 2015

Academic Editor: Arthur Schram

Copyright © 2015 Sarah A. Tulman. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper addresses the paradox of voter turnout, wherein observed voting participation rates are far greater than what rational choice theory would predict. Voters face multiple voting choices, stochastic voting costs, and candidates offering different economic platforms. A combination of two approaches attempts to resolve this paradox: quantal response equilibrium (QRE) analysis, which introduces noise into the decision-making process, and the possibility of ethical (altruism-motivated) voting. A series of laboratory experiments empirically tests the predictions of the resulting model. Participants in the experiments are also given opportunities for communicating online with their immediate neighbors, in order to enhance the chances that subjects would realize the possibility of ethical voting. The results show that ethical voting occurs but gains momentum only in the presence of a vocal advocate and even then it mostly dissipated by the second half of the session. The QRE-based model was able to explain some but not all of the overvoting that was observed, relative to the Nash equilibrium prediction. There is evidence to suggest that communication via the chat feature generated some of the voting and also some of the ethical voting.

## 1. Introduction

The paradox of voter turnout, wherein observed turnout rates in elections are far greater than what is predicted under rational choice theory, has been a focus of researchers in economics and political science for over 50 years (see Geys [1] for a survey of the theoretical literature) and remains unresolved. Under rational choice theory, the probability of casting a pivotal vote—making or breaking a tie—is a key component of the motivation to vote. If the electorate is large then the probability of casting the pivotal vote falls to almost zero, and rational choice theory predicts that no one will vote. This probability declines quickly, so that even elections with relatively small electorates face this issue. And yet, observed turnout rates in elections are far greater than zero. For example, in the 2012 general election in the United States the estimated turnout rate among eligible voters was 58.6%. (All turnout rates are from the United States Election Project, <http://www.electproject.org>.) Even in years without a presidential election, when turnout is generally lower, it is still considerably higher than zero, such as the 41.8% turnout rate in the 2010 House and Senate elections. Therefore, there has

to be some other motivation for voting beyond the probability of one's vote being pivotal.

Why does this paradox matter? As the fields of election forecasting and polling analysis have expanded in order to take advantage of the explosion in available data, interest in these forecasts has rapidly increased. During the 2012 US presidential campaign there was an endless flow of political and polling analysis covered in almost every media outlet, and the presidential campaigns devoted vast resources to “get out the vote” efforts while using the rapidly increasing quantities of available data to microtarget voters. In spite of this increasing sophistication, one thing that can and often does cause election outcomes to diverge from forecasts is the matter of who actually votes on Election Day.

The model presented in this paper and the laboratory experiments which test its predictions focus on the economic motivations for voting. The payoffs that are offered by the candidates, and the equitability (or lack thereof) of these payoffs across groups, can be thought of as their economic platforms. Voters also face a cost associated with voting that models the explicit and implicit costs of going to the polls, such as transportation costs, lost wages due to taking time



off from work, or the opportunity cost of time. I draw from several threads of the literature which are at the intersection of economics and political science, and my model and experiments augment previous work on voting behavior.

The first thread from which my model and experiments are drawn focuses on the use of quantal response equilibrium (QRE), a generalization of the Nash equilibrium that allows for decision errors, to relax the assumption of perfect rationality. Building upon the framework presented in McKelvey and Palfrey [2], and also drawing from Goeree and Holt [3], Levine and Palfrey [4] used laboratory experiments to test voter turnout predictions in the presence of heterogeneous and privately known participation costs. However, in their paper the subjects could only choose between voting and abstaining, whereas in most elections potential voters can choose between at least two candidates, plus the option of abstaining. Would their results hold up if another choice were added so that potential voters could choose between three options: vote for one candidate, vote for the other candidate, and abstain? The model and experimental design presented here extend those of Levine and Palfrey [4] in this manner.

Another thread of the literature focuses on the role of ethical, or altruism-motivated, voting in generating voter turnout. Morton and Tyran [5] separated the motivations for voting choices into selfish voting and ethical voting, following the terminology used in the literature—voting for one's economic self-interest versus voting altruistically if one of the candidates offers a lower payoff but also a more equitable distribution of payoffs across groups of voters; however, their equilibrium predictions were qualitative in nature. By incorporating a similar payoff structure into the QRE-based framework from Levine and Palfrey [4], my model can generate quantitative predictions of turnout rates against which the experimental results can be compared. Additionally, it can be argued that incorporating bounded rationality and noise into the decision-making process may lead to a more realistic representation of a world in which individuals do not always (indeed, almost never) behave perfectly rationally. This combination was also used in Goeree et al. [6].

Another aspect of Morton and Tyran's [5] approach, albeit one that was necessary given the nature of the experiment, was that their experiment was a one-shot game. Although this is a more accurate representation of elections than a repeated game, in an actual election voters usually have several weeks or months leading up to the election during which voters can learn about and discuss the issues and the candidates' platforms. My experiments allowed for these learning effects during the experiment, through the inclusion of 20 rounds (elections).

My experiments also incorporated an online chat feature which allowed subjects to communicate before each round. The initial purpose of this was an attempt to increase the probability that the subjects would grasp the possibility of ethical voting, in treatments where that was relevant—the hope was that the subjects could clarify this (and other points) for each other. In the end, although this effect was indeed observed in certain cases, it also provided insight into subjects' motivations. The chat feature was also found to increase voting participation rates.

Group identity and communication within those groups have also been shown to increase voting participation, as seen in papers by Morton [7], Großer and Schram [8], Schram and van Winden [9], Schram and Sonnemans [10], and Charness et al. [11]. Kittel et al. [12], which was developed concurrently with this paper, incorporated multiple-candidate elections, costly voting as in Levine and Palfrey [4], and pre-voting communication. Although the focus and context of their paper differ from mine, they found that the distribution of earnings was more equitable when intergroup communication was allowed (analogous in some respects to the mixed-type chat groups in this paper). Their results also show that voting participation increases when communication is allowed, which matches the positive effect of communication on voting turnout that I found.

The main findings of this paper are summarized as follows. High rates of voter turnout were observed in the experimental data, especially when a subject voted for the candidate offering that subject's voter type a higher payoff than what the other candidate was offering. Some ethical voting was observed, but not enough to explain the overvoting. Also, most of the overvoting occurred when voters of each type voted in their own economic self-interest and, by its very nature, ethical voting is not a possible motive for that.

One feature that could potentially explain overvoting, unexpectedly, is the chat feature. (This feature was not included in the experimental design for the purpose of increasing voter participation but rather as a method to hopefully spread the word about the possibility of ethical voting and also to gain insight into the motivations behind decisions. This result was unexpected.) Because the chat was repeated before every round and the groups remained fixed throughout the session it set up a dynamic in which there may have been some accountability between chat partners. For example, one might ask the others whether they had voted in the previous round, and if so then for which candidate. The transcripts also showed that subjects would often urge others to vote in a certain way or would agree to vote in a certain way and sometimes would ask about whether/for whom others had voted in the previous round. Additionally, in spite of the differences in design and context, a few patterns were observed that had also appeared in the group identity literature, such as the emergence of subjects who persistently encouraged others to vote (analogous to Schram and van Winden's [9] producers of social pressure).

The rest of the paper proceeds as follows. Section 2 describes the theoretical model and results and the experimental design, which is influenced by the predictions generated by the theoretical model. Section 3 discusses the empirical results. Finally, Section 4 concludes.

## 2. Materials and Methods

**2.1. Model.** This model, in which voter turnout is a participation game (Palfrey and Rosenthal [13]), is an extension of the logit quantal response equilibrium (QRE) model used in Levine and Palfrey [4] and Goeree and Holt [3], incorporating the possibility of ethical voting along the lines of Morton and Tyran [5]. Whereas Levine and Palfrey [4] only allowed

TABLE 1: Payoffs (symmetric).

	Type 1	Type 2
Candidate X	$q$	$q - \theta$
Candidate Y	$q - \theta$	$q$

potential voters (also referred to simply as “voters,” unless a distinction is needed between actual and potential voters) to choose between voting and abstaining, here they have three choices: vote for Candidate X, vote for Candidate Y, and abstain. This better captures the reality of most elections, in which voters choose between two or more candidates in addition to the option of abstaining from voting. This extension of the model also allows for a range of payoff structures, including one which may induce ethical voting through the presence of a “selfish” candidate who offers a high payoff to one group of voters and a low payoff to the other and an “ethical” candidate who offers an equitable distribution of payoffs across groups of voters.

In this model, the voting rule is a simple plurality. Ties are broken fairly, with the winner being picked randomly (e.g., by a coin toss). Payoffs, the distribution of which can be thought of as the candidates’ economic platforms, are paid according to which candidate wins regardless of whether or for whom the participant voted. The payoff structure for each type is common knowledge to everyone regardless of type.

Potential voters are separated into two “types,” which determine the payoff that each one will receive depending on the winner of the election. Each candidate offers different payoffs to the two different voter types (Table 1). This table shows payoffs that are symmetric in both the payoff differences and the actual payoffs. This symmetry is not theoretically necessary—the main takeaway from this table is how the candidates offer different payments to each type—but matches the experimental design used in this research. With this payoff structure, voters of both types would face identical incentives to vote for the candidate offering their type the higher payoff.

Voting is modeled as being costly. This mirrors the costs associated with voting in real-world elections such as transportation costs, childcare, lost wages due to time off from work, or the opportunity cost of whatever someone would otherwise do. This cost is only incurred if someone votes. Costs are independent draws from the same uniform distribution, varying across both voters and elections, and both voter types face the same distribution of costs. In each election, each voter knows his or her own cost before deciding whether to vote but only knows the distribution from which others’ costs—and his or her own costs during future elections—are drawn.

Following Levine and Palfrey [4], a quasi-symmetric equilibrium in this model is a set of four turnout strategies  $(\tau_{1X}, \tau_{1Y}, \tau_{2X}, \tau_{2Y})$  specifying the probabilities that a member of type  $i$  will vote for candidate  $j$ , as a function of the voting cost. For a quasi-symmetric equilibrium, it is assumed that within each type everyone will use the same strategy—although it is possible for members of each type to vote for either candidate, in the Nash equilibrium any voting will be

for the candidate that offers the higher payoff to his or her type.

The aggregate voting probabilities for each type-candidate combination are

$$p_{ij}^* = \int_{-\infty}^{\infty} \tau_{ij}(c) f(c) dc = \int_{-\infty}^{c_{ij}^*} f(c) dc = F(c_{ij}^*) \quad (1)$$

for each combination of voter type  $i$  and candidate  $j$ .

There is a cutpoint cost level  $c_{ij}^*$  below which, in the Nash equilibrium, the probability of voting equals one and above which the probability of voting is zero. At the cutpoint, the voter is indifferent. These cutpoints are expressed as

$$c_{ij}^* = \frac{(V_{ij} - V_{im})}{2} \times (\Pr(\text{make tie})_{ij} + \Pr(\text{break tie})_{ij}), \quad (2)$$

where  $V_{ij}$  is the payoff received by a member of voter type  $i$  if candidate  $j$  wins the election, and where  $m$  indexes the other candidate. This is the difference between the payoff received and the payoff that would have been received if the other candidate had won (the payoff difference). In the Nash equilibrium nobody would vote for the candidate offering one’s type the lower payoff because the payoff difference would be negative, implying a negative cost cutpoint in which case no cost would be low enough to induce voting. These cutpoints are set to zero because costs cannot be negative.

The term  $(1/2) \times (\Pr(\text{make tie})_{ij} + \Pr(\text{break tie})_{ij})$  is the probability that a vote cast by a member of type  $i$  for candidate  $j$  will be pivotal (henceforth referred to as the “pivotal probability”). The one-half is there because, in the presence of two candidates, if someone’s vote *makes* a tie then there is a 50% chance that this voter’s chosen candidate will win (and that the vote will have been pivotal)—the winner of a tie is determined randomly. If someone’s vote *breaks* a tie then there is a 100% chance that his or her chosen candidate will win. However, because there was already a 50% chance that this candidate would have won in the event of a tie, there would only be a 50% chance that the vote that broke the tie changed that candidate’s outcome from losing to winning. Further details about the calculation of the pivotal probability can be found in Appendix A.

The expected payoff differences of voting for candidate  $j$  as a member of type  $i$ , net of costs, are expressed as

$$\pi_{k,ij} = (V_{ij} - V_{im}) \times \Pr(\text{pivotal})_{ij} - c_k, \quad (3)$$

where  $V_{ij}$  is the payoff received by a member of type  $i$  if candidate  $j$  wins the election and  $c_k$  is the voting cost faced by a given person in a given election.  $\pi_{k,ij}$  can be thought of as a component of  $\tilde{\pi}_{k,ij} = \pi_{k,ij} + \varepsilon_{k,ij}$ , where  $\varepsilon_{k,ij}$  is independent and identically distributed extreme value. This produces the logit model that is shown below.

Quantal response equilibrium (QRE) analysis, introduced in McKelvey and Palfrey [2], allows for “noise” in the decision-making process for voting. This process is often influenced by factors such as emotions, perception biases, voter error, unobserved individual heterogeneity, and other

sources of noise. The QRE introduces decision errors via a logit probabilistic choice rule according to which participants make their voting choices.

Under QRE, the best response functions are probabilistic rather than deterministic. Although “better” responses are more likely to be observed than “worse” responses, the “worse” responses are still observed because choice has a stochastic element. Thus, the assumption of perfect rationality is relaxed in favor of bounded rationality. As the amount of noise decreases—and the QRE approaches the Nash equilibrium—voters become increasingly likely to make choices that are consistent with the Nash equilibrium. In its most general form, the logit probability of choice “ $x$ ” takes the form  $P(x) = e^{\lambda\pi_x} / \sum_i e^{\lambda\pi_i}$ , where  $\pi_i$  is the expected payoff from making choice  $i$  and  $\lambda$  is the logit precision parameter. The denominator  $\sum_i e^{\lambda\pi_i}$  sums across all possible choices, summing to 1.

As the precision parameter  $\lambda$  increases, the level of noise decreases and the level of precision in decision making improves. As this happens, the ratio of “better” responses to “worse” responses improves. When  $\lambda$  is very large (very high precision/very low noise), the solution approaches the Nash equilibrium, with all or almost all decisions matching the Nash equilibrium outcome. (Strictly speaking, this convergence only holds for the principal branch of the QRE when there are multiple equilibria. However, for the model presented here the results were robust across a wide range of starting points.) As  $\lambda$  approaches zero (very low precision/very high noise), voting participation decisions approach randomness (50% if there are two choices, 33% if there are three choices, etc.) as the probabilities of “better” responses and “worse” responses become increasingly similar.

In the context of this model, a Nash equilibrium would correspond to a sharp drop in the response function at the cost cutpoint, where the probability of voting would equal one if facing a cost below  $c^*$  and would equal zero if facing a cost above  $c^*$ . For an intermediate level of noise, the cost cutpoint would affect the decision of whether or not to vote but the response function would not be as sharp because noise would lead to some instances of voting (or abstaining) even when the cost is above (or below) the cutpoint. A very low  $\lambda$  (very high noise level/very low level of precision) would correspond to a cutpoint that is effectively irrelevant, where the size of the cost would have very minimal or no effect on the decision of whether or not to vote. This can be seen in Figure 1.

The logit quantal response function, which can be thought of as the logit probability of a member of type  $i$  making voting choice  $j$  (where the choices are vote for  $X$ , vote for  $Y$ , and abstain) and facing cost  $k$ , is shown by

$$p_{k,\text{vote}ij} = \frac{e^{\lambda\pi_{k,ij}}}{e^{\lambda\pi_{k,iX}} + e^{\lambda\pi_{k,iY}} + e^{\lambda\pi_{k,ia}}} \quad (4)$$

This can be simplified, for a type  $i$  voter choosing  $j$  (where the choices are only vote for  $X$  or vote for  $Y$ ) and facing cost  $k$ , as

$$p_{k,ij} = \frac{e^{\lambda\pi_{k,ij}}}{1 + e^{\lambda\pi_{k,iX}} + e^{\lambda\pi_{k,iY}}} \quad (5)$$

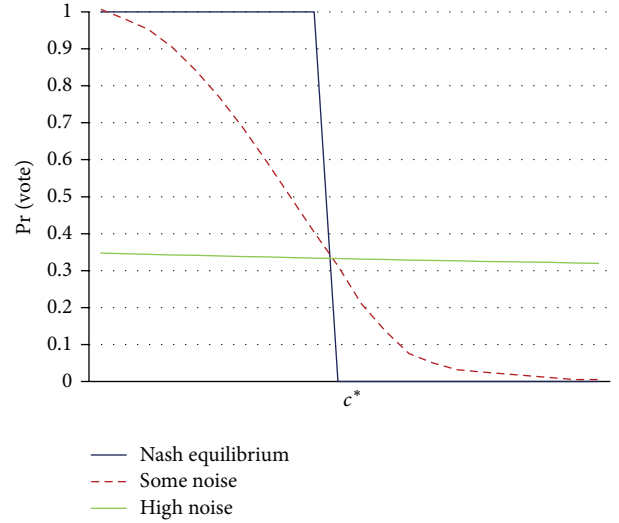


FIGURE 1: Logit response function relative to cost cutpoint  $c^*$ .

TABLE 2: Payoffs (asymmetric).

	Type 1	Type 2
Candidate X	$q + \theta$	$q - \theta$
Candidate Y	$q$	$q$

In the presence of random costs, these response functions have to be averaged across all possible costs (for a given type-candidate combination) in order to arrive at the average response probability of type  $i$  voting for candidate  $j$ . After averaging across costs, the probabilities are  $(p_{iX}, p_{iY}, 1 - p_{iX} - p_{iY})$ , where the third probability is the probability of abstaining.

The logit precision parameter,  $\lambda$ , is estimated using the likelihood function:

$$\begin{aligned} \log L = \sum_l (N_{1X,l} \times \ln(p_{1X,l}) + N_{1Y,l} \\ \times \ln(p_{1Y,l}) + (\text{num\_obs}_{1,l} - N_{1X,l} - N_{1Y,l}) \\ \times \ln(1 - p_{1X,l} - p_{1Y,l}) + N_{2X,l} \\ \times \ln(p_{2X,l}) + N_{2Y,l} \times \ln(p_{2Y,l}) \\ + (\text{num\_obs}_{2,l} - N_{2X,l} - N_{2Y,l}) \\ \times \ln(1 - p_{2X,l} - p_{2Y,l})), \end{aligned} \quad (6)$$

where  $N_{ij,l}$  is the total number of votes by members of type  $i$  for candidate  $j$  that were observed during treatment  $l$  (or session) and  $\text{num\_obs}_{i,l}$  is the total number of observations for someone of type  $i$  in treatment  $l$ . These observations include decisions for any of the three possible choices: vote for Candidate  $X$ , vote for Candidate  $Y$ , and abstain.

One of the contributions of this paper is the incorporation of ethical (altruism-motivated) voting into a QRE framework. Table 2 shows the payoffs for an ethical candidate and a selfish candidate (following the terminology used in the literature).

The selfish candidate offers a high payoff to one voter type and a low payoff to the other type. The ethical candidate offers payoffs that have a more equitable distribution across types. In this respect, the payoff structure follows Morton and Tyran [5]. In Table 2, the ethical candidate offers payoffs that are equal across types. Also, the difference between the high and low payoffs is the same for both types. Neither of these features—the ethical candidate offering exactly equal payoffs or setting payoffs so that the payoff difference is the same for both types—are necessary, but it matches the experimental design that was used in this research.

In Table 2,  $q$  and  $\theta$  stand for the baseline equitable payoff and the payoff difference, respectively. Candidate  $X$  is the selfish candidate and Candidate  $Y$  is the ethical candidate. Candidate  $Y$ 's payoffs maximize the minimum payoff and also minimize the difference between the payoffs received by the two types. Members of Type 1 receive a higher payoff if Candidate  $X$  wins but may be motivated by altruism to vote for Candidate  $Y$  even though they would receive a lower payoff if this candidate won the election. This is ethical voting. On the other hand, members of Type 2 receive a higher payoff if Candidate  $Y$  wins and would have no rational motive for voting for Candidate  $X$  because this candidate offers them a lower payoff and offers inequitable payoffs.

Selfish voting happens when economic self-interest is acted upon by voting for the candidate who offers one's type a higher payoff. This would happen when Type 1 votes for Candidate  $X$  or when Type 2 votes for Candidate  $Y$ . As discussed above, ethical voting happens when altruism outweighs economic self-interest and this would happen—at least potentially—when Type 1 votes for Candidate  $Y$ . It is also possible that this would occur because of noise. This is one example of how the chat feature proves useful for gaining insight into the motives behind voting decisions. There is no rational motive for Type 2 to vote for Candidate  $X$ ; therefore, this voting choice would be due to noise.

One difference between this design and Morton and Tyran's [5] is that in this design the aggregate payoffs are the same across candidates. I designed it this way so that, in the absence of ethical motives, equilibrium predictions would be the same across symmetric and asymmetric payoff structures for a given electorate size and a given noise level.

Ethical voting can be divided into two categories: ethical expressive and ethical instrumental. For ethical expressive voting, utility (or, in this model, implicit payoff) is gained from the act of voting for the ethical candidate regardless of the outcome of the election. This is along the lines of a warm glow parameter, as in Andreoni [14]. It is captured by the  $\alpha$  term in the net expected payoff difference equation below for someone of Type 1 voting for Candidate  $Y$  and facing a person- and election-specific cost  $k$ :

$$\pi_{1Y} = (V_{1Y} - V_{1X}) \times \Pr(\text{pivotal})_{1Y} - c_k + \alpha. \quad (7)$$

Because ethical expressive voting is an additive term, it is independent of the pivotal probability. Therefore, it would increase the turnout rate even as the pivotal probability decreases, mitigating or even possibly outweighing the effect of the declining pivotal probability. The other three net

expected payoff difference equations, for the other type-candidate combinations, remain the same as in (3).

For ethical instrumental voting, utility is gained if the ethical candidate wins. Therefore, the utility/payoff gain is tied to the probability of one's vote for that candidate being pivotal. This is captured in the  $\delta$  term in the equation below:

$$\pi_{1Y} = (V_{1Y} - V_{1X} + \delta) \times \Pr(\text{pivotal})_{1Y} - c_k. \quad (8)$$

In the presence of either ethical instrumental or ethical expressive voting, the estimation procedure for  $\lambda$  and  $\alpha$  or for  $\lambda$  and  $\delta$  is the same as before except that now two parameters are being estimated. In each case either  $\alpha$  or  $\delta$ , respectively, is identified through an exclusion restriction because the ethical voting parameter only appears in one of the four net expected payoff difference equations. Therefore,  $\lambda$  is pinned down by the other three net expected payoff difference equations and then  $\alpha$  or  $\delta$  is pinned down given  $\lambda$ .

In this model, it is not possible to simultaneously model both ethical expressive and ethical instrumental preferences. In that case, the net expected payoff difference equation would be

$$\pi_{1Y} = (V_{1Y} - V_{1X} + \delta) \times \Pr(\text{pivotal})_{1Y} - c_k + \alpha. \quad (9)$$

However, because  $\alpha = \delta \times \Pr(\text{pivotal})_{1Y}$  there are an infinite number of possible combinations and  $\alpha$  and  $\delta$  cannot be separately identified.

**2.2. Experimental Design.** These theoretical predictions were empirically tested through a series of laboratory experiments. All sessions were conducted in the Veconlab Experimental Economics Laboratory at the University of Virginia on undergraduate students, using Veconlab software. There were 180 subjects—109 male and 71 female—across a total of 12 sessions. Each subject could only participate in one session; therefore, all subjects were inexperienced at the start of their respective sessions. Additionally, each session utilized only one treatment, so each subject only participated in one treatment. The average payout per subject was \$35, including a \$6 payment for showing up, and individual payouts depended upon individual decisions as well as the collective voting outcome.

Each session consisted of 20 rounds, or elections. Each subject was randomly assigned to one of two voting types, Type 1 and Type 2, at the start of the session and remained at that same type throughout the entire session in order to avoid reciprocity effects. Within each session, equal numbers of subjects were assigned to both types.

There were three possible choices: vote for Candidate  $X$ , vote for Candidate  $Y$ , and abstain. Anyone from either voter type could vote for either candidate. In the experiment, elections were referred to as rounds and candidates were referred to as options, as in "Option  $X$ " and "Option  $Y$ ," in an attempt to keep the language as neutral as possible. Throughout this paper, "round" and "election" will be used interchangeably and the options will be referred to as "Candidate  $X$ " and "Candidate  $Y$ " except in instances in which the word choice is relevant to the discussion at hand.



As discussed in the previous section, whichever candidate garnered the most votes within a round was declared the winner, with ties decided with a virtual coin flip. After the conclusion of each round, all subjects were told which candidate had won and the total number of votes for each candidate (but not the breakdown of how many of those were cast by members of each type). Each candidate offered different payoffs to each voter type, and subjects were paid the amount that had been promised to their type by the candidate who won in that round, regardless of whether or for which candidate they had voted. At the end of the session, subjects were paid their cumulative earnings.

In each round, each subject drew a voting cost from the same distribution. Costs ranged from \$0 to 0.42, in increments of \$0.02. This increment was chosen so that there were 22 possible costs and 20 rounds, meaning that, on average, there would be a good chance of a subject drawing almost all of the possible costs over the course of the session.

The upper bound of the range of costs was chosen to generate at least some abstention. Along the lines of Levine and Palfrey [4], the payoff from making or breaking a tie would equal  $(V_{ij} - V_{im})/2$  for a member of type  $i$  voting for candidate  $j$ , where  $m$  is the other candidate. For both the asymmetric and symmetric treatments, the payoff from making or breaking a tie would equal \$0.25 for both types, according to the payoffs detailed below. Therefore, for subjects who drew a cost greater than this, abstention was the strictly dominant strategy. If there were a cost equal to \$0.25 then abstaining would be the weakly dominant strategy of a subject who drew this cost, but this was not a possible draw because of the \$0.02 increment size. The highest possible cost was set above the payoff from making or breaking a tie so that there would not exist an equilibrium in which the probability of voting was equal to one (i.e., in order to generate at least some abstention in equilibrium). Note that  $V_{ij}$  is assumed to be the higher payoff. As detailed earlier, there would be no cross-voting in the Nash equilibrium because the payoff differences would be negative, leading to a negative cost cutpoint.

At the end of the session, subjects participated in an exercise to elicit the subjects' perceived probabilities that their vote was/would have been pivotal in the final round, to test the relationship between voting turnout and the (perceived) probability of casting a pivotal vote. An analysis of the elicited probabilities showed a significant and positive relationship between the elicited probabilities and voting decisions—subjects who thought that they had a higher probability of being pivotal were more likely to have voted in the final round and with lower elicited probabilities were less likely to have voted. This section is available from the author upon request.

The subjects' computer screens showed short "backstories" next to each candidate's offered payoffs in the type-candidate payoff chart. The purpose of these was to help the subjects in the treatments with asymmetric payoffs realize the possibility of ethical voting. However, in order to avoid inadvertently encouraging or leading some subjects to vote in a certain way—as had happened in some of the pilot experiments—neutral wording was used here. The new backstories merely repeated the payoffs in words.

The backstories for asymmetric payoffs read as follows: "Option X will implement an investment that pays \$2 to Type 1 voters and \$1 to Type 2 voters" and "Option Y will implement an investment that pays \$1.50 to both Type 1 and Type 2 voters." For symmetric payoffs, the backstories were as follows: "Option X will implement an investment that pays \$1.50 to Type 1 voters and \$1 to Type 2 voters" and "Option Y will implement an investment that pays \$1 to Type 1 voters and \$1.50 to Type 2 voters."

Before the voting began in each round, subjects were able to chat online with their "neighbors" according to the ID number that had been assigned upon first logging into the program. For example, ID 3 could chat with ID 2 and ID 4. Meanwhile, ID 4 could chat with ID 3 and ID 5. The subjects with ID numbers at either end—ID 1 and ID 12 or ID 18, depending on the session size—were in the same chat group, creating a circular network. Subjects could also see the voter type of each of the other two subjects in their chat groups. The chat period lasted exactly one minute in each round, and it was not possible for anyone to submit a voting decision until after the chat ended.

This design models some features of real world discussions about politics. For example, most people only discuss an upcoming election with people whom they know directly (friends, family members, possibly coworkers, etc.). Even for those who broadcast their views on social media, the number of people who are on the receiving end of that is still small relative to the total size of the electorate. Given the small electorate size in this experiment, limiting a subject's chat circle to only two other people roughly mirrors this. Additionally, the overlapping group structure mirrors the fact that social groups are not self-contained, so that ideas or information can potentially be spread to a large number of people.

Furthermore, most of the chat groups contained members of both voting types. This was done to reflect the fact that most people have family members, friends, colleagues, acquaintances, and so forth who hold different political views than they do. (This differs from the focus of much of the group identity literature on the effects of communication within groups of voters.) Some subjects were part of a majority within their chat group and others were the minority, and in a few cases the subject was in a group consisting entirely of that subject's type.

Finally, the chat transcripts are extremely useful for providing insight into the subjects' motivations for voting in a certain way or for abstaining altogether.

The experiment was divided into four treatments, which were designed to test the effects of electorate size—and thereby the role of the probability of casting a pivotal vote—and the possibility of ethical voting. First, the number of subjects per session—which can also be thought of as the size of the electorate—was varied. Half (six) of the sessions had 12 subjects per session, and the remaining sessions had 18 subjects. According to the underlying theory, the turnout rate should fall as the electorate size increases because the probability of one's vote being pivotal decreases. Second, the payoffs were varied so that in half of the sessions the subjects faced symmetric payoffs and in the other half the payoffs were

TABLE 3: Symmetric and asymmetric payoffs.

	Asymmetric		Symmetric	
	Type 1	Type 2	Type 1	Type 2
Candidate X	\$2.00	\$1.00	\$1.50	\$1.00
Candidate Y	\$1.50	\$1.50	\$1.00	\$1.50

asymmetric. These payoffs, which were common knowledge within a session, can be seen in Table 3.

In the treatments with asymmetric payoffs, Candidate X was the selfish candidate, offering a high payoff (\$2.00) to Type 1 subjects and a low payoff (\$1.00) to Type 2 subjects. The ethical candidate, Candidate Y, offered equal payoffs to both voter types. These payoffs maximize the minimum payoff and minimize the difference between the payoffs received by the two groups—indeed, in this design members of both groups would receive the same payoff. The payoff of \$1.50 for each group that was received by subjects of both voter types if Candidate Y won was lower than what Type 1 subjects would receive and higher than what Type 2 subjects would receive if Candidate X won. This design was inspired by Morton and Tyran [5].

To account for the possibility that Candidate X being listed first on the table (or X coming before Y alphabetically) might have affected voting decisions, the roles of ethical and selfish candidates were reversed in roughly every other asymmetric session so that Candidate X offered the equal payoffs and Candidate Y offered the inequitable payoffs. Everything else remained the same. No meaningful difference was observed in the data. In the final data, X and Y were reversed back for these sessions to make the data consistent with the other sessions.

Under this payoff design, members of Type 1 may face conflicting incentives in choosing a candidate. Type 1 subjects would maximize their earnings if Candidate X won but might also decide to vote for Candidate Y for altruistic reasons.

In the symmetric payoff design, the two candidates offered payoffs that were mirror images of each other. This meant that members of both voter types had the same economic incentives to vote for the candidate who was offering one's type the higher payoff. Each candidate was equally inequitable, so there was no possibility of ethical voting here. These treatments served as a control, to measure the effect of noise in the decision-making process in the absence of any potential for ethical voting.

Each of the four treatments was a combination of session size (12 or 18) and payoff structure (asymmetric or symmetric).

The payoffs shown in Table 3 were set such that the equilibrium predictions of voting turnout rates were the same across asymmetric and symmetric payoff structures because the difference between the two candidates' offered payoffs is identical across the two payoff structures so that the net expected payoff difference in functions from (3) remains the same.

Meanwhile, the number of subjects per session would affect the turnout predictions through the effect of the probability of one's vote being pivotal on the net payoff function.

TABLE 4: (a) Predicted turnout probabilities and cost cutpoints (12 subjects). (b) Predicted turnout probabilities and cost cutpoints (18 subjects).

(a)			
	$\lambda = 100$ (Nash)	$\lambda = 7$ (L and P, 2007)	$\lambda = 0.2$ (high noise)
$p_{1X}$	0.2996	0.3069	0.3333
$p_{1Y}$	0	0.0793	0.324
$p_{2X}$	0	0.0793	0.324
$p_{2Y}$	0.2996	0.3069	0.3333
$c_{1X}$	0.122		
$c_{1Y}$	0		
$c_{2X}$	0		
$c_{2Y}$	0.122		

(b)			
	$\lambda = 100$ (Nash)	$\lambda = 7$ (L and P, 2007)	$\lambda = 0.2$ (high noise)
$p_{1X}$	0.2608	0.3069	0.3333
$p_{1Y}$	0	0.0793	0.324
$p_{2X}$	0	0.0793	0.324
$p_{2Y}$	0.2608	0.3069	0.3333
$c_{1X}$	0.122		
$c_{1Y}$	0		
$c_{2X}$	0		
$c_{2Y}$	0.122		

These both decrease (increase) as the electorate size increases (decreases). Therefore, the predicted turnout probabilities for a session with 18 subjects—regardless of whether it is a session with asymmetric or symmetric payoffs—are lower than those for sessions with 12 subjects, all else equal.

Tables 4(a) and 4(b) show the predicted turnout probabilities for the Nash equilibrium (extremely high  $\lambda$ ), for  $\lambda = 7$  (the value that was estimated in Levine and Palfrey, 2007), and for extremely high noise (very low  $\lambda$ ). The predictions for a session with 12 subjects are shown in (a), and the predictions for a session with 18 subjects are shown in (b).

In Tables 4(a) and 4(b),  $p_{ij}$  is the predicted probability that a member of type  $i$  would vote for candidate  $j$ . The probability of a member of type  $i$  abstaining is  $(1 - p_{iX} - p_{iY})$ .

As shown in Tables 4(a) and 4(b), the probability of voting is increasing in the amount of noise (decreasing in the amount of precision). This is true for both own-voting and cross-voting, where own-voting is defined as voting for the candidate offering one's type the higher payoff and cross-voting is defined as voting for the candidate offering one's type the lower payoff. However, it is much more pronounced for cross-voting. In the Nash equilibrium, there is no cross-voting because in the presence of extremely high levels of precision (low noise levels) in the decision-making process, and in the absence of altruism, nobody would choose to vote for the candidate offering them a lower payoff because it would imply a negative expected payoff difference and a negative cost cutpoint. As the level of noise increases/the level of precision falls, it becomes increasingly likely that at least some of the subjects would cross-vote in spite of the negative expected payoff difference.

TABLE 5: Predicted turnout probabilities for different levels of ethical preferences ( $\alpha$ ).

For $\lambda = 7$	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
Type 1 for Candidate X	0.3071	0.2708	0.1606	0.0295	0.0052	0.0009
Type 1 for Candidate Y	0.0793	0.1575	0.3988	0.8232	0.9637	0.9934
Type 2 for Candidate X	0.0793	0.0818	0.1059	0.1629	0.1702	0.1707
Type 2 for Candidate Y	0.3071	0.2921	0.2347	0.175	0.171	0.1708
Candidate X (% of all voters)	0.1932	0.1763	0.13325	0.0962	0.0877	0.0858056
Candidate Y (% of all voters)	0.1932	0.2248	0.31675	0.4991	0.56735	0.5821

For a session with 12 subjects (Table 4(a)) the cost cutpoint is \$0.122, which means that there are seven possible costs—between \$0.00 and \$0.12—that are low enough that someone would vote (for the candidate offering one's type the higher payoff) with probability of 1. Seven out of 22 possible costs equal approximately 30%, which is in line with the turnout probabilities in the Nash equilibrium. For a session of 18 subjects (Table 4(b)), the cost cutpoint is \$0.1047, which means that there are six possible costs—between \$0.00 and \$0.10—that are low enough to induce someone to vote. Six out of 22 possible costs equal approximately 27%, which is in line with the turnout probabilities in the Nash equilibrium.

One other thing to note is that as the noise level increases to the point where subjects are deciding randomly, the highest possible predicted turnout rate will be 33% (33% of each type would choose Candidate X, 33% would choose Candidate Y, and 33% would abstain). This is much lower than the 50% probabilities that would be the case if they could only choose between voting and abstaining, such as in Levine and Palfrey [4]. This will affect the ability of QRE to explain large amounts of overvoting, as will be discussed in the results section.

Additionally, the model predicts that increasing levels of ethical preferences can increase the percentage of the electorate voting for the ethical candidate and decrease the percentage voting for the selfish candidate (Table 5). (As in Table 4, the first four rows of Table 5 measure the percentage of that voters' type who vote for a given candidate. The last two rows, capturing the voting decisions across voter types, measure the percentage of *all* voters.) As the level of altruism increases, ethical voting would reduce the closeness of the race, which in turn would decrease the probability of casting a pivotal vote. This loss of pivotality would not matter for the ethical voters, whose increasingly strong altruistic motivation would outweigh this, but it would lead those who would have voted for the selfish candidate to abstain.

### 3. Results and Discussion

There are several themes that appear throughout the results. The first is that voter turnout far exceeded even the highest possible turnout predicted by QRE. This was largely concentrated in votes for the candidate offering the higher payoff to that subject's voter type and as such is not due to ethical voting. This indicates a strong economic self-interest. In addition, contrary to the predictions of the model—which were that, in the absence of altruism, the number of subjects (size of the electorate) would affect turnout but the symmetry of the payoff structure would not—the number of subjects

had no significant effect and turnout was significantly higher in sessions with symmetric payoffs relative to those with asymmetric payoffs. (However, the lack of a size effect is not entirely surprising considering that the theoretical differences in economically self-interested voting rates were not very large, as seen in Tables 4(a) and 4(b).) Third, some ethical voting (where motives were confirmed through the chat transcripts) and potentially ethical voting were observed; however, it was mostly found only under certain conditions and not enough was observed to explain the overvoting.

**3.1. Terminology.** Before getting into the analysis, some terminology that is used throughout this section needs to be defined. “Turnout” refers to all voting, regardless of type or candidate. “Favored” refers to votes for the candidate who offers that subject's type a higher payoff than what the other candidate offers: Type 1 for Candidate X and Type 2 for Candidate Y. Similarly, “unfavored” refers to votes against their economic self-interest—Type 1 for Candidate Y and Type 2 for Candidate X. The unfavored category does not distinguish between potentially ethical voting—Type 1 voting for Y in a session with asymmetric payoffs—and cross-voting that is done for other reasons.

#### 3.2. Voting Turnout Rates

**3.2.1. Overview.** The overall turnout rates, averaged across all rounds and broken down by type-candidate combinations, can be seen in Table 6(a).

The turnout rates for the candidates offering each type a higher payoff—Candidate X if Type 1 and Candidate Y if Type 2—are much higher than even the highest turnout rates predicted under QRE for any type-candidate combination, which topped out at 33% even in the presence of the highest possible levels of noise. Possible reasons for this overvoting, even relative to QRE predictions, will be addressed later. The second observation is that potentially ethical voting—Type 1 voting for Candidate Y in treatments with asymmetric payoffs—is higher than nonaltruistic cross-voting in these same treatments, which is Type 2 for Candidate X. In the treatments with symmetric payoffs, where there is no possibility of ethical voting, this is not the case.

There are two other notable relationships in the data. The first is that the turnout rate is higher, for all type-candidate combinations, in the treatments with symmetric payoffs than in those with asymmetric payoffs. This contradicts the model's prediction, which was that there would be no difference in turnout between the two payoff structures

TABLE 6: (a) Voting turnout rates (all rounds). (b) Voting turnout rates (rounds 11–20). (c) Voting turnout rates (rounds 1–10).

(a)					
	All treatments	Symmetric	Asymmetric	12	18
Type 1 for Candidate X	0.4911	0.5467	0.4356	0.4500	0.5185
Type 1 for Candidate Y	0.0639	0.0656	0.0622	0.0597	0.0667
Type 2 for Candidate X	0.0544	0.0733	0.0356	0.0653	0.0472
Type 2 for Candidate Y	0.5250	0.5533	0.4967	0.5417	0.5139
(b)					
	All treatments	Symmetric	Asymmetric	12	18
Type 1 for Candidate X	0.4622	0.5022	0.4222	0.3972	0.5056
Type 1 for Candidate Y	0.0344	0.0533	0.0156	0.0167	0.0463
Type 2 for Candidate X	0.0211	0.0200	0.0222	0.0222	0.0204
Type 2 for Candidate Y	0.4911	0.5533	0.4289	0.5250	0.4685
(c)					
	All treatments	Symmetric	Asymmetric	12	18
Type 1 for Candidate X	0.5200	0.5911	0.4489	0.5028	0.5315
Type 1 for Candidate Y	0.0933	0.0778	0.1089	0.1028	0.0870
Type 2 for Candidate X	0.0878	0.1267	0.0489	0.1083	0.0741
Type 2 for Candidate Y	0.5589	0.5533	0.5644	0.5583	0.5593

except for possibly higher turnout for Type 1 voting for Candidate in treatments with asymmetric payoffs (potentially ethical voting). Also, there is no systematic difference in turnout between the two session sizes. This contradicts the model's predictions, in which turnout would be higher in sessions of 12 than in sessions of 18.

The turnout rates in the second half of the data (rounds 11–20) can be seen in Table 6(b).

This focus on the second half of the sessions (largely) sidesteps the effect of the learning curve that the subjects experienced in the early rounds. Once again, the effects of size and symmetry run counter to the model's predictions. Turnout is higher in treatments with symmetric payoffs relative to those with asymmetric payoffs, and session size has no systematic effect.

In rounds 11–20 the turnout rate for potentially ethical voting—Type 1 voting for Candidate Y in treatments with asymmetric payoffs—is even lower than the nonaltruistic cross-voting done by Type 2 for Candidate X, again in the asymmetric treatments. This points to altruism dissipating and economic self-interest taking over fairly quickly. Given that any possible altruism died out by the second half, it would be useful to look at the first half of the data (rounds 1–10). This is shown in Table 6(c).

Here we see potential altruism, with a turnout probability of 0.1089 for Type 1 subjects voting for Candidate Y in asymmetric treatments versus 0.0489 for Type 2 subjects voting for Candidate X in those same treatments. Also, the size and symmetry relationships are the same as before, albeit not as strongly.

It is also helpful to see the trends over time, not just the averages. Figure 2 shows the turnout rate—overall turnout, rather than being broken down into the different type-candidate combinations—over the course of the experiment.

Within Figure 2, the top row shows the turnout rates for 12 subjects (a) and for 18 subjects (b), and the bottom row shows the turnout rates for asymmetric payoffs (c) and symmetric payoffs (d).

A few trends appear in these figures. First, in all treatments the turnout rate decreases as the rounds go on. Second, there is no systematic difference between the 12- and 18-subject turnout rates. This is contrary to the theoretical results, which showed higher turnout for smaller electorate sizes. And finally, there is a clear difference between the treatments with asymmetric versus symmetric payoffs. As before, turnout in treatments with symmetric payoffs is almost always higher. This, too, is contrary to the theoretical results, which were identical across payoff structures for a given electorate size.

**3.2.2. Nonparametric Permutation Test.** A nonparametric permutation test (also known as a randomization test) was used to determine whether the turnout rates are different across treatments. This nonparametric test was useful when looking at session averages because there are only 12 data points. The results from the permutation tests showed the same trends that were observed in the stylized facts: the number of subjects per session does not significantly affect the probability of voting (except for symmetric sessions across all rounds), and subjects in treatments with symmetric payoffs are significantly more likely to vote than are subjects in treatments with asymmetric payoffs.

Details about the test and the results can be found in the Appendix.

**3.2.3. Econometric Analysis.** In the logit analysis, the dependent variables are the different categories of voter turnout—("turnout"), favored, and unfavored. In each of the following

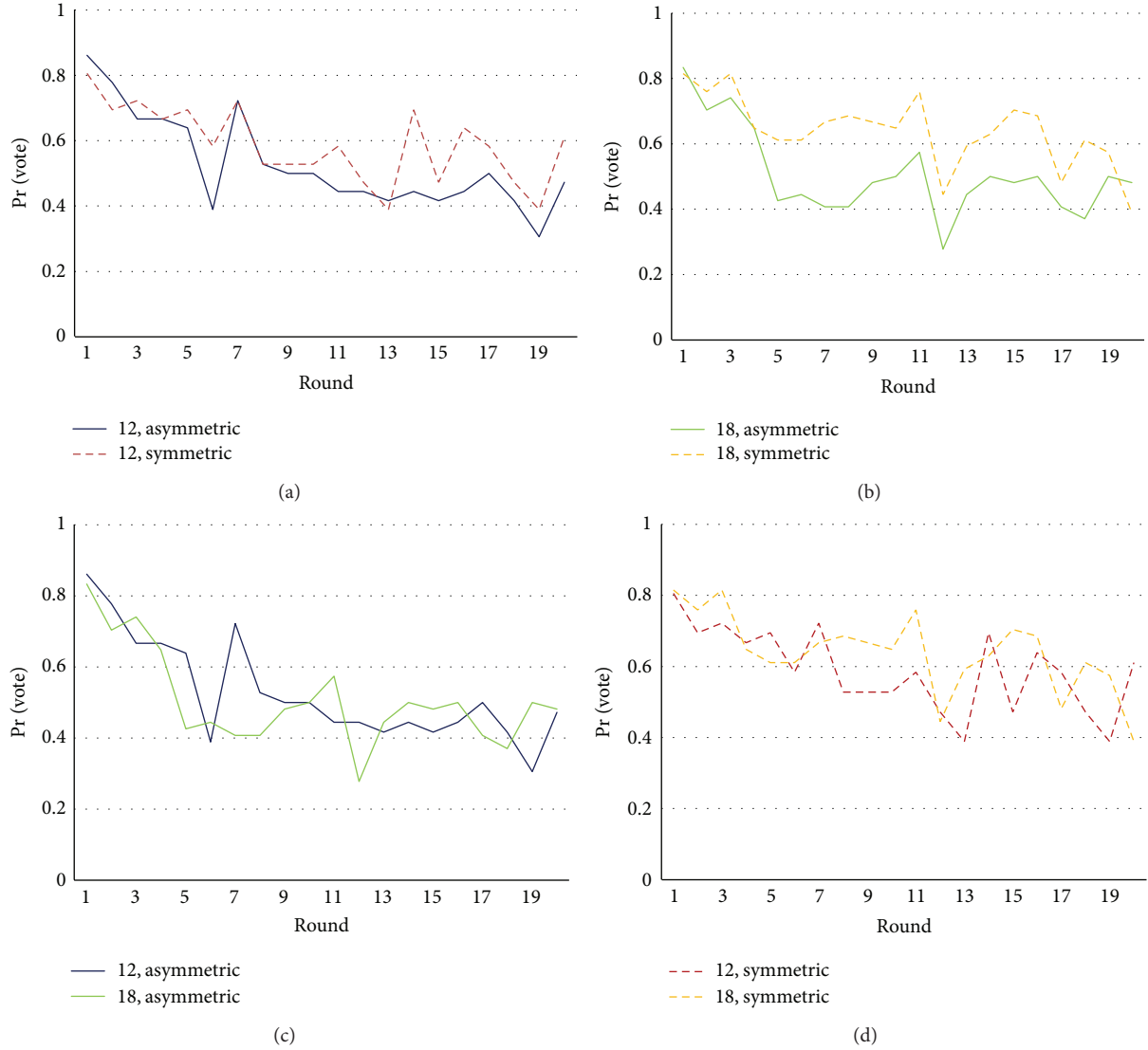


FIGURE 2: Voting turnout rates (top row symmetric versus asymmetric, bottom row 12 versus 18).

regressions, a random effects term was used to account for correlation within each subject's 20 decisions that cannot be attributed to any of the other independent variables—some subjects were inherently more (or less) likely to vote, or to vote in a certain way, than other subjects.

Table 7 shows the results for the following regressions:

$$\begin{aligned} \text{turnout}_{it} = & \beta_0 + \beta_1 (\text{average vote difference})_{it} \\ & + \beta_2 (\# \text{ of previous losses})_{it} + \beta_3 \text{cost}_{it} \\ & + \beta_4 \text{type}_{it} + \beta_5 \text{size}_{it} + \beta_6 \text{asymmetric}_{it} \\ & + \beta_7 \text{round}_{it} + \beta_8 (\text{type 1} \times \text{asymmetric})_{it} \\ & + \beta_9 (\# \text{ same type})_{it} \\ & + \beta_{10} (\# \text{ group favored votes})_{it} \\ & + \beta_{11} (\# \text{ group unfavored votes})_{it} \\ & + w_{\text{turnout},i} + u_{\text{turnout},it}, \end{aligned}$$

$$\begin{aligned} \text{favored}_{it} = & \gamma_0 + \gamma_1 (\text{average vote difference})_{it} \\ & + \gamma_2 (\# \text{ of previous losses})_{it} + \gamma_3 \text{cost}_{it} \\ & + \gamma_4 \text{type}_{it} + \gamma_5 \text{size}_{it} + \gamma_6 \text{asymmetric}_{it} \\ & + \gamma_7 \text{round}_{it} + \gamma_8 (\text{type 1} \times \text{asymmetric})_{it} \\ & + \gamma_9 (\# \text{ same type})_{it} \\ & + \gamma_{10} (\# \text{ group favored votes})_{it} \\ & + \gamma_{11} (\# \text{ group unfavored votes})_{it} \\ & + w_{\text{favored},i} + u_{\text{favored},it}, \end{aligned}$$

$$\begin{aligned} \text{unfavored}_{it} = & \theta_0 + \theta_1 (\text{average vote difference})_{it} \\ & + \theta_2 (\# \text{ of previous losses})_{it} + \theta_3 \text{cost}_{it} \\ & + \theta_4 \text{type}_{it} + \theta_5 \text{size}_{it} + \theta_6 \text{asymmetric}_{it} \end{aligned}$$



TABLE 7: Logit with subject-level random effects and cluster-robust standard errors (marginal effects).

	All rounds			Rounds 11–20			Rounds 1–10
	Turnout	Favored	Unfavored	Turnout	Favored	Unfavored	Unfavored
History: average $ X - Y $	-0.367*** (0.10)	-0.508*** (0.10)	0.037 (0.31)	-0.415* (0.23)	-0.435* (0.23)	-2.863* (1.36)	0.462 (0.34)
Number consecutive losses	-0.008** (0.00)	-0.012*** (0.00)	0.001*** (0.00)	-0.004 (0.00)	-0.009 (0.01)	0.000** (0.00)	-0.001 (0.00)
Type 1	-0.013 (0.06)	-0.002 (0.05)	-0.004 (0.01)	0.011 (0.06)	-0.017 (0.06)	0.000 (0.00)	-0.011 (0.01)
Cost	-1.554*** (0.13)	-1.620*** (0.12)	-0.343* (0.19)	-2.248*** (0.22)	-2.210*** (0.21)	-0.735** (0.36)	-0.254 (0.24)
Asymmetric	-0.129** (0.06)	-0.094* (0.06)	-0.020** (0.01)	-0.144** (0.06)	-0.140** (0.06)	-0.001 (0.00)	-0.031** (0.02)
Size 12	-0.022 (0.04)	-0.031 (0.03)	0.00 (0.01)	-0.04 (0.04)	-0.026 (0.05)	0.000 (0.00)	0.002 (0.01)
Round	-0.009*** (0.00)	-0.001* (0.00)	-0.002*** (0.00)	-0.005 (0.00)	-0.002 (0.00)	-0.000*** (0.00)	-0.004*** (0.00)
Type 1 * asymmetric	-0.023 (0.08)	-0.064 (0.08)	0.016 (0.01)	-0.034 (0.09)	-0.005 (0.09)	0.000 (0.00)	0.035* (0.02)
Number same type	-0.045 (0.04)	-0.032 (0.04)	-0.002 (0.01)	-0.001 (0.04)	0.025 (0.04)	-0.001 (0.00)	-0.002 (0.01)
Number favored	0.118*** (0.02)	0.163*** (0.02)	-0.013*** (0.01)	0.086*** (0.03)	0.095*** (0.03)	-0.000* (0.00)	-0.025*** (0.01)
Number unfavored	0.025 (0.02)	-0.019 (0.01)	0.005*** (0.00)	0.004 (0.02)	0.00 (0.02)	0.000 (0.00)	0.010** (0.01)
LR test ( $p$ value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

Marginal effects elasticities used for continuous variables: history: average  $|X - Y|$ , cost. All others are marginal effects.

$$\begin{aligned}
& + \theta_7 \text{round}_{it} + \theta_8 (\text{type } 1 \times \text{asymmetric})_{it} \\
& + \theta_9 (\# \text{ same type})_{it} \\
& + \theta_{10} (\# \text{ group favored votes})_{it} \\
& + \theta_{11} (\# \text{ group unfavored votes})_{it} \\
& + w_{\text{unfavored},i} + u_{\text{unfavored},it} \cdot
\end{aligned} \tag{10}$$

The last line in Table 7 is the likelihood-ratio test of whether the residual intraclass correlation equals zero (not a marginal effect, even though it was consolidated into this table). This tests whether the random effects term should be omitted. The results show that the use of random effects was appropriate.

Table 7 shows the marginal effects, where marginal effect elasticities were used for continuous variables and the usual marginal effects were used for binary or count variables. This is because “by what percentage does (dependent variable) increase if Asymmetric increases by 10%” is not meaningful, given that Asymmetric is binary. Note that the scales are different, so that an elasticity of 0.34 and a marginal effect of 0.034 (using an arbitrarily chosen example) would both be interpreted as a 0.34% increase in the dependent variable. This is why the results for “History: average  $|X - Y|$ ” and “Cost” look so much larger than the others.

The independent variables are as follows. “History: average  $|X - Y|$ ” measures the average, up until any given round, of the absolute value of the differences between the number of votes received by each candidate in previous rounds. This captures how close the outcomes of the elections had been and might have played a role in subjects’ perception of the probability of their vote being pivotal in the current round. The second variable, “number consecutive losses,” is the number of consecutive rounds, immediately preceding the round in question, that a subject’s type’s favored candidate (offering that type the higher payoff) has lost. This captures discouragement. “Type 1” is the voter type and equals one if Type 1 and zero if Type 2. “Cost” is the voting cost, between \$0 and \$0.42. “Asymmetric” equals one if the observation is from a session with asymmetric payoffs and zero if it is from a session with symmetric payoffs. “Size 12” equals one if the observation is from a session with 12 subjects and equals zero if it is from a session with 18 subjects. “Round” is the round number, between 1 and 20. “Type 1 \* asymmetric” is an interaction term for a Type 1 subject in a session with asymmetric payoffs—the only combination for which ethical voting (Type 1 for Candidate Y) is possible—and equals one if the subject fits that description and zero otherwise. “Number same type” is the number of members of a subject’s chat group (excluding the subject) who are members of the same voter type as the subject. “Number favored” and

“number unfavored” are the number of members of a subject’s chat group (again, excluding the subject) who voted for the subject’s favored or unfavored candidate, respectively, in the current round. I also tested this using versions of “number favored” and “number unfavored” that use a rolling average of the previous 3 rounds, and the results were qualitatively similar in both magnitude and significance. Therefore, I decided to use the version above to focus on the more immediate effect of the most recent chat.

This analysis was done for all rounds and for the second half (rounds 11–20) for overall turnout, favored and unfavored, and for the first half (rounds 1–10) for unfavored only. It is uncommon to focus on the first half of the experimental data, because subjects often experience a learning curve during the first several rounds; however, in this case all of the ethical (motives confirmed through the chat transcripts) and potentially ethical voting occurred in the first half. This occurs in unfavored voting; therefore, this column is included in Table 7.

History: average  $|X - Y|$ , which measures the average number of votes separating the two candidates in previous rounds (the average margin), is significant and negative for both overall turnout and favored turnout, both for all rounds and for rounds 11–20. This means that the farther apart the vote counts had been in the previous rounds the less likely subjects were to vote, or conversely the closer the previous rounds had been the more likely subjects were to vote. Looking at the results for all rounds, if the vote counts in the previous elections were 10% farther apart the turnout rate would fall by 3.67% overall (or conversely, if previous elections were 10% closer the turnout rate would increase by 3.67%) and 5.08% for favored. The results in the second half were qualitatively similar, albeit less strongly significant possibly due to the average becoming less volatile as rounds go on. To make sure that this was not simply because the average incorporates more rounds as the session goes on, so that each incremental round has a smaller impact on the average, I also tried a rolling average of the previous three rounds and got qualitatively similar results.

The negative relationship between the size of the margin of victory in previous elections and the probability of voting in the current election (or the positive relationship between the closeness of previous elections and the probability of voting) may be because of the effect on subjects’ perceptions of the probability of their vote being pivotal. This result, and interpretation, qualitatively matches the results of the belief elicitation from the extended version of this paper (available upon request), which found that subjects’ *ex ante* beliefs about the probability of their vote being pivotal were positively related to the probability of voting. This, combined with the lack of size effect in the data, points to a possibility that voters’ *perceptions* of the probability could be the channel through which the probability of casting a pivotal vote impacts voting participation decisions, rather than the *actual* probability.

The number of consecutive losses immediately preceding a round is significant and negative for both overall turnout and favored turnout; the results show that with each additional consecutive loss the probability of voting falls by 0.9%

and the probability of voting for favored falls by 1.2%. For unfavored, the number of consecutive losses is significant but *positive*. It is not entirely clear why but may reflect the fact that as the number of previous losses adds up, the subjects who have a better understanding of the game may be more likely to abstain, leaving more voters who have a worse understanding and who would be more likely to make voting choices that go against their economic self-interest. This relationship weakens when focusing on only the second half because, in most sessions, losing/winning streaks were shorter in the second half.

The next significant variable is cost, which is highly significant and negative for all three dependent variables. For a 10% increase in voting cost, the probability of voting for either candidate falls by 15.54%, the probability of voting for favored falls by 16.20%, and the probability of voting for unfavored falls by 3.43%. The effect of cost on unfavored voting is both smaller in magnitude and in significance level, possibly because noneconomic factors would play a larger role in unfavored voting decisions and therefore the voting cost would have less impact. The cost effect is even stronger when focusing on only the second half, as subjects have gained a better sense of when it is or is not worth it to vote. In the first half, cost does not significantly affect unfavored voting, possibly because of the ethical (and potentially ethical) voting and the higher level of noise that occurred in the first half, both of which would have made subjects less sensitive to costs.

The results for size and symmetry match those from the permutation test and Figure 2, in that the probability of voting is lower in sessions with asymmetric payoffs relative to sessions with symmetric payoffs—for all rounds, subjects in sessions with asymmetric payoffs were 12.9% less likely to vote at all, 9.4% less likely to vote for favored, and 2.0% less likely to vote for unfavored—and the number of subjects per session is not significant. What could explain the higher turnout probability in sessions with symmetric payoffs, relative to those with asymmetric payoffs? There is no clear answer.

The results in Table 5 show that the model predicts increases in both unfavored voting and overall turnout as the level of altruism increases, but the relationship is negative in the empirical results. However, the drop in favored voting in Table 5 as altruism increases does match the direction of the empirical results. It is possible that, in the absence of any possibility for ethical voting, the subjects may have perceived the elections as being closer, which spurred greater amounts of voting. In the data, the elections were actually closer in sessions with asymmetric payoffs, but work with belief elicitation could potentially uncover a connection to the subjects’ *perceptions* of election closeness rather than the actual margins.

The interaction term of Type 1 and asymmetric captures the voters who have the option of voting ethically. The results here show that potentially ethical voting—Type 1 subjects voting for Candidate Y (unfavored) in sessions with asymmetric payoffs—is 3.5% more likely to occur than nonethical unfavored voting during the first half, which is when almost all of the ethical voting (both potentially ethical and transcript-confirmed ethically motivated) took

TABLE 8: Maximum likelihood estimates.

Parameters	Rounds	Sessions or treatments	Coefficient	Standard errors	<i>t</i> -statistics
$\lambda$	2nd half	All treatments	5.4909	0.5164	10.6332
	1st half	All treatments	1.0853	0.2841	3.8205
	1st half	Asymmetric treatments	1.732	0.4339	3.9704
	1st half	Symmetric treatments	0.5687	0.3767	1.5098
$\lambda$ and $\alpha$	1st half	Session 1, session 7			
		$\lambda$	0.5448	0.8978	0.6068
		$\alpha$	0.129	0.3992	0.323
$\lambda$ and $\delta$	1st half	Session 1, session 7			
		$\lambda$	0.6155	0.8709	0.7068
		$\delta$	0.9748	1.9373	0.5032

place. Note that all unfavored voting in sessions with symmetric payoffs and unfavored voting by Type 2 subjects in sessions with asymmetric payoffs are not even potentially ethical.

It is possible that ethical voting could have had a larger impact if the asymmetric payoff design featured higher aggregate payoffs. Fowler and Kam [15] found that altruists only had a larger incentive to participate, relative to those who are self-interested, when outcomes are perceived as benefitting everyone. Outcomes that are perceived as being merely distributive, without a larger aggregate benefit, are seen as being distributive and altruists gain nothing from merely shifting wealth from one group to another. In the experimental design used here, the aggregate payoffs are identical across candidates so the ethical candidate could be perceived as being distributive.

Round, which captures learning effects, is significant and negative for all three dependent variables when looking at all rounds and also for unfavored in rounds 1–10. As a session progresses the probability of voting falls by 0.9% (for overall turnout), 0.1% (for favored), or 0.2% (for unfavored) with each additional round. This points to the presence of a learning effect as the session proceeds, possibly as subjects figure out that they are less likely to be pivotal than they had anticipated or that it does not make sense to vote when facing a high cost. Then the size and the strength of the significance decline by rounds 11–20, indicating that the learning effect had dissipated because the steepest part of the learning curve happened early in the session.

The next three variables analyze the effect of subjects' chat groups on voting participation decisions. The group affiliation of the members of a subject's chat group did not significantly affect voting decisions, but whether/for whom their chat-mates voted did influence voting decisions. Looking at the results for all rounds, "number favored" increased favored voting by 16.3%, showing that economically self-interested voting participation was substantially increased by the influence of group members' voting decisions to vote for that same candidate. Unfavored voting *decreased* as the number of group members voting for a candidate's favored candidate increased, as voters who might have been leaning towards to vote against type were encouraged to vote for their favored candidate instead.

Unfavored voting *increased* in the number of group members who voted for that candidates' unfavored candidate, possibly due to factors such as group members encouraging each other to vote for the ethical candidate (where applicable), voting in solidarity with group members regardless of whether it is in the subject's best interests, and manipulation. Instances of these, as observed in the chat transcripts, can be found in Section 3.4. However, this 0.5% increase (or 1% in the first half) is much smaller than the effect of "number favored" on favored voting, possibly because it is an easier "sell" to convince people to vote for someone who is promising them more money than the other candidate.

**3.3. Estimating the QRE and Ethical Voting Parameters.** The model that was presented in Section 2 can be used to estimate the QRE precision parameter  $\lambda$  and the ethical voting parameters  $\alpha$  and  $\delta$  from the experimental data. To review,  $\lambda$  is a measure of the level of precision in the decision-making process of whether and for whom to vote. The ethical expressive parameter,  $\alpha$ , measures the utility (or nonpecuniary expected payoff) received from the act of voting ethically. This is along the lines of a warm glow parameter, as in Andreoni [14]. The ethical instrumental parameter,  $\delta$ , measures the utility (or nonpecuniary expected payoff) received from voting ethically if that vote is pivotal and leads to a victory for the ethical candidate.

Table 8 shows the results of the estimation using the experimental data.

First,  $\lambda$  was estimated for the second half (rounds 11–20). Ethical voting had dissipated by then, so limiting it to these rounds made it possible to estimate the QRE precision parameter in the (effective) absence of ethical voting. The estimated value for  $\lambda$  was 5.49, which is lower than what Levine and Palfrey [4] had estimated ( $\lambda = 7$ ), meaning that there was more noise/less precision in the decision-making process here relative to Levine and Palfrey [4]. In that paper, subjects only had a choice between voting and abstaining so the higher noise/lower precision may be due to the addition of a third option (here, the possibility of choosing either candidate).

On a related note, it may also reflect the fact that voting turnout rates for Type 1 voting for Candidate X and Type 2 voting for Candidate Y are substantially above the highest

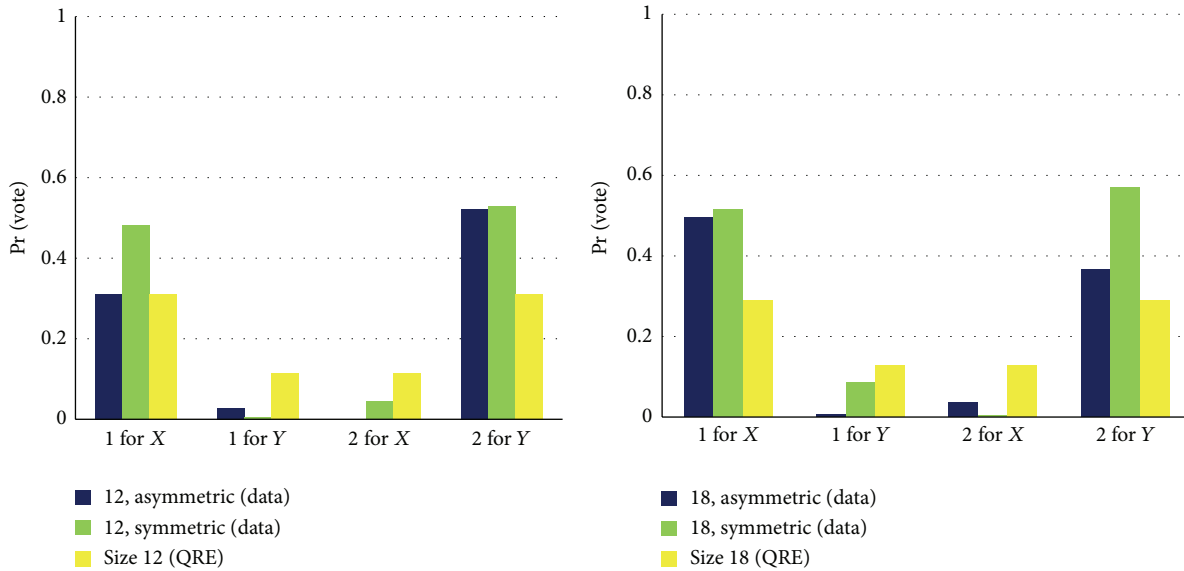


FIGURE 3: QRE versus data, rounds 11–20.

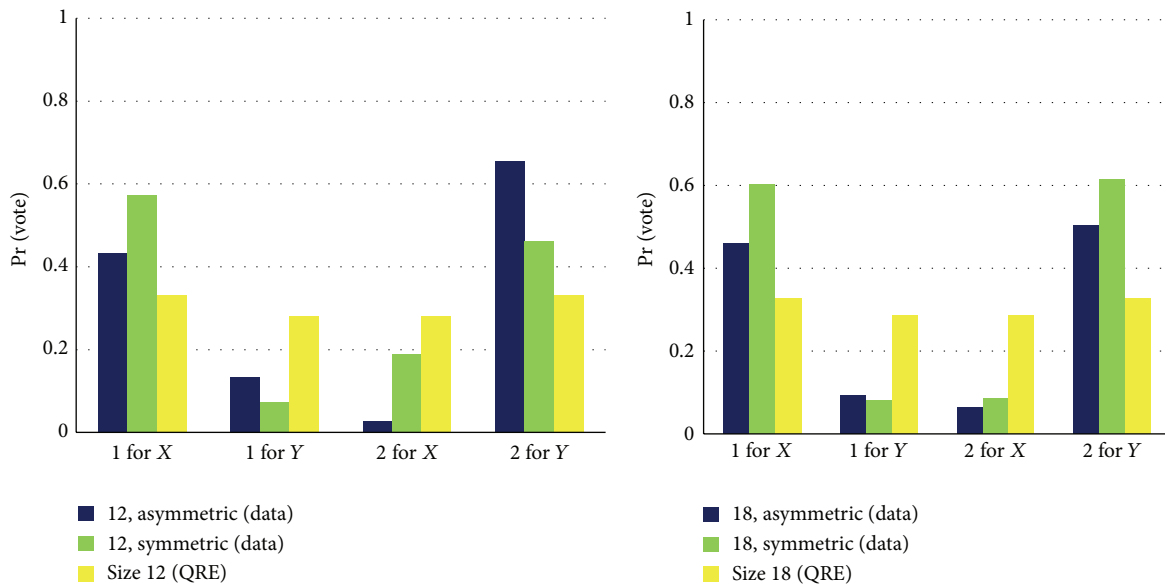


FIGURE 4: QRE versus data, rounds 1–10.

possible predicted turnout under QRE, which was 0.33 in the presence of extremely high noise as shown in Tables 4(a) and 4(b). This is much lower than the 0.5 predicted under high noise in Levine and Palfrey's model (with only two options, when noise is so high/precision is so low that decisions are made randomly the probability of each option approaches 50%), leading to a smaller range within which overvoting could be explained by noise.

Figure 3 illustrates this overvoting, contrasting the QRE turnout predictions for 12 subjects and 18 subjects, given a  $\lambda$  of 5.49, with the observed turnout rates. For favored voting (Type 1 for X and Type 2 for Y), the observed turnout rates are much higher than the predicted rates, and the observed

turnout rates for unfavored voting (cross-voting) are lower than the predicted rates.

Next, as a comparison,  $\lambda$  was estimated for the first half of the data. As expected, the estimate of  $\lambda$  for the first half is considerably smaller—indicating lower levels of precision/higher levels of noise—than those in the second half. This is because of both higher turnout rates and greater levels of cross-voting. A comparison of observed turnout rates and QRE predictions is illustrated in Figure 4. Surprisingly, the estimated precision parameter is lower (higher noise) for sessions with symmetric payoffs, in spite of lacking any possibility of ethical voting. However, looking at the data, there are comparable levels of cross-voting, and the turnout rates of Type 1 for Candidate

$X$  and Type 2 for Candidate  $Y$  are generally higher in the symmetric treatments. This higher turnout in favored voting is what is driving the difference. This is because the estimated  $\lambda$  is pulled lower by the high favored voting rates above and beyond what the model can explain, which is treated as extra noise.

However, the end goal is not only to estimate the QRE precision parameter  $\lambda$  but also to estimate the ethical voting parameters. This is done in the final two sections of Table 8. (As explained earlier, the ethical voting parameters  $\alpha$  and  $\delta$  cannot be separately identified so these are the results for separately estimated models, one with  $\alpha$  and one with  $\delta$ .) Unfortunately, for either of these specifications neither  $\lambda$  nor the respective ethical voting parameter is significant, even when only focusing on the two sessions with the highest rates of overvoting.

**3.4. Chat.** The online chat feature provides insight into the motivations behind subjects' voting decisions. (The chat transcript for all sessions is available upon request.) Two noteworthy themes appear in the chat: the first is the power of the overlapping chat circles to affect outcomes in both useful and harmful ways and the second is peer pressure/peer influence, even to the point of outright manipulation.

**3.4.1. Transmission of Information.** Although at least a small amount of potentially ethical voting was observed in most sessions with asymmetric payoffs, it appeared in a substantial fashion in Sessions 1 and 7. What differentiated these sessions from the other sessions with asymmetric payoffs was that in each case one of the subjects picked up on the possibility of ethical voting and voted accordingly and then broadcasted this realization and acted as a cheerleader of sorts to try to get Type 1 subjects to vote for Candidate  $Y$ . The "cheerleaders" were both Type 1 and Type 2, although their motivations differed according to type. These cheerleaders could be thought of as being thematically related to the producers of social pressure in Schram and van Winden [9]. In Session 1, ID 7—who was Type 1—started this chain of events in the chat before round 1. ID 7 picked up on the idea of ethical voting, "DO NOT BE GREEDY. WE WILL ALL MAKE MORE MONEY IF WE ALL CHOOSE THE LOWER OPTION." (All quotes appear exactly as in the chat transcript.) Strictly speaking, this was not entirely true because the aggregate payoff was the same regardless of which candidate won, and indeed Type 1 made *less* money when Candidate  $X$  won. However, the comment still demonstrates a desire to spread the payoffs more equitably. This subject then repeated this in the chat before round 2. Also in round 2, ID 8—who was a Type 2 and therefore was acting in his or her own economic self-interest—took up ID 7's argument by saying "Type 1 is being greedy," which spread the idea to ID 9—who was also a Type 2—who said, "Please choose option  $Y$ , we receive equal amounts regardless of type." Yet another Type 2, ID 11, said, "vote for option  $y$ . Otherwise Type 2 gets shafted." ID 10 was a Type 1 subject chatting with IDs 9 and 11 and started the chat period in round 2 by stating "Always vote for option  $X$  please. Same deal. We can get our money" but then changed his or her mind and ended up voting for Candidate  $Y$  after

being convinced by IDs 9 and 11. Of course, the fact that ID 10 said, "please" in the chat may indicate that this person was inherently more considerate and therefore maybe more likely to be open to the possibility of voting ethically, as opposed to someone like ID 5 whose sentiment was "sucks to be a 2."

Other rounds followed in a similar fashion. The amount of ethical voting never exceeded 2 subjects per round, and it was always done by IDs 7 and 10—and then only ID 7 after ID 10 started abstaining, but with only 12 subjects that was enough to make or break a tie in a few rounds. The chat also included some amusing attempts to generate ethical voting, such as "a vote for  $X$  is a vote for the bigwig corporations taking our hard-earned tax dollars," "a vote for  $X$  is un-American," and "here is a list of people who would vote for  $x$  in this scenario: bane, john lee hooker, ivan the terrible, judas."

In Session 7 a similar phenomenon occurred. (In Session 7, the selfish and ethical candidates had been switched so that Candidate  $X$  was the ethical candidate and Candidate  $Y$  was the selfish candidate. In this discussion the standard candidate name will be used, surrounded by brackets, rather than what was written in the transcript.) In round 3, ID 5 (Type 2) joked about the inequity of Candidate  $[X]$ 's offered payoffs. After this, ID 6 (Type 1) suggested voting for Candidate  $[Y]$  and ID 7 (Type 1) agreed. This block of voting continued for several more sessions. Also in round 3, ID 15 (Type 2) urged Type 1 subjects to vote for Candidate  $[Y]$  for ethical reasons, saying "but  $[y]$  makes money for everyone." (As explained earlier, this is not entirely true due to equal-sized aggregate payoffs but it demonstrates a desire for a more equitable distribution of payoffs.) However, it did not spread beyond this subject's original chat group.

One difference between Sessions 1 and 7 was the presence in Session 7 of voting that appeared to be ethical but was actually due to subjects not understanding how the payoffs and costs worked. This was the case for IDs 11 and 12, two Type 1 voters who banded together to vote for Candidate  $[Y]$  because they thought that it was to their advantage based on how high or low their costs were. Therefore, some of the seemingly ethical voting was actually due to noise or voter error. This is an example of how voting that appears ethical may not be and illustrates the usefulness of this chat feature for providing insight into motives.

It is also possible for incorrect information to spread and to affect the outcome of an election. In Session 12, ID 3—a Type 2 candidate—started off in the chat before round 1 by saying that everyone should vote for Candidate  $X$ . The chat period ended before he could follow up and explain why, but in round 2 the subject said, "continue with option  $X$  and forward it to your adjacent ID groups. If everyone votes  $X$  throughout, we all win." This strategy worked and by round 3 there were 11 votes for Candidate  $X$ , 1 vote for Candidate  $Y$ , and no abstentions.

The logic behind ID 3's strategy is a mystery because this was a session with symmetric payoffs so neither of the candidates was offering equitable payoffs. Also, as a member of Type 2 this subject received a *lower* payoff if Candidate  $X$  won. In round 5, this same subject wrote, "continue with  $x$  regardless of Type 1 or Type 2, it will all work out to



a guaranteed payout. fwd it to your group.” This subject had very quickly understood the power of the overlapping chat groups and yet apparently had not understood the instructions at the beginning of the session.

At the same time, word was spreading that subjects would not receive a payoff if they abstained or voted for the losing candidate. However, there were also individual conversations in which subjects were testing out the mechanics of the game. For example, in round 5 ID 8 had abstained and ID 9 had voted for Candidate Y, so in the chat before round 6 their discussion focused on whether payoffs were still received even if a subject abstained from voting. ID 8 wrote, “Yes. I did (abstain and still receive the payoff). It sucks that Option Y cannot win. But the better option now is to not vote if you’re Type 2.” ID 1 showed similar insight into this and also into the possibility of free riding in the presence of large margins for the winning candidate, “should we not vote this round? If the pattern stays the same, ppl will vote x and we won’t get charged.” Meanwhile ID 3 continued his or her push for Candidate X.

In the next round, there was chatter demonstrating continued confusion on the part of some subjects and among other subjects a growing understanding of the correct mechanics of the experiment. In this round five subjects ended up abstaining, indicating that the “always vote” strategy was breaking up. However, everyone who did vote chose Candidate X regardless of type. Evidently some were still confused about the link between payoffs and voting decisions, because in the chat before round 8 ID 5 said, “why would 5 people not vote and miss out on this cash cow?” ID 5 was a Type 1, so indeed it had been a cash cow for this subject except for the fact that he or she had incurred the voting cost in every round.

It was at this point that a rare event took place: the experiment was actually paused, and the rules and mechanics of payoffs and costs were explained again, such as the fact that the payoff received is a function of your type and the candidate that wins, not (directly) whether or for whom a subject voted. (This had already been stated several times during the instructions, which had been shown in writing and also read aloud at the start of the session. The instructions had also included a practice question that was specifically designed to make sure that subjects understood this and to correct any misunderstandings before the experiment began.) After this was explained again, voting decisions settled into a more normal pattern.

This incident provides an excellent example how the spread of information can affect outcomes, regardless of whether or not the information is correct. It is also an example of how relatively easy it is to exploit people’s uncertainty, as would have been the case at the beginning of the session, when there is always a learning curve regardless of how thorough the instructions are. It is important to note that this does not necessarily have to be done maliciously—it is unclear whether ID 3’s actions were due to misunderstanding or wanting to cause trouble.

*3.4.2. Peer Pressure/Peer Influence/Manipulation.* A second theme that was seen in the chat transcripts was that of subjects

urging each other to vote and sometimes agreeing to do so. This may have increased the turnout rates beyond what might have occurred without chat and may help to explain the high turnout rates that were observed. Also, because the chat is repeated over many rounds there is accountability to some extent—your chat neighbor might ask if you had voted and, if so, for whom you voted in the previous round. That sense of accountability could also increase voter turnout. And finally, some people feel a sense of solidarity in voting the same way as their friends/family/colleagues/and so forth.

In these experiments, this feeling of solidarity led subjects to vote against their own economic interests, even in the absence of altruistic motives. Usually this only lasted one or two rounds before self-interest took over. However, in a few cases this happened repeatedly. These present a possible parallel to Kittel et al.’s [12] swing voters who were influenced to vote differently than they would have otherwise. In this context, swing voter status was not due to experimental design but rather to certain subjects being more easily influenced than others.

One extreme case occurred in Session 10. ID 3, who was a Type 2 subject, was paired in a chat group with ID 2 and ID 4, both of whom belonged to Type 1. ID 3 wanted to vote with his or her chat group and voted for Candidate X—against the subject’s economic self-interest, but benefitting those of the other group members—in every round throughout the session. The subject asked whether they were all the same type before round 6, in spite of the types having been listed next to ID numbers in the chat interface in every round. However, even after ID 3’s group members clarified that they were both Type 2, ID 3 continued to vote for Candidate X. ID 3 voted for Candidate X in all but one round. Coincidentally, this extra vote for Candidate X either made or broke a tie in each of the first five rounds and set up a streak in which Candidate X won every round of the session, as subjects who would have voted for Y (mostly other Type 2s) got discouraged and stopped voting. I spoke with the subject after the end of the session to find out if this voting behavior resulted from a lack of understanding about the way in which the payoffs worked (being careful to ask about it in terms of wanting to find out whether the instructions were clear enough so that changes could be made, in an attempt to sidestep feelings of defensiveness that could have affected the answers). The subject had indeed understood everything but simply liked the feeling of solidarity from voting with the other chat group members.

In Session 11 there were two Type 1 subjects who often voted for Candidate Y. The payoffs in this session were symmetric, so altruism was not a possible motive. Both of them chatted with ID 15, a Type 2 candidate. ID 15 usually voted for Y but sometimes voted for X, usually when IDs 14 and 16 also voted X, and this subject spent much of the chat trying to convince his or her chat group to vote as a group. When asked about it after the session, IDs 14 and 16 each expressed a desire to vote with their chat groups, similar to the previous example. But then in a separate conversation ID 15 explained his strategy of occasionally voting for Candidate X—especially in the earlier rounds—for the purpose of generating a feeling of solidarity in IDs 14

and 16. This subject then would urge IDs 14 and 16 to vote for Candidate Y, which maximized ID 15's payoff. The end result was that ID 15 only voted for Candidate X four times, whereas ID 14 voted for Candidate Y eight times and ID 16 voted for Candidate Y twelve times, so on a net basis this strategy gained many more votes than it lost for Candidate Y.

This was the only incident of strategic manipulation that was observed in any of the sessions, at least as far as the transcripts indicated (and a bit of luck in running into the subject shortly after the session). However, it does illustrate the ability of a savvy political strategist to manipulate others who may not be as savvy, even those whose interests diverge from those of the strategist. It is also thematically related to the producers of social pressure in Schram and van Winden [9].

#### 4. Conclusions

The analysis presented above attempted to resolve the paradox of voter turnout by incorporating ethical voting into a quantal response equilibrium- (QRE-) based model which allows for noise in the decision-making process and then testing the model's predictions using a series of laboratory experiments. The benefit of this modeling approach is that it generates predicted voting turnout probabilities, which were used in fine-tuning the experimental design, and against which the results of the experiments could be compared.

High rates of voter turnout were indeed observed. In fact, the turnout rates for some of the type-candidate combinations were so high that QRE-based analysis could not account for all of the overvoting, relative to the Nash equilibrium predictions. This is due at least in part to the presence of three voting choices—vote for Candidate X, vote for Candidate Y, and abstain—as opposed to the two choices of vote and abstain that were present in Levine and Palfrey [4]. Given a cost-payoff ratio that generates at least some abstention in equilibrium, the highest possible payoff occurs in the presence of extremely high levels of noise. With three choices, the probability of choosing any one of them is 0.33. This is considerably lower than the 0.50 when there are only two choices, which would not have explained all of the overvoting observed for some of the type-candidate combinations but would have at least explained more of it.

Ethical voting was not able to explain the overvoting, either. Some ethical voting was observed, but not enough to explain the extent of overvoting. Also, most of the overvoting occurred when subjects voted for the candidate offering their type a higher payoff. Ethical voting, which is defined as voting against one's own economic self-interest in order to vote for the candidate offering payoffs that are distributed more equitably across voter types, by its very nature does not fall into this category.

The online chat feature that subjects used for communicating with their virtual neighbors before each round (election) was shown to have increased voter participation. In this chat feature, subjects chatted online with the two subjects with adjacent ID numbers within the experiment. This mirrored the way in which people discuss politics

and upcoming elections with their friends, families, and coworkers.

Additionally, because the chat was repeated before every round it also set up a dynamic in which it may have generated some accountability between chat partners, for example, by asking one's chat group whether they had voted in the previous round, and if so then for which candidate(s). It was also observed in the chat transcripts that subjects would often ask others to vote in a certain way or agree to vote in a certain way and then sometimes follow up afterwards, which supports that conjecture. Through these channels, subjects' voting decisions were (mostly positively) influenced by the voting decisions of their group members.

There are several possibilities for future research in this area, such as comparing sessions with a chat feature versus no communication or the use of different chat group structures, for example, the current chat setup versus one in which subjects only communicated with members of their own voter type. Another possibility would be to focus on only ethical voting or only communication to clarify the individual impacts of these features on turnout rates and patterns. It is possible that the inclusion of both communication and (the potential for) ethical voting in this paper's experimental design obscured the importance of ethical voting. The chat feature was originally included for the insight that it would provide rather than being a tool to increase voting participation, and its impact on turnout was unexpected. Because of this, the experimental design did not carefully test the relative impacts of these two features. It would be interesting to design treatments to test for the individual effects, which I leave for future research.

Several other themes were also observed in the results, in which the patterns in the observed turnout rates differed from those that had been predicted. There was a strong cost effect on abstention rates, as predicted. However, there was no significant size effect, in contrast to the predictions in which turnout declined as the session size increased and the probability of casting a pivotal vote decreased. Second, the turnout rate was higher in sessions with symmetric payoffs (in which there was no possibility of ethical voting) than in sessions with asymmetric payoffs (in which there was a possibility of ethical voting). This contradicts the predicted outcome; however, the reason for this pattern is unclear. Belief elicitation might provide some insight into this but would require further research.

## Appendices

### A. Pivotal Probabilities

The term  $(1/2) \times (\Pr(\text{make tie})_{ij} + \Pr(\text{break tie})_{ij})$  is the probability that a vote cast by a member of type  $i$  for candidate  $j$  will be pivotal (henceforth referred to as the "pivotal probability"). The one-half is there because, in the presence of two candidates, if someone's vote *makes* a tie then there is a 50% chance that this voter's chosen candidate will win (and that the vote will have been pivotal)—remember that the winner of a tie is determined randomly. If someone's vote *breaks* a tie then there is a 100% chance that his or her chosen

candidate will win. However, because there was already a 50% chance that this candidate would have won in the event of a tie, there would only be a 50% chance that the vote that broke the tie changed that candidate's outcome from losing to winning. Therefore, the sum of the probabilities of making a tie and of breaking a tie is multiplied by one-half.

The pivotal probabilities are analogous to those in Levine and Palfrey [4] and Goeree and Holt [3] but complicated by the fact that it is possible for votes received by each of the candidates to vote for either candidate to come from either type of voters. For example, even in the relatively straightforward example of 1 vote (total) for Candidate X and 1 vote (total) for Candidate Y, there are numerous possibilities for how this outcome arose: both votes came from Type 1 voters; Type 1 voted for Candidate X and Type 2 voted for Candidate Y; Type 2 voted for Candidate X and Type 1 voted for Candidate Y; and both votes came from Type 2 voters. Additionally, there is a separate pivotal probability for each type-candidate combination because it is possible for members of either type to vote for either candidate.

One example is the probability that a member of Type 1 will be pivotal by voting for Candidate X, modeled as a multinomial probability:

$$\begin{aligned}
 & \Pr(\text{pivotal})_{1X} \\
 &= \frac{1}{2} \left( \sum_{n_Y=1}^{\min(N_1, N_2)} \sum_{n_{2Y}=0}^{n_Y} \sum_{n_{2X}=0}^{n_Y-1} \binom{N_2}{n_{2X}, n_{2Y}} \right. \\
 & \quad \cdot \binom{N_1-1}{n_Y-1-n_{2X}, n_Y-n_{2Y}} \\
 & \quad \times p_{2X}^{n_{2X}} p_{2Y}^{n_{2Y}} (1-p_{2X}-p_{2Y})^{N_2-n_{2X}-n_{2Y}} \\
 & \quad \cdot p_{1X}^{n_Y-1-n_{2X}} p_{1Y}^{n_Y-n_{2Y}} \\
 & \quad \cdot (1-p_{1X}-p_{1Y})^{N_1-(n_Y-n_{2X})-(n_Y-n_{2Y})} \\
 & \quad + \sum_{n_Y=0}^{\min(N_1, N_2)-1} \sum_{n_{2Y}=0}^{n_Y} \sum_{n_{2X}=0}^{n_Y} \binom{N_2}{n_{2X}, n_{2Y}} \\
 & \quad \cdot \binom{N_2-1}{n_Y-n_{2X}, n_Y-n_{2Y}} \\
 & \quad \times p_{2X}^{n_{2X}} p_{2Y}^{n_{2Y}} (1-p_{2X}-p_{2Y})^{N_2-n_{2X}-n_{2Y}} \\
 & \quad \cdot p_{1X}^{n_Y-n_{2X}} p_{1Y}^{n_Y-n_{2Y}} \\
 & \quad \cdot (1-p_{1X}-p_{1Y})^{N_1-1-(n_Y-n_{2X})-(n_Y-n_{2Y})} \Big), \tag{A.1}
 \end{aligned}$$

where  $N_i$  is the total number of potential voters of type  $i$ ,  $n_j$  is the total number of votes for candidate  $j$  (cast by voters of either type),  $n_{ij}$  is the total number of votes by members of type  $i$  for candidate  $j$ , and  $p_{ij}$  is the probability of a member of type  $i$  voting for candidate  $j$ .

The first term within the brackets is the probability that a member of Type 1 will make a tie by voting for Candidate X. This means that before this incremental voting decision is made, there is one fewer vote for Candidate X than there is for Candidate Y, or  $n_X = n_Y - 1$ . The second term is the probability that a member of Type 1 will break a tie by voting for Candidate X. The fact that one vote would break a tie means that there is currently a tie and  $n_X = n_Y$ .

For each term within the brackets, the triple sum allows the calculation to sum over every possible combination of votes, for every possible number of votes. The outer sum loops over every possible value of  $n_Y$  (the total number of votes for Candidate Y), and therefore also over every possible value of  $n_X$ . The inner two sums loop over every possible combination of  $n_{2Y}$  and  $n_{2X}$  (the number of votes cast by members of Type 2 for Candidates Y and X, resp.). Because  $n_{1Y}$  and  $n_{1X}$ —the number of votes cast by members of Type 2 for Candidates X and Y, respectively—are simply  $(n_Y - n_{2Y})$  and  $(n_X - n_{2X})$ , these summations also loop over every possible combination of  $n_{1Y}$  and  $n_{1X}$ .

In the first term,  $\binom{N_2}{n_{2X}, n_{2Y}} p_{2X}^{n_{2X}} p_{2Y}^{n_{2Y}} (1-p_{2X}-p_{2Y})^{N_2-n_{2X}-n_{2Y}}$  is the probability that there will be exactly  $n_{2X}$  and  $n_{2Y}$  votes if all  $N_2$  members of Type 2 have already made their voting (or abstention) decisions.  $\binom{N_1-1}{n_Y-1-n_{2X}, n_Y-n_{2Y}} p_{1X}^{n_Y-1-n_{2X}} p_{1Y}^{n_Y-n_{2Y}} (1-p_{1X}-p_{1Y})^{N_1-(n_Y-n_{2X})-(n_Y-n_{2Y})}$  is the probability that there will be exactly  $n_{1X}$  and  $n_{1Y}$  votes if all  $N_1 - 1$  other members of Type 1 have already voted.

The second term follows the same format.

The other probabilities of a member of type  $i$  being pivotal by voting for candidate  $j$  equal

$$\begin{aligned}
 & \Pr(\text{pivotal})_{1Y} \\
 &= \frac{1}{2} \left( \sum_{n_X=1}^{\min(N_1, N_2)} \sum_{n_{2X}=0}^{n_X} \sum_{n_{2Y}=0}^{n_X-1} \binom{N_2}{n_{2X}, n_{2Y}} \right. \\
 & \quad \cdot \binom{N_1-1}{n_X-n_{2X}, n_X-1-n_{2Y}} \\
 & \quad \times p_{2X}^{n_{2X}} p_{2Y}^{n_{2Y}} (1-p_{2X}-p_{2Y})^{N_2-n_{2X}-n_{2Y}} \\
 & \quad \cdot p_{1X}^{n_X-n_{2X}} p_{1Y}^{n_X-1-n_{2Y}} \\
 & \quad \cdot (1-p_{1X}-p_{1Y})^{N_1-(n_X-n_{2X})-(n_X-n_{2Y})} \\
 & \quad + \sum_{n_X=0}^{\min(N_1, N_2)-1} \sum_{n_{2X}=0}^{n_X} \sum_{n_{2Y}=0}^{n_X} \binom{N_2}{n_{2X}, n_{2Y}} \\
 & \quad \cdot \binom{N_2-1}{n_X-n_{2X}, n_X-n_{2Y}} \\
 & \quad \times p_{2X}^{n_{2X}} p_{2Y}^{n_{2Y}} (1-p_{2X}-p_{2Y})^{N_2-n_{2X}-n_{2Y}} \\
 & \quad \cdot p_{1X}^{n_X-n_{2X}} p_{1Y}^{n_X-n_{2Y}} \\
 & \quad \cdot (1-p_{1X}-p_{1Y})^{N_1-1-(n_X-n_{2X})-(n_X-n_{2Y})} \Big),
 \end{aligned}$$

$\Pr(\text{pivotal})_{2X}$ 

$$\begin{aligned}
&= \frac{1}{2} \left( \sum_{n_Y=1}^{\min(N_1, N_2)} \sum_{n_{1Y}=0}^{n_Y} \sum_{n_{1X}=0}^{n_Y-1} \binom{N_1}{n_{1X}, n_{1Y}} \right. \\
&\quad \cdot \binom{N_2-1}{n_Y-1-n_{1X}, n_Y-n_{1Y}} \\
&\quad \times p_{1X}^{n_{1X}} p_{1Y}^{n_{1Y}} (1-p_{1X}-p_{1Y})^{N_1-n_{1X}-n_{1Y}} \\
&\quad \cdot p_{2X}^{n_Y-1-n_{1X}} p_{2Y}^{n_Y-n_{1Y}} \\
&\quad \cdot (1-p_{2X}-p_{2Y})^{N_2-(n_Y-n_{1X})-(n_Y-n_{1Y})} \\
&\quad + \sum_{n_Y=0}^{\min(N_1, N_2)-1} \sum_{n_{1Y}=0}^{n_Y} \sum_{n_{1X}=0}^{n_Y} \binom{N_1}{n_{1X}, n_{1Y}} \\
&\quad \cdot \binom{N_2-1}{n_Y-n_{1X}, n_Y-n_{1Y}} \\
&\quad \times p_{1X}^{n_{1X}} p_{1Y}^{n_{1Y}} (1-p_{1X}-p_{1Y})^{N_1-n_{1X}-n_{1Y}} \\
&\quad \cdot p_{2X}^{n_Y-n_{1X}} p_{2Y}^{n_Y-n_{1Y}} \\
&\quad \cdot (1-p_{2X}-p_{2Y})^{N_2-1-(n_Y-n_{1X})-(n_Y-n_{1Y})} \Big),
\end{aligned}$$

 $\Pr(\text{pivotal})_{2Y}$ 

$$\begin{aligned}
&= \frac{1}{2} \left( \sum_{n_X=1}^{\min(N_1, N_2)} \sum_{n_{1X}=0}^{n_X} \sum_{n_{1Y}=0}^{n_X-1} \binom{N_1}{n_{1X}, n_{1Y}} \right. \\
&\quad \cdot \binom{N_2-1}{n_X-n_{1X}, n_X-1-n_{1Y}} \\
&\quad \times p_{1X}^{n_{1X}} p_{1Y}^{n_{1Y}} (1-p_{1X}-p_{1Y})^{N_1-n_{1X}-n_{1Y}} \\
&\quad \cdot p_{2X}^{n_X-n_{1X}} p_{2Y}^{n_X-1-n_{1Y}} \\
&\quad \cdot (1-p_{2X}-p_{2Y})^{N_2-(n_X-n_{1X})-(n_X-n_{1Y})} \\
&\quad + \sum_{n_X=0}^{\min(N_1, N_2)-1} \sum_{n_{1X}=0}^{n_X} \sum_{n_{1Y}=0}^{n_X} \binom{N_1}{n_{1X}, n_{1Y}} \\
&\quad \cdot \binom{N_2-1}{n_X-n_{1X}, n_X-n_{1Y}} \\
&\quad \times p_{1X}^{n_{1X}} p_{1Y}^{n_{1Y}} (1-p_{1X}-p_{1Y})^{N_1-n_{1X}-n_{1Y}} \\
&\quad \cdot p_{2X}^{n_X-n_{1X}} p_{2Y}^{n_X-n_{1Y}} \\
&\quad \cdot (1-p_{2X}-p_{2Y})^{N_2-1-(n_X-n_{1X})-(n_X-n_{1Y})} \Big).
\end{aligned}$$

(A.2)

TABLE 9: Voting turnout rates by gender.

	Male	Female
Turnout	0.54633	0.5993
Favored	0.50275	0.5162
Unfavored	0.04358	0.0831
Ethical	0.0375	0.01833

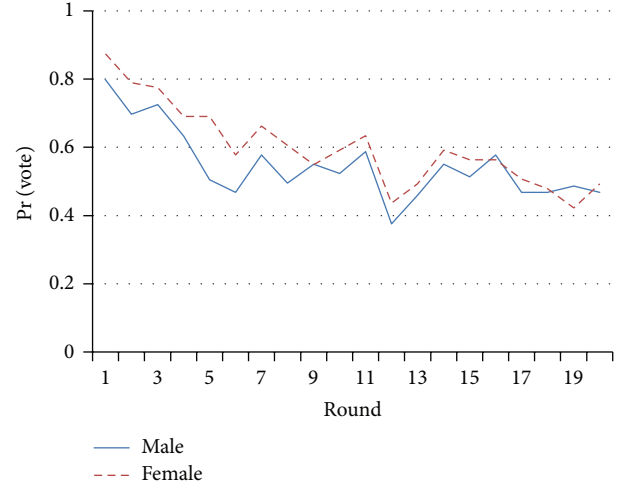


FIGURE 5: Voting turnout rates by gender.

These pivotal probabilities are then used to calculate the expected payoff differences, net of costs.

## B. Gender Differences

Another area of interest is the effect of gender on voting decisions. In various nonparametric and parametric analyses gender was not found to have any significant effect on voting turnout, either for overall, for favored, or for unfavored. The turnout rates are summarized in Table 9 and Figure 5 illustrates the closeness of these turnout rates over time.

## C. Nonparametric Permutation Test

Nonparametric permutation tests were used to test the significance of the difference between the means of two independent samples, when the sample sizes are small. The null hypothesis is that all of the observations are from the same population. Therefore, rejecting the null hypothesis would mean that all of the observations could not be from the same population, and the fact that the observations occurred specifically in their respective treatments cannot be attributed to coincidence. This was tested by reassigning the observations between the treatments, then comparing the differences in the means of the permutations with those in the original. The null hypothesis is rejected if the observed difference between treatments is large relative to the hypothetical differences in the other possible permutations.

First, looking at session-level average turnout rates for overall turnout, shown in Table 10, we see the same relationships as before in the averages for all rounds and for

TABLE 10: Voting turnout rates by session.

Treatment	Session	1st round	Average	Average (2nd half)
12, asymmetric	Session 1	0.8333	0.5625	0.4417
	Session 3	0.9167	0.5000	0.3750
	Session 9	0.8333	0.5208	0.4750
12, symmetric	Session 2	0.5833	0.6292	0.6500
	Session 4	0.8333	0.5292	0.4667
	Session 12	1.0000	0.6083	0.4750
18, asymmetric	Session 5	0.9444	0.5306	0.4778
	Session 7	0.7778	0.4778	0.4833
	Session 10	0.7778	0.5111	0.4000
18, symmetric	Session 6	0.7222	0.6306	0.6222
	Session 8	1.0000	0.6306	0.5556
	Session 11	0.7222	0.6583	0.5833

TABLE 11: Permutation test results ( $p$  value)—overall turnout rates.

	First round only		Average		Average (2nd half)	
	1-tailed	2-tailed	1-tailed	2-tailed	1-tailed	2-tailed
12: asymmetric versus symmetric	0.30	0.75	0.1*	0.200	0.150	0.300
18: asymmetric versus symmetric	0.50	1.00	0.05**	0.1*	0.05**	0.1*
Asymmetric: 12 versus 18	0.30	0.60	0.250	0.500	0.200	0.450
Symmetric: 12 versus 18	0.40	0.90	0.05**	0.1*	0.150	0.350
12 versus 18 (asymmetric and symmetric)	0.4142	0.8270	0.2767	0.5520	0.1357	0.2719
Asymmetric versus symmetric (12 and 18)	0.3419	0.6848	0.003***	0.006***	0.0038***	0.0078***

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

the second half (rounds 11–20). The first round of voting was also included because in most real-world elections one would only vote a single time.

The nonparametric permutation test then analyzes whether there really is a difference across treatments. The results are shown in Table 11. The results are reported as  $p$  values.

These results show the same trends that were observed in the stylized facts: the number of subjects per session does not significantly affect the probability of voting (except for symmetric sessions across all rounds), and subjects in treatments with symmetric payoffs are significantly more likely to vote than are subjects in treatments with asymmetric payoffs. Results for the first round only are not significant, which is not surprising. As opposed to an election outside of the lab, where potential voters have weeks or months to learn about the candidates and issues and to talk with their friends/family/colleagues, in the experiment there is quite a bit of learning that occurs during the first few rounds which may obscure any meaningful relationships in the data.

Next, the data for favored turnout and unfavored turnout were analyzed in the same manner. Given that most of the voting was for the favored candidate, the results found here were very similar to those for overall turnout. For unfavored turnout, no significant differences were found even after using the averages for rounds 1–10 (because that is when almost all of the ethical voting happened).

## D. Instructions

The subjects saw the following set of instructions at the start of the experiment. These were also read aloud. The instructions below are from a 12-person session with asymmetric payoffs, so the instructions for the other treatments would vary with respect to the number of subjects and/or the payoff charts but would otherwise be the same. At the top of each page, the subject sees “Instructions (ID = —), Page — of 6”.

### Page 1

- (i) *Matchings*. The experiment consists of a series of *rounds*. You will be matched with the *same* group of 11 other people in each round. The decisions that you and the 11 other people make will determine the amounts earned by each of you.
- (ii) *Voting Decisions*. At the beginning of each round, you will be asked to consider a vote for one of 2 options: Option X and Option Y.
- (iii) *Earnings*. The votes that are cast by the members of your group will determine a winning option, which will determine your earnings in a manner to be explained next.
- (iv) *Please Note*. Your earnings will depend on the outcome that receives the most votes (with ties decided at random), regardless of whether or not you voted for the winning option.



- (v) *Communications*. There will be a *chat room* open prior to the voting process, which allows you to send and receive messages. Voters will be identified by their ID numbers, and you are free to discuss any aspect of the voting process during the 1-minute chat period.
- (vi) *Communications Network*. You will be able to communicate with only two other voters in your group, who are your neighbors in the sense of having ID numbers adjacent to your ID. You should avoid inappropriate language or attempts to arrange side payments.

## Page 2

- (i) *Voting Sequence*. The option that receives the most votes will determine earnings for all voters (irrespective of how each individual voted).
- (ii) *Ties*. In the event of a tie, one of the tied options will be selected at random, with each tied option having an equal chance of being selected.
- (iii) *Your Earnings*. The option selected in the voting process will determine your earnings. For example, your earnings for each outcome in the first round are

Option X: \$(individual payoff),

Option Y: \$(individual payoff).

- (iv) *Voting Cost*. If you do vote, you will incur a cost. This cost will change randomly from round to round, and it will vary randomly from person to person. Each person's voting cost will be a number that is randomly selected from a range between \$0.00 and \$0.42. You will find out your voting cost before you decide whether to vote. You will not know anyone else's voting cost. Your total earnings will be the amount determined by the voting outcome, minus your voting cost (if any).

## Page 3

- (i) *Individual Differences*. There are 12 voters in your group, who are divided into 2 "types," as shown in the table below. Your type and your earnings for each possible outcome are indicated in the *bright blue column*.
- (ii) *Payoff Backstory*.

Option X will implement an investment that pays \$2 to Type 1 voters and \$1 to Type 2 voters; Option Y will implement an investment that pays \$1.50 to both Type 1 and Type 2 voters.

Payoffs for all voters		
Outcome	type 1 (you)	type 2
Option X	\$2.00	\$1.00
Option Y	\$1.50	\$1.50
Number	6	6
Voters:	(you and 5 others)	

Page 4. Please select the best answer. Your answer will be checked for you on the page that follows.

Question 1. Suppose that your voting cost is \$0.21 and you are of Type 1 in this round (different people will have different types). If you decide to vote and incur this cost, your earnings will be

- ☐ \$1.79 if Option X receives the most votes and \$1.29 if Option Y receives the most votes;
- ☐ \$1.79 if you voted for Option X and Option X wins, \$1.29 if you voted for Option Y and Option Y wins, and \$0.00 if the option you voted for does not win.

Submit answers

Payoffs for all voters

Outcome	type 1 (you)	type 2
Option X	\$2.00	\$1.00
Option Y	\$1.50	\$1.50
Number	6	6
Voters:	(you and 5 others)	

## Page 5

Question 1. Suppose that your voting cost is \$0.21 and you are of Type 1 in this round (different people will have different types). If you decide to vote and incur this cost, your earnings will be

- ☐ \$1.79 if Option X receives the most votes and \$1.29 if Option Y receives the most votes;
- ☐ \$1.79 if you voted for Option X and Option X wins, \$1.29 if you voted for Option Y and Option Y wins, and \$0.00 if the option you voted for does not win.

Your answer is *Correct*. You receive the payoff from the relevant (blue) column in the payoff table, regardless of whether you voted or not. In this case, the cost of voting is deducted since you voted in this example.

Continue

Payoffs for all voters

Outcome	type 1 (you)	type 2
Option X	\$2.00	\$1.00
Option Y	\$1.50	\$1.50
Number	6	6
Voters:	(you and 5 others)	

## Page 6 (instructions summary page)

- (i) You will be matched with the *same* group of 11 other people in each round.
- (ii) All participants will begin by finding out their own voting cost, which will be a randomly determined number between \$0.00 and \$0.42. This cost will vary

from person to person. After seeing your voting cost, you will decide whether to incur this cost and vote. The option that receives the most votes will be selected. Your earnings will depend on the option selected, although your voting cost (if any) will be subtracted.

- (iii) *Please Note.* Your earnings will be determined by the outcome that receives the most votes (with ties decided at random), regardless of whether or not you voted for the winning option.
- (iv) Each round will begin with a 1-minute chat period. You will be able to communicate with only two other voters in your group, who are your “neighbors” in the sense of having ID numbers adjacent to your ID number. You will have to press the Update button manually to retrieve other’s messages as they arrive.
- (v) There will be a number of rounds in this part of the experiment. Your earnings for each round will be calculated for you and added to previous earnings.

## Disclaimer

The views expressed are those of the author and should not be attributed to the U.S. Department of Agriculture or the Economic Research Service.

## Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

This paper was based on a chapter from a Ph.D. dissertation completed at the University of Virginia under the helpful supervision of Charles Holt, Federico Ciliberto, and Simon Anderson. The author would also like to thank Craig Volden, Wayne-Roy Gayle, Beatrice Boulu-Reshef, Sean Sullivan, Angela Smith, Jura Liaukonyte, Jonathan Williams, Christopher Clapp, Nathaniel Higgins, Christopher Burns, James Williamson, Michael Harris, Charles Hallahan, and the participants of the Experimental Economics Workshop and the Political Psychology Working Group, both at the University of Virginia, for their helpful comments. The author is grateful to Arthur Schram and two anonymous referees for their comments and suggestions. In addition, the author thanks the Political Psychology Working Group, the Bankard Fund for Political Economy, and the National Science Foundation (Award no. BCS-0904795) for their financial support and the Department of Economics and the Veconlab Experimental Economics Laboratory at the University of Virginia for the experimental infrastructure.

## References

- [1] B. Geys, “‘Rational’ theories of voter turnout: a review,” *Political Studies Review*, vol. 4, no. 1, pp. 16–35, 2006.

- [2] R. D. McKelvey and T. R. Palfrey, “Quantal response equilibria for normal form games,” *Games and Economic Behavior*, vol. 10, no. 1, pp. 6–38, 1995.
- [3] J. K. Goeree and C. A. Holt, “An explanation of anomalous behavior in models of political participation,” *American Political Science Review*, vol. 99, no. 2, pp. 201–213, 2005.
- [4] D. K. Levine and T. R. Palfrey, “The paradox of voter participation? A laboratory study,” *The American Political Science Review*, vol. 101, no. 1, pp. 143–158, 2007.
- [5] R. B. Morton and J.-R. Tyran, “Ethical versus selfish motivations and turnout in small and large electorates,” Working Paper, New York University, 2012.
- [6] J. K. Goeree, C. A. Holt, and S. K. Laury, “Private costs and public benefits: unraveling the effects of altruism and noisy behavior,” *Journal of Public Economics*, vol. 83, no. 2, pp. 255–276, 2002.
- [7] R. B. Morton, “Groups in rational turnout models,” *American Journal of Political Science*, vol. 35, no. 3, pp. 758–776, 1991.
- [8] J. Großer and A. Schram, “Neighborhood information exchange and voter participation: an experimental study,” *The American Political Science Review*, vol. 100, no. 2, pp. 235–248, 2006.
- [9] A. Schram and F. van Winden, “Why People Vote: free riding and the production and consumption of social pressure,” *Journal of Economic Psychology*, vol. 12, no. 4, pp. 575–620, 1991.
- [10] A. Schram and J. Sonnemans, “Why people vote: experimental evidence,” *Journal of Economic Psychology*, vol. 17, no. 4, pp. 417–442, 1996.
- [11] G. Charness, L. Rigotti, and A. Rustichini, “Individual behavior and group membership,” *American Economic Review*, vol. 97, no. 4, pp. 1340–1352, 2007.
- [12] B. Kittel, W. Luhan, and R. Morton, “Communication and voting in multi-party elections: an experimental study,” *The Economic Journal*, vol. 124, pp. 196–225, 2014.
- [13] T. R. Palfrey and H. Rosenthal, “A strategic calculus of voting,” *Public Choice*, vol. 41, no. 1, pp. 7–53, 1983.
- [14] J. Andreoni, “Impure altruism and donations to public goods: a theory of warm-glow giving,” *The Economic Journal*, vol. 100, no. 401, pp. 464–477, 1990.
- [15] J. H. Fowler and C. D. Kam, “Beyond the self: social identity, altruism, and political participation,” *The Journal of Politics*, vol. 69, no. 3, pp. 813–827, 2007.

## Research Article

# Preference for Efficiency or Confusion? A Note on a Boundedly Rational Equilibrium Approach to Individual Contributions in a Public Good Game

Luca Corazzini<sup>1,2</sup> and Marcelo Tyszler<sup>3</sup>

<sup>1</sup>Department of Law Science and History of Institutions, University of Messina, Piazza XX Settembre 4, 98122 Messina, Italy

<sup>2</sup>ISLA, Bocconi University, Via Roentgen 1, 20136 Milan, Italy

<sup>3</sup>Center for Research in Experimental Economics and Political Decision Making (CREED), University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, Netherlands

Correspondence should be addressed to Marcelo Tyszler; [marcelo@tyszler.com.br](mailto:marcelo@tyszler.com.br)

Received 5 December 2014; Revised 25 March 2015; Accepted 29 March 2015

Academic Editor: Dimitris Fotakis

Copyright © 2015 L. Corazzini and M. Tyszler. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

By using data from a voluntary contribution mechanism experiment with heterogeneous endowments and asymmetric information, we estimate a quantal response equilibrium (QRE) model to assess the relative importance of efficiency concerns versus noise in accounting for subjects overcontribution in public good games. In the benchmark specification, homogeneous agents, overcontribution is mainly explained by error and noise in behavior. Results change when we consider a more general QRE specification with cross-subject heterogeneity in concerns for (group) efficiency. In this case, we find that the majority of the subjects make contributions that are compatible with the hypothesis of preference for (group) efficiency. A likelihood-ratio test confirms the superiority of the more general specification of the QRE model over alternative specifications.

## 1. Introduction

Overcontribution in linear public good games represents one of the best documented and replicated regularities in experimental economics. The explanation of this apparently irrational behaviour, however, is still a debate in the literature. This paper is aimed at investigating the relative importance of noise versus preference for efficiency. In this respect, we build and estimate a quantal response equilibrium (henceforth, QRE [1]) extension of the model presented by Corazzini et al. [2]. This boundedly rational model formally incorporates both preference for efficiency and noise. Moreover, in contrast to other studies that investigate the relative importance of error and other-regarding preferences, the QRE approach explicitly applies an equilibrium analysis.

To reconcile the experimental evidence with the standard economic framework, social scientists developed explanations based on refinements of the hypothesis of “other-regarding preferences”: reciprocity [3–6], altruism and spitefulness [7–9], commitment and Kantianism [10, 11], norm compliance [12], and team-thinking [13–15].

Recently, an additional psychological argument to explain agents’ attitude to freely engage in prosocial behavior is gaining increasing interest: the hypothesis of preference for (group) efficiency. There is evidence showing that experimental subjects often make choices that increase group efficiency, even at the cost of sacrificing their own payoff [16, 17]. Corazzini et al. [2] use this behavioral hypothesis to explain evidence from linear public good experiments based on prizes (a lottery, a first price all pay auction, and

a voluntary contribution mechanism used as a benchmark), characterized by endowment heterogeneity and incomplete information on the distribution of incomes. In particular, they present a simple model in which subjects bear psychological costs from contributing less than what is efficient for the group. The main theoretical prediction of their model when applied to linear public good experiments is that the equilibrium contribution of a subject is increasing in both her endowment and the weight attached to the psychological costs of (group-) inefficient contributions in the utility function. The authors show that this model is capable of accounting for overcontribution as observed in their experiment, as well as evidence reported by related studies.

However, as argued by several scholars, rather than being related to subjects' kindness, overcontribution may reflect their natural propensity to make errors. There are several experimental studies [18–23] that seek to disentangle other-regarding preferences from pure noise in behavior by running *ad hoc* variants of the linear public good game. A general finding of these papers is that “warm-glow effects and random error played both important and significant roles” [20, p. 842] in explaining overcontribution.

There are several alternative theoretical frameworks that can be used to model noise in behavior (bounded rationality) and explain experimental evidence in strategic games. Two examples are the “level- $k$ ” model (e.g., [24–26]) and (reinforcement) learning models (e.g., [27]). In the “level- $k$ ” model of iterated dominance, “level-0” subjects choose an action randomly and with equal probability over the set of possible pure strategies while “level- $k$ ” subjects choose the action that represents the best response against level- $(k - 1)$  subjects. Level- $k$  models have been used to account for experimental results in games in which other-regarding preferences do not play any role, such as  $p$ -Beauty contests and other constant sum games. Since in public good games there is a strictly dominant strategy of no contribution, unless other-regarding preferences are explicitly assumed, “level- $k$ ” models do not apply. Similar arguments apply to learning models. In the basic setting, each subject takes her initial choice randomly and with equal probability over the set of possible strategies. As repetition takes place, strategies that turn out to be more profitable are chosen with higher probability. Thus, unless other-regarding preferences are explicitly incorporated into the utility function, repetition leads to the Nash equilibrium of no contribution.

The QRE approach has the advantage that even in the absence of other-regarding preferences it can account for overcontribution in equilibrium. Moreover, we can use the model to assess the relative importance of noise versus efficiency concerns.

We start from a benchmark model in which the population is homogeneous in both concerns for (group) efficiency and the noise parameter. We then allow for heterogeneity across subjects by assuming the population to be partitioned into subgroups with different degrees of efficiency concerns but with the same value for the noise parameter.

In the QRE model with a homogeneous population, we find that subjects' overcontribution is entirely explained by

noise in behavior, with the estimated parameter of concerns for (group) efficiency being zero. A likelihood-ratio test strongly rejects the specification not allowing for randomness in contributions in favor of the more general QRE model. A different picture emerges when heterogeneity is introduced in the QRE model. In the model with two subgroups, the probability of a subject being associated with a strictly positive degree of preference for (group) efficiency is approximately one-third. This probability increases to 59% when we add a third subgroup characterized by an even higher efficiency concern. A formal likelihood-ratio test confirms the superiority of the QRE model with three subgroups over the other specifications. These results are robust to learning processes over repetitions. Indeed, estimates remain qualitatively unchanged when we replicate our analysis on the last 25% of the experimental rounds. The rest of this paper is structured as follows. In Section 2, we describe the experimental setting. In Section 3, we present the QRE extension of the model based on the preference for (group) efficiency hypothesis. Section 4 reports results from our statistical analysis. Section 5 concludes the paper.

## 2. The Experiment

We use data from three sessions of a voluntary contribution mechanism reported by Corazzini et al. [2]. Each session consisted of 20 rounds and involved 16 subjects. At the beginning of each session, each subject was randomly and anonymously assigned, with equal chance, an endowment of either 120, 160, 200, or 240 tokens. The endowment was assigned at the beginning of the experiment and was kept constant throughout the 20 rounds. The experiment was run in a strangers condition [28] such that, at the beginning of each round, subjects were randomly and anonymously rematched in groups of four players. This procedure was common knowledge. Thus, in each round, subjects made their choices under incomplete information on the distribution of the endowments in their group. In each round, every subject had to allocate her endowment between an individual and a group account. The individual account implied a private benefit such that, for each token a subject allocated to the individual account, she received two tokens. On the other hand, tokens in the group account generated monetary returns to each of the group members. In particular, each subject received one token for each token allocated by her or by any other member of her group to the group account. Thus, the marginal per capita return used in the experiment was 0.5. At the beginning of each round, the experimenter exogenously allocated 120 tokens to the group account, independently of subjects' choices, thus implying 120 extra tokens for each group member. At the end of each round, subjects received information about their payoffs. Tokens were converted to euros using an exchange rate of 1000 points per euro. Subjects, mainly undergraduate students of economics, earned 12.25 euros on average for sessions lasting about 50 minutes. The experiment took place in May 2006 in the Experimental Economics Laboratory of the University



of Milan Bicocca and was computerized using the z-Tree software [29].

The features of anonymity and random rematching narrow the relevance of some “traditional” behavioral hypotheses used to explain subjects’ overcontribution. For instance, they preclude subjects’ possibility to reciprocate (un)kind contributions of group members [30]. Moreover, under these conditions, subjects with preferences for equality cannot make compensating contributions to reduce (dis)advantageous inequality [31, 32]. Rather, the hypothesis of preference for (group) efficiency as a particular form of warm-glow [8, 9] appears as a more plausible justification.

### 3. Theoretical Predictions and Estimation Procedure

Consider a finite set of subjects  $P = \{1, 2, \dots, p\}$ . In a generic round, subject  $i \in P$ , with endowment  $w_i \in N^+$ , contributes  $g_i$  to the group account, with  $g_i \in N^+$  and  $0 \leq g_i \leq w_i$ . The monetary payoff of subject  $i$  who contributes  $g_i$  in a round is given by

$$\pi_i(w_i, g_i) = 2(w_i - g_i) + 120 + g_i + G_{-i}, \quad (1)$$

where  $G_{-i}$  is the sum of the contributions of group members other than  $i$  in that round. Given (1), if subjects’ utility only depends on the monetary payoff, zero contributions are the unique Nash equilibrium of each round. In order to explain the positive contributions observed in their experiment, Corazzini et al. [2] assume that subjects suffer psychological costs if they contribute less than what is optimal for the group. In particular, psychological costs are introduced as a convex quadratic function of the difference between a subject’s endowment (i.e., the social optimum) and her contribution. In the VCM, player  $i$ ’s (psychological) utility function is given by

$$u_i(w_i, g_i, \alpha_i) = \pi_i(w_i, g_i) - \alpha_i \frac{(w_i - g_i)^2}{w_i}, \quad (2)$$

where  $\alpha_i$  is a nonnegative and finite parameter measuring the weight attached to the psychological costs,  $(w_i - g_i)^2/w_i$ , in the utility function. Notice that psychological costs are increasing in the difference between a subject’s endowment and her contribution. Under these assumptions, in each round, there is a unique Nash equilibrium in which individual  $i$  contributes:

$$g_i^{\text{NE}} = \frac{2\alpha_i - 1}{2\alpha_i} w_i. \quad (3)$$

The higher the value of  $\alpha_i$ , the higher the equilibrium contribution of subject  $i$ . The average relative contribution,  $g_i/w_i$ , observed in the VCM sessions is 22%, which implies  $\alpha = 0.64$ .

Following McKelvey and Palfrey [1], we introduce noisy decision-making and consider a Logit Quantal Response extension of (2). In particular, we assume subjects choose their contributions randomly according to a logistic quantal

response function. Namely, for a given endowment,  $w_i$ , and contributions of the other group members,  $G_{-i}$ , the probability that subject  $i$  contributes  $g_i$  is given by

$$q_i(w_i, g_i, \alpha_i, \mu) = \frac{\exp\{u_i(w_i, g_i, \alpha_i)/\mu\}}{\sum_{g_j=0}^{w_i} \exp\{u_i(w_i, g_j, \alpha_i)/\mu\}}, \quad (4)$$

where  $\mu \in \mathfrak{R}_+$  is a noise parameter reflecting a subject’s capacity of noticing differences in expected payoffs.

Therefore each subject  $i$  is associated with a  $w_i$ -dimensional vector  $\underline{q}_i(w_i, \underline{g}_i, \alpha_i, \mu)$  containing a value of  $q_i(w_i, g_i, \alpha_i, \mu)$  for each possible contribution level  $g_i \in \underline{g}_i \equiv \{0, \dots, w_i\}$ . Let  $\{\underline{q}_i(w_i, \underline{g}_i, \alpha_i, \mu)\}_{i \in P}$  be the system including  $\underline{q}_i(w_i, \underline{g}_i, \alpha_i, \mu)$ ,  $\forall i \in P$ . Notice that since others’ contribution,  $G_{-i}$ , enters the r.h.s of the system, others’  $q_i$  will also enter the r.h.s. A fixed point of  $\{\underline{q}_i(w_i, \underline{g}_i, \alpha_i, \mu)\}_{i \in P}$  is, hence, a quantal response equilibrium (QRE),  $\{\underline{q}_i^{\text{QRE}}(w_i, \underline{g}_i, \alpha_i, \mu)\}_{i \in P}$ .

In equilibrium, the noise parameter  $\mu$  reflects the dispersion of subjects’ contributions around the Nash prediction expressed by (3). The higher the  $\mu$ , the higher the dispersion of contributions. As  $\mu$  tends to infinity, contributions are randomly drawn from a uniform distribution defined over  $[0, w_i]$ . On the other hand, if  $\mu$  is equal to 0, the equilibrium contribution collapses to the Nash equilibrium. (more specifically, for each subject  $i$  equilibrium contributions converge to  $q_i(w_i, g_i^{\text{NE}}, \alpha_i, 0) = 1$  and  $q_i(w_i, g_i, \alpha_i, 0) = 0$ ,  $\forall g_i \neq g_i^{\text{NE}}$ .)

In this framework, we use data from Corazzini et al. [2] to estimate  $\alpha$  and  $\mu$ , jointly. We proceed as follows. Our initial analysis is conducted by using all rounds ( $n = 20$ ) and assuming the population to be homogeneous in both  $\alpha$  and  $\mu$ . This gives us a benchmark that can be directly compared with the results reported by Corazzini et al. [2]. In our estimation procedure, we use a likelihood function that assumes each subject’s contributions to be drawn from a multinomial distribution. That is,

$$L_i(w_i, \underline{g}_i, \alpha, \mu) = \frac{n!}{\prod_{g_j=0}^{w_i} n(g_j)!} \prod_{g_k=0}^{w_i} q_i^{\text{QRE}}(w_i, g_k, \alpha, \mu)^{n(g_k)}, \quad (5)$$

where  $n(g_j)$  is the number of times that subject  $i$  contributed  $g_j$  over the  $n$  rounds of the experiment, and similarly for  $n(g_k)$ . The contribution of each person to the log-likelihood is the log of expression (5). The Maximum Likelihood procedure consists of finding the nonnegative values of  $\mu$  and  $\alpha$  (and corresponding QRE) that maximize the summation of the log-likelihood function evaluated at the experimental data. In other words, we calculate the multinomial probability of the observed data by restricting the theoretical probabilities to QRE probabilities only.

We then extend our analysis to allow for cross-subject heterogeneity. In particular, we generalize the QRE model above by assuming the population to be partitioned into  $S$



TABLE 1: Homogeneous population (all rounds).

	Data	CFS	(1) $\bar{\mu}, \bar{\alpha}$	(2) $\mu, \bar{\alpha}$	(3) $\mu, \alpha$
$\mu$	—	—	1	21.83 [19.69; 24.34]	41.59 [39.11; 44.34]
$\alpha$	—	0.64	0.64	0.64	0 [0; 0.01]
(Predicted) avg. contributions					
Overall endowments	37.91	39.38	39.41	60.24	37.91
$w_i = 120$	34.02	26.25	26.34	44.84	34.12
$w_i = 160$	24.53	35.00	35.03	55.68	37.67
$w_i = 200$	47.50	43.75	43.76	65.57	39.48
$w_i = 240$	45.57	52.50	52.50	74.86	40.36
$\log ll$			-8713.95	-3483.79	-3170.69
Obs.	960	960	960	960	960

This table reports average contributions as well as estimates and predictions from various specifications of the model based on the efficiency concerns assumption using all 20 rounds of the experiment. CFS refers to the specification not accounting for noise in subjects' contributions while (1), (2), and (3) are Logit Quantal Response extensions of the model. In (1)  $\alpha$  and  $\mu$  are constrained to 0.64 and 1, respectively. In (2), the value on  $\alpha$  is set to 0.64, while  $\mu$  is estimated through (5). Finally, (3) refers to the unconstrained model in which both  $\alpha$  and  $\mu$  are estimated through (5). The table also reports, for each specification, the corresponding log-likelihood. Confidence intervals are computed using an inversion of the likelihood-ratio statistic, at the 0.01 level, subject to parameter constraints.

subgroups that are characterized by the same  $\mu$  but different  $\alpha$ . In this case, the likelihood function becomes

$$L_i \left( w_i, \underline{g}_i, \alpha_1, \alpha_2, \dots, \alpha_S, \gamma_1, \gamma_2, \dots, \gamma_S, \mu \right) = \sum_{s=1}^S \gamma_s \frac{n!}{\prod_{j=0}^{w_i} n(g_j)!} \prod_{g_k=0}^{w_i} q_i^{\text{QRE}}(w_i, g_k, \alpha_s, \mu)^{n(g_k)}, \quad (6)$$

where  $\gamma_1, \gamma_2, \dots, \gamma_S$ , with  $\sum_{s=1}^S \gamma_s = 1$ , are the probabilities for agent  $i$  belonging to the subgroup associated with  $\alpha_1, \alpha_2, \dots, \alpha_S$ , respectively. This allows us to estimate the value of  $\mu$  for the whole population, the value of  $\alpha_1, \alpha_2, \dots, \alpha_S$  for the  $S$  subgroups, and the corresponding probabilities,  $\gamma_1, \gamma_2, \dots, \gamma_S$ . For identification purposes we impose that  $\alpha_s \leq \alpha_{s+1}$ . The introduction of one group at a time accompanied by a corresponding likelihood-ratio test allows us to determine the number of  $\alpha$ -groups that can be statistically identified from the original data. In the following statistical analysis, estimates account for potential dependency of subject's contributions across rounds. Confidence intervals at the 0.01 level are provided using the inversion of the likelihood-ratio statistic, subject to parameter constraints, in line with Cook and Weisberg [33], Cox and Hinkley [34], and Murphy [35].

## 4. Results

Using data from the 20 rounds of the experiment, Table 1 reports (i) average contributions (by both endowment type and overall) observed in the experiment, (ii) average contributions as predicted by the model not accounting for noise in subjects' contributions, and (iii) estimates as well as average contributions from different parameterizations of the Logit Quantal Response extension of the model. In particular, specification (1) refers to a version of the model in which both  $\alpha$  and  $\mu$  are constrained to be equal to benchmark values

based on Corazzini et al. [2]. Under this parameterization,  $\alpha$  is fixed to the value computed by calibrating (3) on the original experimental data, 0.64, while  $\mu$  is constrained to 1. (Table 4 shows the Maximum Likelihood estimation value of  $\alpha$  when we vary  $\mu$ . It is possible to see that for a large range of values of  $\mu$  this value is close to 0.64. We choose  $\mu = 1$  as a sufficiently low value in which the estimated  $\alpha$  is close to 0.64 and thus provide a noisy version of the base model which can be used for statistical tests.)

As shown by the table, specification (1) closely replicates predictions of the original model presented by Corazzini et al. [2] not accounting for noise in subjects' contributions. In specification (2),  $\alpha$  is fixed to 0.64, while  $\mu$  is estimated by using (4). The value of  $\mu$  increases substantially with respect to the benchmark value used in specification (1). A likelihood-ratio test strongly rejects specification (1) that imposes restrictions on the values of both  $\alpha$  and  $\mu$  in favor of specification (2) in which  $\mu$  can freely vary on  $\mathfrak{R}_+$  (LR = 10460.33;  $\Pr\{\chi^2(1) > \text{LR}\} < 0.01$ ). However, if we compare the predicted average contributions of the two specifications, we find that specification (1) better approximates the original experimental data. This is because a higher value of the noise parameter spread the distributions of contributions around the mean. Therefore even with mean contributions further from the data (induced by the fixed value of  $\alpha$ ) the spread induced by the noise parameter in specification (2) produces a better fit. This highlights the importance of taking into account not only the average (point) predictions but also the spread around it. It also suggests that allowing  $\alpha$  to vary can improve fit.

In specification (3),  $\alpha$  and  $\mu$  are jointly estimated using (5), subject to  $\alpha \geq 0$ . If both parameters can freely vary over  $\mathfrak{R}_+$ ,  $\alpha$  reduces to zero and  $\mu$  reaches a value that is higher than what was obtained in specification (2). As confirmed by a likelihood-ratio test, specification (3) fits the experimental data better than both specification (1)

TABLE 2: Homogeneous population (last 5 rounds).

	Data	CFS	(1) $\bar{\mu}, \bar{\alpha}$	(2) $\mu, \bar{\alpha}$	(3) $\mu, \alpha$
$\mu$		—	1	11.63 [9.80; 13.95]	26.91 [24.14; 30.17]
$\alpha$		0.59	0.59	0.59	0 [0; 0.03]
(Predicted) avg. contributions					
Overall endowments	25.94	26.39	26.91	44.78	25.94
$w_i = 120$	21.13	17.59	18.44	34.55	25.05
$w_i = 160$	17.63	23.46	24.04	41.68	26.01
$w_i = 200$	35.28	29.32	29.71	48.30	26.30
$w_i = 240$	29.70	35.19	35.45	54.60	26.39
$\log ll$			-1675.03	-987.54	-885.62
Obs.	240	240	240	240	240

This table reports average contributions as well as estimates and predictions from various specifications of the model based on the efficiency concerns assumption using the last 5 rounds of the experiment only. The same remarks as in Table 1 apply.

TABLE 3: Heterogeneous subjects (all and last 5 rounds).

	$\mu, \alpha_1, \text{ and } \alpha_2 \ (n = 20)$	$\mu, \alpha_1, \text{ and } \alpha_2 \ (n = 5)$	$\mu, \alpha_1, \alpha_2, \text{ and } \alpha_3 \ (n = 20)$	$\mu, \alpha_1, \alpha_2, \text{ and } \alpha_3 \ (n = 5)$
$\mu$	28.50 [25.88; 31.26]	15.07 [12.90; 17.64]	22.14 [20.56; 23.95]	14.25 [12.04; 16.85]
$\alpha_1$	0 [0; 0.01]	0 [0; 0.02]	0 [0; 0.01]	0 [0; 0.02]
$\alpha_2$	0.53 [0.46; 0.60]	0.54 [0.47; 0.61]	0.43 [0.39; 0.46]	0.48 [0.40; 0.56]
$\alpha_3$			1.04 [0.92; 1.16]	0.76 [0.53; 1.01]
$\gamma_1$	0.66 [0.53; 0.78]	0.63 [0.50; 0.75]	0.41 [0.33; 0.46]	0.59 [0.49; 0.64]
$\gamma_2$			0.50 [0.43; 0.55]	0.34 [0.23; 0.40]
(Predicted) avg. contributions				
Overall endowments	37.15	25.57	38.51	25.72
$w_i = 120$	32.06	22.17	31.97	22.07
$w_i = 160$	36.04	24.63	36.78	24.68
$w_i = 200$	39.01	26.77	40.83	26.99
$w_i = 240$	41.48	28.73	44.45	29.13
$\log ll$	-3112.06	-865.75	-3083.35	-865.16
Obs.	960	240	960	240

This table reports estimates and predictions from two specifications of the model with efficiency concerns accounting for cross subject heterogeneity in the value of  $\alpha$ . The analysis is conducted both by including all experimental rounds and by focusing on the last five repetitions only. Parameters are estimated through (6). Given the linear restriction  $\sum_{s=1}^S \gamma_s = 1$ , we only report estimates of  $\gamma_1, \gamma_2, \dots, \gamma_{S-1}$ . Confidence intervals are computed using an inversion of the likelihood-ratio statistic, at the 0.01 level, subject to parameter constraints.

(LR = 11086.54;  $\Pr\{\chi^2(2) > LR\} < 0.01$ ) and specification (2) (LR = 626.21;  $\Pr\{\chi^2(1) > LR\} < 0.01$ ). Thus, under the maintained assumption of homogeneity, our estimates suggest that contributions are better explained by randomness in subjects' behavior rather than by concerns for efficiency.

In order to control for learning effects, we replicate our analysis using the last five rounds only.

Consistent with a learning argument, in both specifications (2) and (3), the values of  $\mu$  are substantially lower than the corresponding estimates in Table 1. Thus, repetition reduces randomness in subjects' contributions. The main results presented above are confirmed by our analysis on the last five periods. Looking at specification (3), in the model with no constraints on the parameters, the estimated value of  $\alpha$  again drops to 0. Also, according to a likelihood-ratio test, specification (3) explains the data better than both

specifications (1) (LR = 1578.83;  $\Pr\{\chi^2(2) > LR\} < 0.01$ ) and (2) (LR = 203.85;  $\Pr\{\chi^2(1) > LR\} < 0.01$ ).

These results seem to reject the preference for (group) efficiency hypothesis in favor of pure randomness in subjects' contributions. However, a different picture emerges when we allow for cross-subject heterogeneity. In Table 3 we drop the assumed homogeneity. We consider two models with heterogeneous subjects: the first assumes the population to be partitioned into two subgroups ( $S = 2$ ) and the second into three subgroups ( $S = 3$ ). (We have also estimated a model with  $S = 4$ . However, adding a fourth subgroup does not significantly improve the goodness of fit of the model compared to the specification with  $S = 3$ . In particular, with  $S = 4$ , the point estimates for the model with all periods are  $\mu = 21.81$ ,  $\alpha_1 = 0$ ,  $\alpha_2 = 0.38$ ,  $\alpha_3 = 0.61$ ,  $\alpha_4 = 1.04$ ,  $\gamma_1 = 0.39$ ,  $\gamma_2 = 0.42$ , and  $\gamma_3 = 0.09$ .) As before, we conduct

TABLE 4

$\mu$	$\alpha$	Log-likelihood
1000.00	0	-3637.64
500.00	0	-3591.79
333.33	0	-3548.7
250.00	0	-3508.34
200.00	0	-3470.67
166.67	0	-3435.64
142.86	0	-3403.19
125.00	0	-3373.25
111.11	0	-3345.76
100.00	0	-3320.64
90.91	0	-3297.8
83.33	0	-3277.17
76.92	0	-3258.65
71.43	0	-3242.16
66.67	0	-3227.61
62.50	0	-3214.93
58.82	0	-3204.01
55.56	0	-3194.79
52.63	0	-3187.18
50.00	0	-3181.1
40.00	0	-3171.22
30.30	0.14	-3192.52
20.00	0.33	-3247.22
10.00	0.50	-3444.42
9.09	0.52	-3488.26
8.00	0.53	-3555.57
7.04	0.55	-3634.45
5.99	0.56	-3753.67
5.00	0.57	-3916.89
4.00	0.58	-4173.35
3.00	0.59	-4615.92
2.00	0.60	-5547.35
1.00	0.61	-8506.13
0.90	0.61	-9181.67
0.80	0.61	-10032.07
0.70	0.61	-11133.38
0.60	0.61	-12612.63
0.50	0.61	-14699.13
0.40	0.61	-17852.64

This table reports Maximum Likelihood estimates of  $\alpha$  for selected values of  $\mu$  (see (5)). The last column reports the corresponding log-likelihood value.

our analysis both by including all rounds of the experiment and by focusing on the last five repetitions only.

We find strong evidence in favor of subjects' heterogeneity. Focusing on the analysis over all rounds, according to the model with two subgroups, a subject is associated with  $\alpha_1 = 0$  with probability 0.66 and with  $\alpha_2 = 0.53$  with probability 0.34. Results are even sharper in the model with three subgroups: in this case  $\alpha_1 = 0$  and the two other  $\alpha$ -parameters are strictly positive:  $\alpha_2 = 0.43$  and  $\alpha_3 = 1.04$ . Subjects are associated with these values with probabilities

0.41, 0.50, and 0.09, respectively. Thus, in the more parsimonious model, the majority of subjects contribute in a way that is compatible with the preference for (group) efficiency hypothesis. These proportions are in line with findings of previous studies [18, 21, 22] in which, aside from confusion, social preferences explain the behavior of about half of the experimental population.

Allowing for heterogeneity across subjects reduces the estimated randomness in contributions: the value of  $\mu$  reduces from 41.59 in specification (3) of the model with homogeneous population to 28.50 and 22.14 in the model with two and three subgroups, respectively. According to a likelihood-ratio test, both the models with  $S = 2$  and  $S = 3$  fit the data better than the (unconstrained) specification of the model with homogeneous subjects (for the model with  $S = 2$ ,  $LR = 117.25$ ;  $\Pr\{\chi^2(2) > LR\} < 0.01$ , whereas for the model with  $S = 3$ ,  $LR = 174.66$ ;  $\Pr\{\chi^2(4) > LR\} < 0.01$ ). Moreover, adding an additional subgroup to the model, with  $S = 2$ , significantly increases the goodness of fit of the specification ( $LR = 57.42$ ;  $\Pr\{\chi^2(2) > LR\} < 0.01$ ). As before, all these results remain qualitatively unchanged when we control for learning processes and we focus on the last 5 experimental rounds.

In order to check for the robustness of our results in Table 3, we have also estimated additional specifications accounting for heterogeneity in both concerns for (group) efficiency and noise in subjects' behavior. Although the log-likelihood of the model with both sources of heterogeneity significantly improves in statistical terms, the estimated values of the  $\alpha$ -parameters remain qualitatively the same as those reported in the third column of Table 3.

## 5. Conclusions

Is overcontribution in linear public good experiments explained by subjects' preference for (group) efficiency or, rather, does it simply reflect their natural attitude to make errors? In order to answer this fundamental question, we estimate a quantal response equilibrium model in which, in choosing their contributions, subjects are influenced by both a genuine concern for (group) efficiency and a random noise in their behavior.

In line with other studies, we find that both concerns for (group) efficiency and noise in behavior play an important role in determining subjects' contributions. However, assessing which of these two behavioral hypotheses is more relevant in explaining contributions strongly depends on the degree of cross-subject heterogeneity admitted by the model. Indeed, by estimating a model with homogeneous subjects, the parameter capturing concerns for (group) efficiency vanishes while noise in behavior entirely accounts for overcontribution. A different picture emerges when we allow the subjects to be heterogeneous in their concerns for efficiency. By estimating a model in which the population is partitioned into three subgroups that differ in the degree of concerns for efficiency, we find that most of the subjects contribute in a way that is compatible with the preference for (group) efficiency hypothesis. A formal likelihood-ratio

test confirms the supremacy of the QRE model with three subgroups over the other specifications.

Previous studies [18–23] tried to disentangle the effects of noise from other-regarding preferences by mainly manipulating the experimental design. Our approach adds a theoretical foundation in the form of an equilibrium analysis. In contrast to studies which focus mostly on (direct) altruism, we follow Corazzini et al. [2] and allow for preference for efficiency. Our results are in line with the literature in the sense that we also conclude that a combination of noise and social concerns plays a role. Our results, however, are directly supported by a sound theoretical framework proven valid in similar settings (e.g., [36]).

Recent studies [37, 38] have emphasized the importance of admitting heterogeneity in social preferences in order to better explain experimental evidence. In this paper we show that neglecting heterogeneity in subjects' social preferences may lead to erroneous conclusions on the relative importance of the love for (group) efficiency hypothesis with respect to the confusion argument. Indeed, as revealed by our analysis, the coupling of cross-subject heterogeneity in concerns for (group) efficiency with noise in the decision process seems to be the relevant connection to better explain subjects' contributions.

## Appendix

Table 4 shows the Maximum Likelihood value of  $\alpha$  and the log-likelihood according to (5) as  $\mu$  decreases from 1000 to 0.4. As shown by the table, for high values of  $\mu$ , the estimated value of  $\alpha$  is 0. When  $\mu$  is equal to 10, the estimated value of  $\alpha$  is 0.50. Moreover, for  $\mu$  lower than 2.00, the estimated value of  $\alpha$  is 0.61. For the specification tests presented in Section 4, we set  $\mu = 1$ . This is a sufficiently low value of  $\mu$  in order to generate a noisy version of the base model. Two arguments indicate why this choice is valid. First, for a range of values including  $\mu = 1$ , the estimated  $\alpha$  is stable. Moreover, since the log-likelihood of a model with  $\alpha = 0.61$  and  $\mu = 1$  is higher than that corresponding to a model with  $\mu = 0.4$  (and similarly for  $\alpha = 0.64$ ), the choice of any  $\mu$  lower than 1 for the benchmark value would only reinforce the results of Section 4. More specifically, both likelihood-ratio statistics comparing specifications (1) with specifications (2) and (3) of Tables 1 and 2 would increase.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors thank Arthur Schram, Jens Grosser, and participants to the 2010 Credexa Meeting at the University of Amsterdam, 2011 IMEBE in Barcelona, and 2011 Annual Meeting of the Royal Economic Society in London for useful comments and suggestions.

## References

- [1] R. D. McKelvey and T. R. Palfrey, "Quantal response equilibria for normal form games," *Games and Economic Behavior*, vol. 10, no. 1, pp. 6–38, 1995.
- [2] L. Corazzini, M. Faravelli, and L. Stanca, "A prize to give for: an experiment on public good funding mechanisms," *The Economic Journal*, vol. 120, no. 547, pp. 944–967, 2010.
- [3] R. Sugden, "Reciprocity: the supply of public goods through voluntary contributions," *The Economic Journal*, vol. 94, no. 376, pp. 772–787, 1984.
- [4] H. Hollander, "A social exchange approach to voluntary contribution," *The American Economic Review*, vol. 80, no. 5, pp. 1157–1167, 1990.
- [5] A. Falk and U. Fischbacher, "A theory of reciprocity," *Games and Economic Behavior*, vol. 54, no. 2, pp. 293–315, 2006.
- [6] U. Fischbacher, S. Gächter, and E. Fehr, "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, vol. 71, no. 3, pp. 397–404, 2001.
- [7] D. K. Levine, "Modeling altruism and spitefulness in experiments," *Review of Economic Dynamics*, vol. 1, no. 3, pp. 593–622, 1998.
- [8] J. Andreoni, "Giving with impure altruism: applications to charity and ricardian equivalence," *Journal of Political Economy*, vol. 97, no. 6, pp. 1447–1458, 1989.
- [9] J. Andreoni, "Impure altruism and donations to public goods: a theory of warm-glow giving," *Economic Journal*, vol. 100, no. 401, pp. 464–477, 1990.
- [10] J. J. Laffont, "Macroeconomic constraints, economic efficiency and ethics: an introduction to Kantian economics," *Economica*, vol. 42, no. 168, pp. 430–437, 1975.
- [11] M. Bordinon, "Was Kant right? Voluntary provision of public goods under the principle of unconditional commitment," *Economic Notes*, vol. 3, pp. 342–372, 1990.
- [12] M. Bernasconi, L. Corazzini, and A. Marenzi, "Expressive obligations in public good games: crowding-in and crowding-out effects," Working Papers, University of Venice "Ca' Foscari", Department of Economics, 2010.
- [13] M. Bacharach, *Beyond Individual Choice: Teams and Frames in Game Theory*, Edited by N. Gold and R. Sugden, Princeton University Press, 2006.
- [14] R. Sugden, "The logic of team reasoning," *Philosophical Explorations*, vol. 6, no. 3, pp. 165–181, 2003.
- [15] R. Cookson, "Framing effects in public goods experiments," *Experimental Economics*, vol. 3, no. 1, pp. 55–79, 2000.
- [16] G. Charness and M. Rabin, "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, vol. 117, no. 3, pp. 817–869, 2002.
- [17] D. Engelmann and M. Strobel, "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments," *American Economic Review*, vol. 94, no. 4, pp. 857–869, 2004.
- [18] J. Andreoni, "Cooperation in public-goods experiments: kindness or confusion?" *The American Economic Review*, vol. 85, no. 4, pp. 891–904, 1995.
- [19] T. R. Palfrey and J. E. Prisbrey, "Altruism, reputation and noise in linear public goods experiments," *Journal of Public Economics*, vol. 61, no. 3, pp. 409–427, 1996.
- [20] T. R. Palfrey and J. E. Prisbrey, "Anomalous behavior in public goods experiments: how much and why?" *The American Economic Review*, vol. 87, no. 5, pp. 829–846, 1997.

- [21] J. Brandts and A. Schram, "Cooperation and noise in public goods experiments: applying the contribution function approach," *Journal of Public Economics*, vol. 79, no. 2, pp. 399–427, 2001.
- [22] D. Houser and R. Kurzban, "Revisiting kindness and confusion in public goods experiments," *American Economic Review*, vol. 92, no. 4, pp. 1062–1069, 2002.
- [23] J. K. Goeree, C. A. Holt, and S. K. Laury, "Private costs and public benefits: unraveling the effects of altruism and noisy behavior," *Journal of Public Economics*, vol. 83, no. 2, pp. 255–276, 2002.
- [24] D. O. Stahl and P. W. Wilson, "On players' models of other players: theory and experimental evidence," *Games and Economic Behavior*, vol. 10, no. 1, pp. 218–254, 1995.
- [25] T.-H. Ho, C. Cambrer, and K. Weigelt, "Iterated dominance and iterated best response in experimental 'p-Beauty Contests,'" *American Economic Review*, vol. 88, no. 4, pp. 947–969, 1998.
- [26] D. O. Stahl and E. Haruvy, "Level- $n$  bounded rationality in two-player two-stage games," *Journal of Economic Behavior & Organization*, vol. 65, no. 1, pp. 41–61, 2008.
- [27] I. Erev and A. E. Roth, "Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria," *American Economic Review*, vol. 88, no. 4, pp. 848–881, 1998.
- [28] J. Andreoni, "Why free ride? Strategies and learning in public goods experiments," *Journal of Public Economics*, vol. 37, no. 3, pp. 291–304, 1988.
- [29] U. Fischbacher, "Z-Tree: zurich toolbox for ready-made economic experiments," *Experimental Economics*, vol. 10, no. 2, pp. 171–178, 2007.
- [30] M. Rabin, "Incorporating fairness into game theory and economics," *The American Economic Review*, vol. 83, no. 5, pp. 1281–1302, 1993.
- [31] E. Fehr and K. M. Schmidt, "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.
- [32] G. E. Bolton and A. Ockenfels, "ERC: a theory of equity, reciprocity, and competition," *American Economic Review*, vol. 90, no. 1, pp. 166–193, 2000.
- [33] R. D. Cook and S. Weisberg, "Confidence curves in nonlinear regression," *Journal of the American Statistical Association*, vol. 85, no. 410, pp. 544–551, 1990.
- [34] D. R. Cox and D. V. Hinkley, *Theoretical Statistics*, Chapman and Hall, London, UK, 1974.
- [35] S. A. Murphy, "Likelihood ratio-based confidence intervals in survival analysis," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1399–1405, 1995.
- [36] J. K. Goeree and C. A. Holt, "An explanation of anomalous behavior in models of political participation," *American Political Science Review*, vol. 99, no. 2, pp. 201–213, 2005.
- [37] U. Fischbacher and S. Gächter, "Heterogeneous social preferences and the dynamics of free riding in public goods," CeDEx Discussion Paper 2006-01, 2006.
- [38] M. Erlei, "Heterogeneous social preferences," *Journal of Economic Behavior and Organization*, vol. 65, no. 3-4, pp. 436–457, 2008.



## Research Article

# Competition, Income Distribution, and the Middle Class: An Experimental Study

**Bernhard Kittel,<sup>1</sup> Fabian Paetzel,<sup>2</sup> and Stefan Traub<sup>2</sup>**

<sup>1</sup>*Department of Economic Sociology and FOR 2104, University of Vienna, 1090 Vienna, Austria*

<sup>2</sup>*Department of Economics and FOR 2104, Helmut-Schmidt-University, 22043 Hamburg, Germany*

Correspondence should be addressed to Bernhard Kittel; [bernhard.kittel@univie.ac.at](mailto:bernhard.kittel@univie.ac.at)

Received 5 December 2014; Revised 28 January 2015; Accepted 17 February 2015

Academic Editor: Klarita Gërxhani

Copyright © 2015 Bernhard Kittel et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We study the effect of competition on income distribution by means of a two-stage experiment. Heterogeneous endowments are earned in a contest, followed by a surplus-sharing task. The experimental test confirms our initial hypothesis that the existence of a middle class is as effective as institutional hurdles in limiting the power of the less able in order to protect the more able players from being expropriated. Furthermore, majoritarian voting with a middle class involves fewer bargaining impasses than granting veto rights to the more able players and, therefore, is more efficient.

## 1. Introduction

In recent years, public awareness of an increasingly lopsided distribution of income and wealth in Western countries has strongly increased [1]. Most OECD countries have witnessed growing inequality over the past 20 years [2]. In particular, the gap between the bottom and the top deciles of the household income distribution has risen dramatically. The decile ratio currently amounts to about 1:15 in the US and 1:9 in the OECD-34, and even in a Nordic welfare state like Sweden it is close to 1:6. Given the fact that the typical OECD income distribution is right-skewed [3], it does not come as a surprise that income tax progression is much higher in the US than in Sweden [4].

Observers diagnose the formation, as well as deliberate establishment, of a winner-take-all society in which the middle class is gradually eroded [5–7]. A few superstars, which may (or may not) be the most able practitioners in a particular area, receive all the profit while the others' efforts are in vain (e.g., [8–10]). Some commentators expect this development to result in increasing societal division, distributive struggle, and violent conflict [11].

On the other hand, the fall of communist regimes in 1989 has demonstrated the crippling effects of excessive

egalitarianism on economic efficiency and growth [12, 13]. Societies, or groups, thus need to strike a balance between both excessive inequality and excessive equality. In principle, either an institutional or a structural condition is sufficient to protect a society against excesses from both sides. On the one hand, institutional rules of collective decision-making giving veto power to all stakeholders [14, 15] enforce a consensual decision and thereby make the distributional struggle an issue of explicit negotiation. On the other hand, the existence of a neutral but opportunistic middle class serves as a buffer between the upper and the lower classes because they fear expropriation from both sides and will thus side with the weaker group in case of conflict [16–18]. Moreover, not being the main target of redistribution, members of the middle class will let their decisions be guided more by moral values such as justice norms than the other two socioeconomic groups [19–21].

The two conditions, however, differ with respect to their decision efficiency. Although the unanimity rule minimizes external costs due to the inclusion of all members of the society, it also allows every member to hold up the decision until the outcome satisfactorily represents its particular interests. For this reason, this arrangement is almost inexistent at

the level of nation states and is notorious for its detrimental effects in the security council of the United Nations. Because of its crippling implications, the unanimity rule has been replaced by qualified majority for many policy areas in the European Council by the Lisbon Treaty in 2007. In contrast, the position-based interests of the middle class limit the costs burdened upon the minority because its members will shift sides as soon as demands on the minority become expropriatory. The existence of a middle class thus quasiautomatically corrects for both excessive inequality and equality without generating the hold-up effect of an institutional rule.

In this paper, we experimentally test the composite hypothesis following from this argument that, in democratic and competitive organizations, both the existence of a middle class and institutional hurdles can protect the most able organization members from the demand of the least able for excessive income distribution policies (*Protection Hypothesis*); yet, setting up institutional hurdles is inferior to relying on a middle class as institutional hurdles involve efficiency losses due to bargaining impasses (*Efficiency Hypothesis*).

In our experiment, we directly test whether a neutral middle class is as effective as an institutional rule to avoid excessive expropriation or extremely low redistribution. The second goal of the paper is to check if the existence of a middle class leads to higher efficiency than institutional hurdles.

The experiment was composed of two consecutive tasks. During the first task, subjects took part in a multiple-prize rank-order contest that involved a simple cognitive ability task. The contest rewarded effort by assigning the first prize to the most able subject, the second prize to the second most able subject, and so on. Following Moldovanu and Sela [22], the contest ensured that more able subjects should expend more effort in order to win it. Hereby, we focus on the impact of the skewness of the prize scheme on effort and its interaction with ex-post income distribution. During the second task players first deliberated on sharing a surplus (using a group chat). As soon as the deliberation time had elapsed, they simultaneously and anonymously made their binding proposals for a distribution parameter. The proposal that achieved the quorum was played out and the surplus shares were added to the prize money won in the contest. If a group missed the quorum, the players received only their prizes.

We find that the *Protection Hypothesis* is clearly supported by our data. The existence of a neutral middle class is as effective as institutional hurdles. The *Efficiency Hypothesis* is also supported. Institutional hurdles come at the costs of significantly less efficiency in terms of higher default rates. The corroboration of both hypotheses allows us to give a clear recommendation for the rich. The rich should brace against the erosion of the middle class because the middle class is the efficient protection of the rich.

The rest of the paper is organized as follows. Section 2 links our paper to the literature. Section 3 introduces the theoretical framework of our experiment and derives testable hypotheses. Section 4 explains how the models were transferred into the experimental lab. Section 5 reports the results of the experiment. Section 6 discusses the results and concludes the paper.

## 2. Literature Review

If we assume that subjects solely follow self-interest and that self-interest is directly expressed in a social contract with a corresponding level of redistribution, we find ourself in the world of the median voter theorem [23]. Following this theory, an increasing skewness must result in larger pressure for equalization because the median voter's income declines and hence her interest in redistribution increases [23, 24]. Disregarding the possibility of revolutionary redistribution (tax rates exceeding one), the most extreme form of equalization within a society is to implement, by means of collective decision, a measure which results in equal payoffs. This result, however, is likely to generate socially inefficient outcomes because it undermines the willingness of the more able to expend effort [25].

Our paper contributes to the growing literature analyzing why the rich do not get expropriated. Besides the above-mentioned protection through institutional hurdles or the existence of a neutral middle class, another stream in the literature stresses different explanations why the poor median voter does not necessarily follow the median voter's prediction. The "prospects of upward mobility" hypothesis [26] brings forward the argument that some voters who have an income below the mean expect that their future income will be above the mean. The expectation of these voters works against votes for high levels of redistribution. Roemer [27] and Roemer and Lee [28] show that voting about redistribution is also affected by religion and race. The more fragmented the population is with regard to both dimensions, the lower is the willingness to redistribute. Alesina and Angeletos [13] and Fong [29] show that if the poor believe that the rich are rich because they had invested high efforts and have high abilities, the pressure for redistribution is low. According to extensive empirical and experimental literature, people tend to accept some amount of inequality if it can be attributed to individual effort and ability (e.g., [30–35]). Klor and Shayo [36] find that if the poor identify themselves with the nation (rich and poor) the poor do not want to harm members of the nation and therefore do not expropriate the rich. Moreover, Glaeser et al. [18] highlight that, beyond being protected from expropriation through institutions which are maintained to enforce low levels of redistribution, the better-off are able to manipulate institutions to their own profit. Our paper adds an additional explanation of why the poor cannot enforce high levels of redistribution. We argue that the middle class acts as an uninvolved spectator not affected by redistribution that does not have an interest in excesses from either side.

We contribute to a recent wave of research investigating voting on redistribution. This literature sheds some light on the voting decision by using experiments. The main methodological advantage of using experiments is that it allows us to detect how preferences affect the voting-decision. A stylized fact generated by this literature is that voters are willing to sacrifice their own payoffs in order to achieve a more equal payoff distribution, which has been explained by social preferences or social identity. Tyran and Sausgruber [37] reported supporting evidence for inequity theory [38]

but emphasized the importance of being pivotal for voting on redistribution. Similarly, Höchtl et al. [39] showed that for empirically plausible cases inequality-averse voters may not matter for redistribution outcomes. Moreover, they did not observe efficiency-loving voters. Bolton and Ockenfels [40] investigated the trade-off between equity and efficiency motives in a voting game with three voters. They found that twice as many voters were willing to give up their own payoff in favor of an equal distribution as compared to a more efficient but unequal distribution. Messer et al. [41] studied the impact of majority voting on the provision of a public good. They detected substantial concerns for efficiency in the subjects' behavior but found little support for inequality aversion and maximin preferences. Paetzel et al. [42] investigated whether voters would be willing to sacrifice their own payoff in order to implement not only an efficiency-increasing but also an inequality-increasing reform. This case can be thought of as the reverse of voting on efficiency-reducing but inequality-decreasing redistribution. They showed that efficiency preferences of potential reform-losers outweighed the inequality aversion of reform-gainers. Klor and Shayo [36] reported a trade-off between social-identity concerns and payoffs maximization: a subset of their sample systematically deviated from monetary payoff maximization towards a tax rate benefiting their group. Such deviations can neither be explained by efficiency concerns nor by inequality aversion.

Recently, Balafoutas et al. [43] reported on an experiment similar to ours (heterogenous initial endowments, quiz versus random assignment, majority vote, and neutral middle player), which generated a number of interesting results. For example, the players with the highest and the lowest endowments were mainly driven by material self-interest. Low-endowment players, however, signalled their willingness to cooperate by increasing their contributions if the redistribution rate was determined by majority vote.<sup>1</sup> Our experiment differs in various important respects: the size of the pie to be redistributed is exogenously given, our game involves a communication phase before voting on the tax, and we vary both the initial distribution of endowments and the institutional background in terms of the quorum.

A bit against the trend, Esarey et al. [44] and Durante et al. [45] still found redistribution to be strongly related to the self-interest of voters. Durante et al. [45] showed that support for redistribution vanishes if taxation was associated with costs and deadweight losses and if participants earned their incomes in a real effort task.

### 3. Theoretical Framework

In this section, we introduce the theoretical framework of our experiment in order to derive testable hypotheses. The experiment was composed of two consecutive tasks, namely, a contest and a demand game. The contest is explained in the first subsection. It was used for allocating subjects' initial endowments. The protection hypothesis presupposes that initial endowments adequately reflect subjects' abilities. We did not induce abilities, but subjects took part in a multiple-prize rank-order contest that involved a simple knowledge

test. Assuming that subjects knew their own abilities to solve the knowledge test and had a guess about the distribution of abilities within the sample, which was made up of their fellow students, the contest ensured that more able subjects would expend more effort in order to win it. Effort was rewarded by assigning the first prize to the most able subject, the second prize to the second most able subject, and so on. The theoretical setup for the contest was taken from Moldovanu and Sela [22]. Sheremeta et al. [46] showed the rank-order contest to generate higher effort than lottery and proportional contests.<sup>2</sup>

The demand game is explained in the second subsection. It was used for investigating the impact of different allocations of initial endowments and institutional setups on the acknowledgment of abilities and effort, respectively. Subjects first deliberated on sharing a surplus. As soon as the deliberation time had elapsed, they simultaneously and anonymously made their binding proposals for a distribution parameter. The proposal that achieved the quorum was played out and the surplus shares were added to the prize money won in the contest. If a group missed the quorum, the subjects received only their prizes. Such coordination and bargaining games with preplay communication (also known as cheap talk) have intensively been studied in the literature (for surveys see, e.g., [47–49]). The task determined the distribution parameter by vote and required a communication phase prior to finalizing the proposals such that group members could coordinate towards a distribution parameter capable of winning the necessary majority. In the context of our experiment, the results of Roth [50, 51] are highly relevant, who showed that cheap talk focuses subjects' attention on a small number of fairness norms in unstructured bargaining experiments (see also [52]). More specifically, subjects identify initial bargaining positions that have some special reason for being credible and these serve as focal points.

In the third subsection, we derive testable hypotheses from the theoretical models. We first identify three potential focal points of the demand game, namely, the egalitarian solution, equal sharing, and proportional sharing. Then, each experimental setup is assigned its most likely sharing rule by theoretical consideration. We also analyze the possibility of disagreement. Finally, we argue by backward induction whether and how the sharing rule enacted in the demand game could impact subjects' willingness to expend effort in the contest.

**3.1. The Contest.** Following Moldovanu and Sela [22], we consider a contest with three prizes  $\pi_j$ , where  $\pi_1 > \pi_2 \geq \pi_3 \geq 0$ . The total prize money  $\Pi = \sum_{j=1}^3 \pi_j$  has been exogenously fixed by the contest organizer. The set of contestants is given by  $\mathcal{K} = \{A, B, C\}$ . Individual effort is denoted by  $x_i$ , where we set  $x_A > x_B > x_C$ . The contest success function is perfectly discriminatory; that is, contestant A is awarded the first prize  $\pi_1$ , contestant B is awarded the second prize  $\pi_2$ , and contestant C is awarded the third prize  $\pi_3$ . Making an effort involves a cost  $c_i \gamma(x_i)$ .<sup>3</sup>  $\gamma$  is a strictly increasing function with  $\gamma(0) = 0$ .  $c_i > 0$  is an ability parameter. The lower is  $c_i$ , the higher is the ability of contestant  $i$  and the lower is the cost of effort.

Individual abilities are private information. They are independently drawn from the closed interval  $[m, 1]$ , where  $m > 0$ , and have a continuous distribution function  $F(c)$  which is common knowledge. To put it simply, we assume that ability follows a uniform distribution  $F(c) = ((1/(1 - m))c - m/(1 - m))I_{[m,1]}(c) + I_{(1,\infty)}(c)$ , where  $I$  is an indicator function taking a value of one if  $c$  is inside the stated interval and otherwise taking a value of zero. It can be shown (see Appendix C in [22]) that in the Nash equilibrium individual effort is given by

$$x(c) = \gamma^{-1} \left( \sum_{j=1}^3 \pi_j \int_c^1 -\frac{1}{a} F'_j(a) da \right), \quad (1)$$

where

$$F'_j(a) = \frac{(3-1)!}{(j-1)!(3-j)!} (1-F(a))^{3-j-1} F(a)^{j-2} \cdot F'(a) ((1-3)F(a) + (j-1)). \quad (2)$$

In order to interpret (1), one has to realize that a subject's probability of winning prize  $j$ ,  $1 \leq j \leq 3$ , given her ability  $a \in [m, 1]$ , can be represented by an order statistic

$$F_j(a) = \frac{(3-1)!}{(j-1)!(3-j)!} (1-F(a))^{3-j} F(a)^{j-1}. \quad (3)$$

This equation takes into account that subject  $j$  meets 2 competitors of whom  $3-j$  are less able than  $j$  and  $j-1$  are more able. Equation (2) is the first derivative of (3) with respect to ability. Hence, the payoff-maximizing equilibrium effort stated in (1) is a function of the weighted sum of the three prizes, and the weights differ from subject to subject according to their abilities  $c$ . Shaked and Shanthikumar [53] have shown that, if  $\gamma(c)$  is a strictly increasing function, effort is a monotonically decreasing function of the cost of ability  $c$  and is zero at  $c = 1$ . Hence, the most able subject will expend the highest effort and win the first prize, the medium subject will win the second prize, and the least able subject will expend the lowest effort (but not necessarily zero effort) and win the third prize. Since there is a unique relation between ability, effort, and prize, in the following we will call the most able subject the *A*-player, the medium able subject the *B*-player, and the least able subject the *C*-player.

The bidding function (1) is too complex to be solved analytically for the "structural parameters" of the prize scheme that are of interest for us, namely, its mean, variance, and skewness. Nevertheless, clear predictions concerning subjects' effort reactions can be derived for each parameter change by numerical simulation. In Appendix A, we report the results of simulating a contest with  $c = \{0.4, 0.6, 0.8\}$  and  $m = 0.2$ , assuming linear cost functions. The predicted outcome for each prize scheme applied in the experiment is shown in Table 4. The columns list groups of three hypothetical subjects with their abilities, expected winning probabilities, expected returns, efforts, ranks, prizes, costs, and net payoffs (recall that subjects did not know the abilities of the other subjects, but supposedly only the distribution

of ability). Table 5, also located in Appendix A, summarizes the information given in Table 4 by listing each prize scheme with its parametrization.<sup>4</sup> Additionally it states the mean and the coefficient of variation of predicted effort. As can be seen from the table, predicted effort doubles if the prize money is doubled and if the coefficient of variation of the prize scheme is doubled, and it is almost halved when the prize scheme becomes right-skewed. Furthermore, within-group effort variation increases with a right-skewed prize scheme.

**3.2. The Demand Game.** The second task is a variation of the Nash demand game [54]. After awarding the prizes to the contestants, the contest organizer asks the contestants to make an arbitrary number of proposals  $s = \{s_A, s_B, s_C\}$  for sharing a surplus  $S = \Pi$ ; that is, total prize money is doubled. However, the surplus is made available only if one of the proposals is supported by a qualified majority of the group of contestants. If none prevails, the contestants receive only their prizes  $\pi^D = \{\pi_A, \pi_B, \pi_C\}$  and exit the game (default option). If one of the proposals becomes accepted, it is made binding and the contestants receive their gross payoffs  $\pi^\alpha = \{\pi_A + s_A, \pi_B + s_B, \pi_C + s_C\}$ . The following restriction applies to the set of feasible distributions of  $S$ :

$$s_i^\alpha = \alpha \pi_i + (1 - \alpha) \left( \frac{2}{3} \Pi - \pi_i \right) \quad \forall i \in \{A, B, C\}, \quad 0 \leq \alpha \leq 1. \quad (4)$$

It is straightforward to show that any feasible agreement  $\alpha \in [0, 1]$  would be weakly preferred by each contestant over the default option if  $s_i \geq 0$  or  $0 \leq \pi_i \leq (2/3)\Pi$ . Hence, assuming that the latter condition holds, the default option can never be a Nash equilibrium. Conversely, any agreement is a Pareto-efficient Nash equilibrium. We refer to  $\alpha$  as the distribution parameter in the following.<sup>5</sup>

This unstructured bargaining protocol, in which subjects could first freely communicate and thereafter had to submit their binding proposals for  $\alpha$  simultaneously and anonymously, is ideal for studying the distribution of the surplus and its interaction with the contest for two reasons (for a critical survey of bargaining theories and their experimental tests, see, e.g., [55]): First, existing experimental evidence clearly suggests that subjects coordinate on a small number of focal points or fairness norms in such settings.<sup>6</sup> Distributions that could potentially serve as focal points in our experiment are addressed below. Second, the bargaining protocol allows for studying, apart from conflict of interest, the impact of different institutional settings in terms of voting rules on the frequency of disagreement as required for testing the *Efficiency Hypothesis*.<sup>7</sup> Sequential bargaining protocols such as Rubinstein's [56] and, in its finite version, Baron and Ferejohn's [57] are inappropriate for studying our research questions because they exhibit unique subgame perfect Nash equilibria, and they assign proposer and responder roles to the players.

**3.3. Focal Points and Hypotheses.** The demand game outlined in the previous subsection has three potential focal points which correspond to different distribution principles or



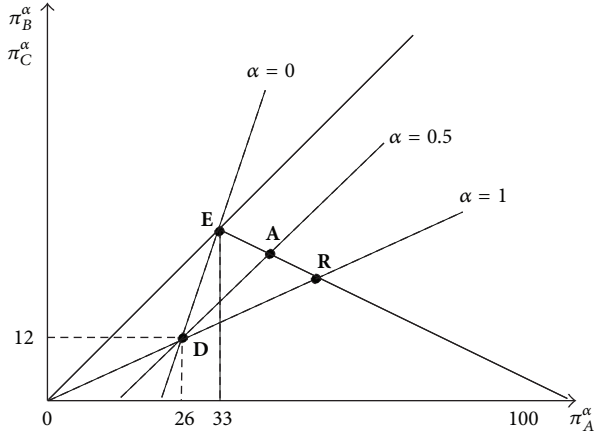


FIGURE 1: Sharing a surplus. The figure shows the gross payoffs  $\pi_i^\alpha$  of players  $i = A$  (horizontal axis) and  $B$  and  $C$  (vertical axis) for sharing a surplus of  $S = \Pi = 50$  with a right-skewed distribution of prizes  $\pi = \{26, 12, 12\}$ .  $\overline{ER}$  represents the bargaining set as well as the Pareto frontier.  $D$  marks the default option (no agreement). Three focal points  $E$  (egalitarian),  $A$  (equal sharing), and  $R$  (proportional sharing) are depicted together with their corresponding values of the distribution parameter  $\alpha$ .

fairness norms. Figure 1 gives a self-explaining example for a right-skewed distribution of prizes  $\pi^D = (26, 12, 12)$ . Total prize money and surplus are given by  $\Pi = S = 50$ . The “budget line”  $\overline{ER}$  represent the bargaining set as well as the Pareto frontier. With regard to  $\alpha$  we can distinguish three interesting special cases. First, for  $\alpha = 1$ , we obtain a surplus distribution  $s^R = \{s_A^R, s_B^R, s_C^R\}$ , where  $s_i^R = \pi_i \forall i \in \{A, B, C\}$ , that is proportional to the prize money earned in the contest and, thus, has a strong positive correlation with effort. In comparison to the distribution of prizes  $\pi^D = \{\pi_A, \pi_B, \pi_C\}$  (point  $D$ ) the resulting gross payoff distribution  $\pi^R = \{2\pi_A, 2\pi_B, 2\pi_C\}$  (point  $R$  in the figure) preserves the degree of inequality measured by any relative (scale invariant) inequality measure. The equity standard underlying  $\pi^R$  is well known as *proportional sharing* [58]. It targets the individual opportunity costs of the players [59].

Second, for  $\alpha = 0.5$ , surplus is shared equally  $s_i^A = (1/3)\Pi$ , proportional to the size of the surplus, and uncorrelated with effort. Compared with  $\pi^D$  the distribution  $\pi^A = \{\pi_A + (1/3)\Pi, \pi_B + (1/3)\Pi, \pi_C + (1/3)\Pi\}$  does not change any absolute (translation invariant) inequality measure. The respective focal point is marked by point  $A$  in Figure 1. Moulin [58, page 162] argued that “any symmetric solution concept would divide equally the surplus.” The *equal sharing* rule rewards cooperation rather than effort.

In the third case,  $\alpha = 0$  yields  $s_i^E = (2/3)\Pi - \pi_i$ , which is proportional to the gap between the equal share in total resources (prize money and surplus) and own prize.  $s_i^E$  is negatively correlated with effort. An equal distribution of gross payoffs  $\pi^E = \{(2/3)\Pi, (2/3)\Pi, (2/3)\Pi\}$  (point  $E$ ) occurs (*egalitarian* solution). Here, the surplus is devoted to wiping out all inequality with regard to gross payoff by imposing a distribution principle which resembles Rawls’

[60] difference principle (though the demand game does not involve a veil of ignorance). However, if erroneously applied to gross income instead of net income, this principle penalizes effort by reversing the rank ordering produced by the contest.

As regards different setups of the demand game, we have the following expectations. The *unanimity rule* assigns every subject the same veto power regardless of her performance in the contest. Hence, subjects deal separately with the different tasks. In the default option of the demand game, the experimenter retains the surplus; that is, all subjects are treated alike in being given nothing. In line with the literature (see, e.g., [61]), we therefore conjecture the so-called symmetric solution, equal sharing ( $\alpha = 0.5$ ), to be the only distribution standard compatible with the unanimity rule.

If the demand game is played with *simple majority voting* instead of the unanimity rule, we have to carefully distinguish between *symmetric* and *right-skewed* prize schemes. Let  $\bar{\pi} = (1/3)\Pi$  denote the equal share. The first derivative of (4) with respect to  $\alpha$  is given by

$$\frac{\partial s_i^\alpha}{\partial \alpha} = 2(\pi_i - \bar{\pi}). \quad (5)$$

A payoff-maximizing subject would therefore vote for  $\alpha = 1$  if her prize money exceeds the equal share, she would vote for  $\alpha = 0$  if her prize money falls short of the equal share, and she would be indifferent if her prize money equals  $\bar{\pi}$ . In our experimental setup, right-skewed prize schemes consist of two prizes falling short of  $\bar{\pi}$  and one prize exceeding it. Hence, the  $B$ - and  $C$ -players are expected to agree on  $\alpha = 0$ , giving rise to the egalitarian solution. In contrast to this, if a symmetric prize scheme applies, the  $B$ -player is indifferent because she cannot change her personal surplus share by voting on  $\alpha$ . The  $B$ -player’s vote on the distribution parameter might be guided by social preferences. Inequity aversion would induce the  $B$ -player to agree with the  $C$ -player on an  $\alpha < 0.5$ . On the other hand, if she acknowledges the higher ability of the  $A$ -player, she might agree with the  $A$ -player on an  $\alpha > 0.5$ . The  $B$ -player might not be interested in  $\alpha$  at all and just randomizes between the proposals of  $A$  and  $C$  in order to prevent the default option. Since, we have no a priori information about the distribution of  $B$ ’s preferences, we conjecture that any feasible agreement  $\alpha \in [0, 1]$  is equally likely. Hence, on average, we should observe that  $\alpha = 0.5$ , the equal sharing distribution standard. These considerations lead to a testable version of the *Protection Hypothesis*.

- (H1) Both unanimity rule and simple majority voting with a symmetric prize scheme result in a moderate level of redistribution ( $\alpha = 0.5$ , equal sharing). Simple majority voting with a right-skewed prize scheme leads to the lowest possible value of the distribution parameter; namely  $\alpha = 0$ , the egalitarian solution.

Recall that any agreement  $\alpha \in [0, 1]$  is a Pareto-efficient Nash equilibrium. Nevertheless, default might occur if subjects decide to veto an agreement in order to prevent “unfair” outcomes. Then, they would sacrifice potential payoff in exchange for keeping the respective prize scheme intact. Obviously, subjects’ veto power is highest if the decision



mode is the unanimity rule. Furthermore, it is more difficult for three subjects to coordinate on a joint distribution standard than just for two as in the case, where  $\alpha$  is determined by simple majority vote. These considerations lead to a testable version of the *Efficiency Hypothesis*.

- (H2) Fixing the distribution standard by the unanimity rule involves more group disagreement in terms of default than simple majority voting and, therefore, is less efficient.

Finally, we have to address the issue whether subjects change their behavior in the contest when anticipating the results of the demand game. Large distribution parameters  $1 \geq \alpha \geq 0.5$  definitely preserve the rank ordering determined by the contest in terms of net payoff. A risk neutral contestant would therefore plan her effort independent from the surplus sharing task. Small distribution parameters  $0.5 > \alpha \geq 0$  would punish effort and possibly reverse the rank ordering in terms of net payoff if  $\alpha$  is sufficiently close to zero. In that case, the marginal return on effort becomes negative and effort is expected to diminish strongly. Hence, we can complement the Protection Hypothesis and the *Efficiency Hypothesis* by the following hypothesis.

- (H3) Subjects on average expend less effort in contests with right-skewed prize schemes if the decision mode in the demand game is simple majority voting.

## 4. Experimental Design

The experiment was held in the experimental economics laboratory at the University of Bremen and the University of Oldenburg using z-Tree [62].<sup>8</sup> In total, 216 subjects participated in 12 sessions in the 90-minute long experiments. Subjects were recruited using the recruitment system ORSEE [63]. About half of the subjects were female (45.6%). 37% of the participants were students of economics, 18% were students of social sciences, and the rest came from all other fields. Participants were on average in the sixth semester. At the outset, subjects were drawn from the subject pool and randomly assigned to a single session. No subject was allowed to take part more than once. We used anonymous random matching of the subjects assigned to a certain treatment into their respective groups. Hence, individual decisions should be independent from one another.

**4.1. Treatment and Round Structure.** The experiment involved four treatments that differed in

- (i) the way subjects were allocated their initial *endowments*  $\pi^D$  (contest, random);
- (ii) the *quorum* required for fixing the distribution parameter  $\alpha$  (majority, unanimity).

Our BASELINE treatment is contest  $\times$  majority. Subjects first took part in a contest and then determined the distribution parameter for sharing the surplus by majority vote.<sup>9</sup> In the VETO treatment unanimity was required, that is, every contestant was given the same veto power. We additionally

conducted BASELINE and VETO without a contest in the first step. Here, initial endowments were randomly allocated to the subjects (CONTROL-B and CONTROL-V treatments). Each treatment involved eight rounds. Rounds differed as to the first three moments (mean, variance, and skewness) of the prize schemes. In each round, the prize scheme was presented to the subjects before the knowledge test started.

Given the structure of our experiment, treatment effects regarding the impact of the quorum and endowment allocation procedure are investigated at the between-subjects level. Treatment effects regarding the impact of the shape of the prize scheme are investigated at the within-subjects level. This procedure was chosen, on the one hand, in order not to confuse subjects with different voting and allocation mechanisms and, on the other hand, to study the impact of the variation of the moments of the prize scheme on the subjects' effort and distribution preferences. Apart from this argument, the number of subjects needed for the experiment shrunk by a factor of eight.

**4.2. First Task: Knowledge Test.** Subjects participated under time pressure in a contest. More specifically, they were presented ten questions differing in complexity which had been taken from an intelligence test. Questions were displayed in sequential order. Participants had a time restriction of 15 seconds per question. Going back and forth between questions was not possible.<sup>10</sup> For each right (wrong) answer they obtained (lost) one point. Questions that remained unanswered after the time limit of 150 seconds yielded zero points. After having completed the test, subjects were randomly matched with two other subjects into groups of three. A subject's success in terms of points collected determined his or her rank  $i \in \{A, B, C\}$  in the group. A random number drawn from the unit interval was added to the number of collected points in order to avoid ties. Participants were informed about the procedure but were not informed whether they were actually affected.

In each round, each rank was endowed with a prize in tokens exchangeable for money at the end of the experiment. The respective amount was taken from the left of Table 1. Note that the number stated in the first column of the table is given for reference purposes only. The prize schemes were presented in randomized order to each group.<sup>11</sup> After the end of the quiz, subjects learned their own ranks and the ranks of the other group members, and received their prize money according to the relevant prize scheme.<sup>12</sup> They did not receive information about the absolute number of correctly solved tasks of any group member. The group assignment was renewed in every round (stranger design) in order to avoid supgame effects.<sup>13</sup>

An important feature of the set of prize schemes is that symmetric prize schemes render the *B*-player "neutral" because her final payoff is independent of  $\alpha$ . To see this, we plug  $\pi_B = (1/3)\Pi$  into (4), which yields  $s_B^\alpha = (1/3)\Pi$  for all  $\alpha \in [0, 1]$ . Hence, it is only her "impartial" distribution attitude, which decides whether or not to agree, later on in the surplus-sharing task, with player *A* or *C* on a specific  $\alpha$ , and not self-interest. As opposed to this, right-skewed prize

TABLE 1: Set of prize schemes.

Number	Rank			Parameters		
	A	B	C	Mean	Coefficient of Variation	Skewness
1	33	17	0	l	h	s
2	67	33	0	h	h	s
3	25	17	8	l	l	s
4	50	33	17	h	l	s
5	36	7	7	l	h	r
6	72	14	14	h	h	r
7	26	12	12	l	l	r
8	52	24	24	h	l	r

Note. L: low, h: high, s: symmetric, and r: right-skewed.

schemes discriminate against the *B*-player by allocating the same prize money to her and *C*. Since *B* tried harder than *C*, it is likely that her net payoff from the contest will fall below *C*'s. Regardless of that, *B*'s expected utility is higher than *C*'s because she is more able and; therefore, she will expend more effort.

If subjects were assigned to the Control treatments their ranks  $i \in \{A, B, C\}$  were determined by a random number generator instead of a contest. Everything else, in particular the parametrization of the rounds, stayed the same.

**4.3. Second Task: Surplus Sharing.** The second task involved a phase of preplay communication. Subjects were allowed to chat with their group mates about the distribution parameter to be fixed. For this purpose, we arranged a separate chat room for each group. A calculator for the gross payoff distribution was available during the whole task. The chat was available for three minutes at maximum. Afterwards, subjects were prompted to type in, simultaneously and anonymously, their preferred distribution parameter. The subjects were informed that if there was no agreement on a distribution parameter according to the necessary quorum, they would only receive their prize money. In BASELINE at least two group members had to type in exactly the same distribution parameter. In VETO all group members had to type in exactly the same distribution parameter in order to find an agreement. Note that the parametrization of the experiment secured that the distribution of the surplus for a given prize scheme was equivalent to redistributing half of total income (the sum of prize money and surplus). Each distribution standard  $\alpha$  corresponded to a specific level of redistribution  $\tau = 1 - \alpha$ .<sup>14</sup> For example, the egalitarian distributional standard  $\alpha = 0$  corresponded to full redistribution  $\tau = 100$  percent. We favored the design of the experiment as a distributional problem over its presentation as a redistribution task because the chosen design highlights that in case of group default subjects are left alone with their individual outcomes from the contest and nobody can benefit from belonging to a group or society.

At the end of the experiment, after the eighth round had been played, one round was randomly selected for payoff.

The distribution parameter agreed upon by the group was applied for calculating the gross payoff. If subjects failed to reach an agreement in the surplus sharing task, they received only their prizes. Finally, tokens were converted at a rate of 4:1 into Euros. All payments were made in cash and in private. The minimum and the maximum possible payoffs were zero and 36 Euros, respectively. On average, subjects earned €12.01 ( $\approx$ US\$ 13.67) plus a show-up fee of €5 ( $\approx$ US\$ 5.69) for about 90 minutes of work.

## 5. Results

We present the results of our experiment in three steps. First, we investigate individual and group effort, then we compare the default rates between different treatments and prize schemes, and, finally, we report the results of testing the hypotheses regarding the outcome of the vote and relate them to effort choice. Furthermore, the outcome of the surplus sharing task after the contest is compared with a setup where the contest was replaced by a random generator.

**5.1. Effort.** BASELINE and VETO were conducted each with 72 subjects, randomly forming  $72 \times 8/3 = 192$  groups. We start off with the analysis of individual effort expended in the contest. The results of running separate fixed effects (FE) regressions for both treatments are presented in Table 2. We regressed individual effort on the parameters of the prize schemes, mean (0 = low, 1 = high), variance (0 = low, 1 = high), and skewness (0 = symmetric, 1 = right skewed). The expected signs of their coefficients are given in the last column of the table. The fit of these regressions is very low, which is not too surprising given the usual noise in experimental data and given the fact that we did not induce abilities. Varying mean and variance did not have a significant influence on individual effort.<sup>15</sup> For skewness, we obtain a rather interesting pattern. In BASELINE, subjects behaved as expected by significantly reducing effort with right-skewed prize schemes. As opposed to this, in VETO subjects increased effort as compared to symmetric prize schemes. In the following, we will therefore concentrate on skewness, quorum, and their interaction.

Mean group effort across all treatments is 8.25 points. Figure 2 shows how group effort varied between treatments and prize schemes, each bar representing 96 group observations. As indicated by the FE regression, the mean difference between VETO and BASELINE is negative and insignificant ( $-0.13$ , SE: 0.51,  $P = 0.399$ ).<sup>16</sup> Due to the opposing effect of skewness between treatments, the mean effort difference between symmetric and right-skewed prize schemes is insignificant too (0.27, SE: 0.51,  $P = 0.300$ ). In fact, mean group effort, like individual effort, decreases by 1.66 points (SE: 0.68) in BASELINE, which is highly significant ( $P \leq 0.01$ ), while it increases by 1.13 points (SE: 0.77) in VETO, which is significant at the 10% level ( $P = 0.065$ ).

We conclude that hypothesis **H3** is clearly supported by our data both at the individual and at the group level. Subjects expended less effort in contests with right-skewed

TABLE 2: Individual effort by treatment.

	Coefficient	SE	$P$	Expected sign
Baseline				
Mean	-0.281	0.231	0.228	+
Variance	0.038	0.209	0.855	+
Skewness	-0.552	0.194	0.006	-
Constant	3.170	0.200	0.000	
$R^2$	0.0132			
$F$	3.76		0.015	
Veto				
Mean	-0.313	0.208	0.137	+
Variance	-0.063	0.181	0.731	+
Skewness	0.375	0.222	0.095	-
Constant	2.729	0.180	0.000	
$R^2$	0.008			
$F$	1.96		0.128	

Note. Fixed effects regression ( $n = 72$  in both regressions). Endogenous variable: effort. SE: robust standard errors. Subject and period dummies.  $P$ : significance level of a two-tailed  $t$ -test.

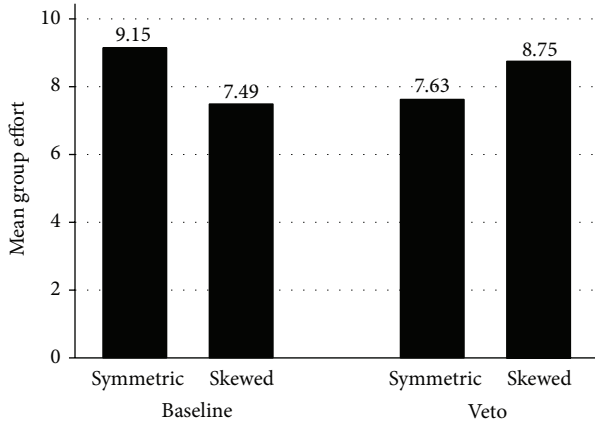


FIGURE 2: Mean effort. The figure shows mean group effort in points by treatment and prize scheme.  $n = 96$  for each bar.

prize schemes when the surplus sharing task involved the simple majority rule. Additionally, we observed the opposite effect for the VETO treatment, though the effect was less pronounced. In order to analyze our subjects' effort reactions to the anticipation of surplus distribution in the second part of the experiment more closely, we compiled the respective information on mean effort, separated by rank and treatment, in Table 6 in Appendix A. First focusing on simple majority voting, it becomes apparent that the decline of group effort was mainly due to the less able subjects  $B$  and  $C$  who expended significantly less effort when the contest involved a right-skewed prize scheme ( $B$ -player:  $P = 0.004$ ;  $C$ -player:  $P = 0.043$ ), while the most able subject  $A$  also showed a negative but insignificant reaction ( $P = 0.326$ ). The unanimity rule induced all subjects to expend a bit more

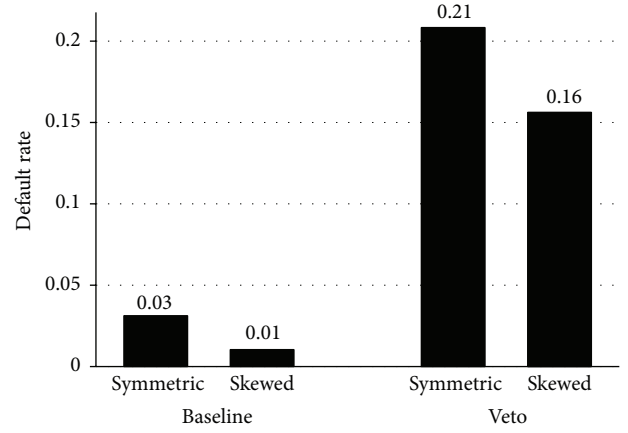


FIGURE 3: Default rates. The figure shows the default rate by treatment and prize scheme.  $n = 96$  for each bar.

effort when the contest involved a right-skewed prize scheme, but all differences are insignificant ( $A$ -player:  $P = 0.167$ ;  $B$ -player:  $P = 0.216$ ;  $C$ -player:  $P = 0.114$ ). Hence, we refrain from speculating about the reasons for this effect.

**5.2. Default.** Before moving on to analyzing the distribution parameter, we have to filter out those groups that failed to come to an agreement. The overall default rate is 10.2%. Figure 3 shows the default rates by treatment and prize scheme. As expected by the *Efficiency Hypothesis*, the unanimity requirement for  $\alpha$  made default more likely in VETO than in BASELINE (mean difference: 16.1%, SE: 3.0%,  $P \leq 0.01$ ). At first sight, default rates also seem to be lower for right-skewed prize schemes, but the difference turned out to be insignificant at usual significance levels (3.6%, SE: 3.1%,  $P = 0.119$ ). Applying the test separately to each treatment does not change the result (BASELINE: 2.1%, SE: 2.1%,  $P = 0.157$ ; VETO: 5.2%, SE: 5.6%,  $P = 0.176$ ). We conclude that **H2** is clearly supported by the data. Note that groups that failed at finding an agreement on  $\alpha$  exhibited significantly lower group effort (0.769 points, SE: 0.276,  $P = 0.003$ ). Default was not associated with higher within-group effort variation (mean difference: -7.0%, SE: 51.2%,  $P = 0.554$ ).

**5.3. Distribution.** After excluding groups that defaulted, 345 group observations remained, 188 (93/95) in BASELINE (symmetric/right-skewed) and 157 (76/81) in VETO. The mean of the distribution parameter across all groups is 0.366; that is, the average distribution attitude is a mixture between equal sharing (point A in Figure 1) and the difference principle (point E). However, the black bars in Figure 4 show  $\alpha$ s of around 0.5 (equal sharing) for all situations except for BASELINE (majority) combined with right-skewed prize schemes. There,  $\alpha$  is very close to zero, that is, the egalitarian distribution standard. Comparing means between treatments (VETO versus BASELINE: 0.142, SE: 0.029,  $P \leq 0.01$ ) and between prize schemes (symmetric versus skewed: 0.236, SE: 0.027,  $P \leq 0.01$ ) supports the *Protection Hypothesis* that

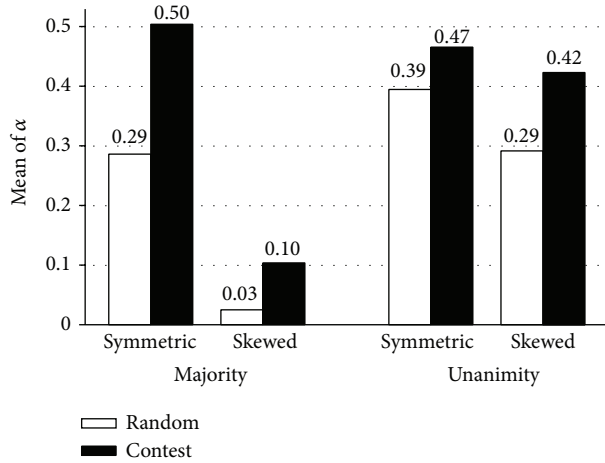


FIGURE 4: Distribution parameter. The figure shows the mean of the distribution parameter  $\alpha$  by treatment and prize scheme. Black bars represent BASELINE and VETO (contest);  $n = 93, 95, 76$ , and 81 group observations. White bars represent CONTROL (random endowment);  $n = 48, 48, 47$ , and 41 group observations.

unanimity requirement and symmetric prize schemes make strong (re)distribution less likely. The effect of skewness is more extreme in BASELINE but significant in both treatments (BASELINE: 0.400, SE: 0.038,  $P \leq 0.01$ ; VETO: 0.043, SE: 0.029,  $P = 0.069$ ).

We conclude that **H1** is clearly supported by the data. Additionally, we take a glance at the final proposals for  $\alpha$  submitted by the different players (the respective figures can be taken from the right panel of Table 6 in Appendix A). As implied by the *Protection Hypothesis*, it is mainly the B-player who is responsible for this result. When moving from symmetric to right-skewed prize schemes, her  $\alpha$  drops from 0.51 – punctuating her neutrality – to 0.10 ( $P \leq 0.01$ ). It is also interesting to see that both the A-player and the B-player made concessions: when moving to a right-skewed prize scheme the A-player “offered”  $\alpha = 0.4$  instead of  $\alpha = 0.59$  ( $P \leq 0.01$ ); when moving to a symmetric prize scheme the B-player “offered”  $\alpha = 0.41$  instead of  $\alpha = 0.1$  ( $P \leq 0.01$ ). In the VETO treatment, we observe significant differences between symmetric and right-skewed prize schemes with regard to the submitted  $\alpha$  too (A-player:  $P = 0.052$ ; B-player:  $P = 0.026$ ; C-player:  $P = 0.032$ ). Although all results are relatively close to equal sharing ( $\alpha = 0.4$ ), subjects seem to have perceived the right skewed prize scheme less fair than the symmetric and, therefore, were willing to compensate the losers with a slightly larger share of the surplus. Finally, we would like to note that the dispersion of submitted distribution parameters was highest in BASELINE with right-skewed prize schemes, pointing to huge friction within the respective groups.

One could argue that the  $\alpha$ -values represented by the black bars do not relate to subjects’ approval of individual effort or ability in the contest, but to their acceptance of the prize-schemes specified by the experimenter. In order to check for that objection, we set up the CONTROL treatment, where subjects were randomly endowed with the same set of “prizes” as in BASELINE and VETO. 72 subjects participated

in CONTROL. The white bars in Figure 4 represent the mean  $\alpha$ -values of the groups that did not default. There are two striking observations. First, the average height of the white bars is significantly lower than the black ones (0.247 versus 0.366, mean difference:  $-0.119$ , SE: 0.025,  $P \leq 0.01$ ). This means that randomly endowing subjects gave rise to a more egalitarian distribution attitude. Groups strove towards wiping out inequality caused by chance rather than effort. Second, the white bars seem to replicate the black-bar pattern produced by the contest. The situation with majority vote and right-skewed initial endowment exhibits the lowest distribution parameter of only 0.03.

Table 3 gives the difference of the group mean of  $\alpha$  between contest and random endowment by quorum and prize scheme. All mean differences exhibit the expected sign and are significantly greater than zero at least at the 5 percent level. Figure 4 and Table 3 indicate that the difference between contest and random endowment is largest for situations involving symmetric endowments/prize schemes and majority vote. This conjecture is supported by the data, the difference of the mean differences of  $\alpha$  between symmetric  $\times$  majority and symmetric  $\times$  unanimity is 0.147 (SE: 0.069,  $P = 0.017$ ), right-skewed  $\times$  majority is 0.087 (SE: 0.068,  $P = 0.100$ ), and right-skewed  $\times$  unanimity is 0.139 (SE: 0.064,  $P = 0.015$ ). All other pairwise difference tests are insignificant. In other words, if at all, individual effort in terms of success in a contest is acknowledged by a group majority only if the prize scheme is symmetric and the distribution of the surplus is determined by the majority vote. Despite the general acceptance of the outcome of the contests, groups on average agreed on a distribution parameter close to 0.5, making equal sharing the dominant distribution motive. This setup was also associated with highest group effort (9.15, see Figure 2). The acceptance of the outcome of the contest shrunk dramatically with a right-skewed prize scheme. In this situation the majority agreed willingly on counteracting ex-ante inequality by imposing the difference principle. Hence, default rates shrunk, but at the cost of lower effort. The unanimity requirement obviously shifted the focus away from effort towards cooperation in order to obtain the surplus. In that situation groups geared themselves to equal sharing combined with a small correction for ex-ante inequality. The respective  $\alpha$ -values fall below 0.5 and are almost independent from the shape of the prize scheme/endowment distribution.

## 6. Conclusion

In this paper, we experimentally test if both the existence of a middle class and institutional hurdles can protect the most able members of the society from the demand of the least able for excessive income distribution policies (*Protection Hypothesis*). We operationalized the institutional rule by the quorum needed for a group decision (majority versus unanimity) and the social structure by the existence of a middle class in contrast to a winner-take-all society. Moreover, we have argued that setting up institutional hurdles is inferior to relying on a middle class as institutional



TABLE 3: Contest versus random endowment.

Prize scheme	Quorum	
	Majority	Unanimity
Symmetric	0.218	0.071
	(0.057)	(0.035)
	$P \leq 0.01$	$P = 0.022$
Right-skewed	0.079	0.131
	(0.029)	(0.031)
	$P \leq 0.01$	$P \leq 0.01$

Note. First row: mean difference of  $\alpha$  between contest and random. Second row: standard errors. Third row: significance level of a one-tailed  $t$ -test.

hurdles involve efficiency losses due to bargaining impasses (*Efficiency Hypothesis*).

The experiment directly tests our main hypotheses. We find that the *Protection Hypothesis* is clearly supported by our data. The existence of a neutral middle class is as effective as institutional hurdles to protect the rich from being expropriated. Hereby, the middle class balances the interests of the poor and the rich. The middle class does not have an interest in excesses from one side because redistribution shifts money from the rich to the poor. We argue that the middle class can be interpreted as an uninvolved spectator who has no interest to favor one group over another. The *Efficiency Hypothesis* is also supported. Institutional hurdles come at the cost of significantly less efficiency in terms of higher default rates. Hereby, a higher default rate means that the payoff for both the rich and the poor is lower. Even, at the individual level, we could observe that subjects expend more effort if the prize scheme involved a moderating middle player. The corroboration of both hypotheses allows us to draw the conclusion that the middle class is the efficient protection of the rich.

We contribute to the literature about why the rich do not get expropriated. Leaving aside the branch of the literature which summarizes the reasons of why the poor do not expropriate the rich (e.g., beliefs, identity, and social preferences of the poor), we study a constellation in which the rich can actively protect themselves from being expropriated. Although any extrapolation of our laboratory results has to be done carefully, we note that our finding that collective decisions have been less equalizing and more efficient under the majority rule with a neutral middle player than in groups with a right-skewed prize structure resonates well with the empirical data, supporting the warnings about the impact of rising inequality on societal stability and economic growth in contemporary societies. Coming back to the empirical findings in the introduction: not only the middle class itself but also the rich should have an interest in bracing against the erosion of the middle class.

## Appendices

### A. Further Tables

See Tables 4, 5, and 6.

### B. Instructions

*B.1. Welcome to the Experiment and Thank You for Participating! Please do not talk to other participants during the entire experiment!* During the experiment, you have to make several decisions. Your individual payoff depends on your own decision and the decisions of your group members due to the following rules. You will be paid individually, privately, and in cash after the experiment. During the experiment, we will talk about Tokens as the experimental currency. After the experiment, Tokens will be transferred into Euros with the following exchange rate:

$$10 \text{ Tokens} = 2.50 \text{ Euros.} \quad (\text{B.1})$$

Please take your time reading the instructions and making your decisions. You are not able to influence the duration of the experiment by rushing through your decisions because you always have to wait until the remaining participants have reached their decisions. The experiment is completely anonymous. At no time during the experiment or afterwards will the other participants know which role you were assigned to and how much you have earned.

If you have any questions please raise your hand. One of the experimenters will come to you and answer your questions privately. Following this rule is very important. Otherwise the results of this experiment will be worthless for scientific purposes. You will receive a show-up fee of 5 Euros for your participation. Depending on your decisions and the decisions of the other participants you can additionally earn up to 144 Tokens (36 Euros). The expected duration of the experiment is 90 minutes. The exact course of the experiment will be described in the following.

*B.2. Detailed Information.* The experiment consists of 8 rounds which all follow the same course. In each round participants will be randomly and repeatedly assigned to groups of three members. Your payoff will be determined only by your own decisions and decisions of the other group members. Decisions of the other groups do not affect your payment. No participant receives information about other subjects' final choices, although group members have the possibility to communicate via chat to find an agreement. The final decisions are anonymous within each group.

After the group assignment, a knowledge test follows (*in the control treatments, ranks were assigned by a random generator instead of the knowledge test*). The more questions you answer correctly in the knowledge test, the higher your potential payoff at the end of the experiment will be. The knowledge test consists of various tasks. For each correctly solved task you receive one point; each question you answered incorrectly, a point is subtracted, and for not answering questions you do not get any points. Based on the collected points you will be assigned a rank within the group. The player with the highest score in the knowledge test gets the highest rank (rank 1) and also the highest token endowment. The player with the lowest number of collected points will be assigned the lowest rank (rank 3) and also the lowest token endowment.



TABLE 4: Predictions for the contest.

Prize scheme	Player	Ability	Expected winning probabilities (%)			Expected return	Effort	Rank	Prize	Cost	Net payoff
	$i$	$c_i$	1st	2nd	3rd	$E(\pi_i)$	$x_i$	$j$	$\pi_j$		
1	1	0.4	56.2	37.5	6.2	10	38	A	33	15	18
	2	0.6	25.0	50.0	25.0	4	21	B	17	13	4
	3	0.8	6.2	37.5	56.2	1	9	C	0	8	-8
2	1	0.4	56.2	37.5	6.2	19	77	A	67	31	36
	2	0.6	25.0	50.0	25.0	8	42	B	33	25	8
	3	0.8	6.2	37.5	56.2	2	18	C	0	15	-15
3	1	0.4	56.2	37.5	6.2	13	20	A	25	8	17
	2	0.6	25.0	50.0	25.0	10	11	B	17	7	10
	3	0.8	6.2	37.5	56.2	8	5	C	8	4	4
4	1	0.4	56.2	37.5	6.2	27	38	A	50	15	35
	2	0.6	25.0	50.0	25.0	21	21	B	33	12	21
	3	0.8	6.2	37.5	56.2	18	9	C	17	7	10
5	1	0.4	56.2	37.5	6.2	12	29	A	36	11	25
	2	0.6	25.0	50.0	25.0	8	10	B	7	6	1
	3	0.8	6.2	37.5	56.2	7	2	C	7	2	5
6	1	0.4	56.2	37.5	6.2	24	57	A	72	23	49
	2	0.6	25.0	50.0	25.0	16	20	B	14	12	2
	3	0.8	6.2	37.5	56.2	14	4	C	14	3	11
7	1	0.4	56.2	37.5	6.2	14	14	A	26	6	20
	2	0.6	25.0	50.0	25.0	13	5	B	12	3	9
	3	0.8	6.2	37.5	56.2	12	1	C	12	1	11
8	1	0.4	56.2	37.5	6.2	29	28	A	52	11	41
	2	0.6	25.0	50.0	25.0	25	10	B	24	6	18
	3	0.8	6.2	37.5	56.2	24	2	C	24	2	22

Note. Predictions for a 3-player contest with linear cost functions and maximum ability,  $m = 0.2$ .

TABLE 5: Predicted effort by prize scheme.

Prize scheme	Mean	Coefficient of variation	Skewness	Effort	
				Mean	Variation
1	0	1	0	22.9	.6
2	1	1	0	45.9	.6
3	0	0	0	11.9	.6
4	1	0	0	22.5	.6
5	0	1	1	13.6	1.0
6	1	1	1	27.2	1.0
7	0	0	1	6.6	1.0
8	1	0	1	13.1	1.0

Note. 0 = low/symmetric, 1 = high/right-skewed. Effort predictions according to Table 4.

The knowledge test is designed with a time restriction of 150 seconds for 10 screens with different tasks. The tasks come from different fields. Each question has only one correct answer. You have 15 seconds per question. Please pay attention to the time restriction in the upper right corner

TABLE 6: Effort and distribution proposal by treatment.

Prize scheme	Subject's rank					
	Effort			$\alpha$		
	A	B	C	A	B	C
Simple majority voting						
Symmetric	4.39	3.14	1.63	0.59	0.51	0.41
	(0.22)	(0.20)	(0.32)	(0.03)	(0.03)	(0.03)
Right-skewed	4.26	2.38	0.85	0.40	0.10	0.10
	(0.17)	(0.20)	(0.30)	(0.04)	(0.02)	(0.02)
Unanimity rule						
Symmetric	3.92	2.55	1.16	0.48	0.49	0.47
	(0.21)	(0.21)	(0.31)	(0.02)	(0.02)	(0.02)
Right-skewed	4.24	2.80	1.72	0.43	0.41	0.43
	(0.26)	(0.22)	(0.35)	(0.02)	(0.03)	(0.02)

Note. The figures give the subjects' mean scores in the intelligence test (per round) and the mean of the distribution parameter  $\alpha$  typed in after the chat phase. Standard errors are in parentheses.

of the screen. By clicking on the OK button you can get to the next screen. After a total of 150 seconds, the knowledge

test stops and the collected points are summed up. The ranks and the associated endowments will be provided after the knowledge test, explanations, and further instructions.

**B.3. After the Knowledge Test.** You have collected points in the knowledge test. According to the collected points of all group members each group member will be assigned a rank and a certain endowment of tokens. The ranking reflects directly the performance in the knowledge test of the group's members (*in the control treatment the ranks and endowment were assigned randomly*). The token endowment varies from round to round. A total of 8 rounds will be played. The sequential arrangement of token distributions will be chosen randomly.

**B.4. Decision about Payoff Distribution by Choosing a Distribution Parameter.** In the experiment, you will see on the next screen your own rank and token endowment and the ranks and token endowments of your group members. In the following you and your group members have to determine the distribution parameter. You should set a distribution parameter which distributes tokens between the group members in such a way that your preferences are met. The payoff of the group member  $X = 1, 2, 3$  at the end of the experiment is then calculated as follows:

$$\text{Payoff } X = \text{Tokens } X \cdot (1 - t) + \frac{(\text{Tokens } 1 + \text{Tokens } 2 + \text{Tokens } 3) \cdot t}{3} \quad (\text{B.2})$$

Distribution only takes place when at least 2 (3, *under unanimity rule*) group members choose the same distribution parameter. The chosen distribution parameter determines the final distribution within the group in that period. Sample calculation: player 1 has 3 tokens, player 2 has 5 tokens, and player 3 has 10 tokens. Given that order, players chose the following distribution parameters: 81% (50%, *under unanimity rule*), 50%, and 50%. The distribution parameter, which determines the level of redistribution, is thus 50%. The payoff for player 1 is

$$\text{Payoff } 1 = 3 \cdot (1 - 0.5) + \frac{(3 + 5 + 10) \cdot 0.5}{3} = 1.5 + 3 = 4.5. \quad (\text{B.3})$$

The computer calculates the payoffs under the assumption that the distribution parameter you make is the relevant distribution parameter which will be implemented.

If not at least 2 (3, *under unanimity rule*) players from the group typed in the same distribution parameter, no distribution takes place and players receive only their token endowment determined by their rank in the knowledge test. In the example above, this would mean that if the players chose distribution parameters of 96%, 51%, and 7% and therefore no agreement is reached, player 1 receives only 1.5 tokens. The same would apply to the endowments of the remaining players.

**B.5. Possibility to Communicate (Open or Restricted).** You have the possibility to chat with the other group members in a joint chat room. You can chat freely (*restricted in the way that you can only type in numbers between 0 and 100, under restricted communication rule*) and discuss which distribution parameter is to be implemented, but you are not allowed to reveal your identity.

Chat time is restricted to 3 minutes. If not at least 2 (3, *under unanimity rule*) players have entered the same distribution parameters in time and have confirmed their chosen distribution parameters by clicking the OK button, then no distribution will take place.

The calculator is available to you for trying out different distribution parameters. To get an impression about the chat structure, the endowments and the tokens, and the calculator, a sample screen is given below. In the upper part of the screen you see the group chat room. The rank number of each player is displayed directly above and within the chat window. In the lower part of the screen the ranks and tokens of all group members before and after distribution are displayed. In the right field, you can enter as often as you wish different distribution parameters to calculate the final token distribution. In this example, the final distribution is calculated as follows: the distribution parameter is 45%. Hence, 45% of 6 = 2.7, 45% of 10 = 4.5, and 45% of 4 = 1.8 are subtracted from the players' accounts, resulting in a sum of 9 Tokens. Dividing the sum by 3 gives 3 Tokens to be distributed to each group member. The final distribution then is given by  $6 - 2.7 + 3 = 6.3$ ,  $10 - 4.5 + 3 = 8.5$ , and  $4 - 1.8 + 3 = 5.2$  Tokens. For example, using the equation given above, the payoff of player 3 would be given by

$$\text{Payoff} = 4 \cdot (1 - 0.45) + \frac{(4 + 6 + 10) \cdot 0.45}{3} = 2.2 + 3 = 5.2. \quad (\text{B.4})$$

Pressing the OK button confirms and completes the decision.

**B.6. Calculation of the Payoffs.** One of the 8 rounds is randomly selected for payoff. Each round could therefore be payoff relevant. Individual payoffs will be calculated according to the chosen distribution parameter and the rules stated above. The experiment will begin shortly. If you have questions, please raise your hand and wait quietly until an experimenter comes to you. Speaking with other participants is strictly prohibited throughout the experiment. Thank you and have fun in the experiment.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgments

The authors thank two anonymous referees and the editors, in particular Klarita Gërzhani, for their comments and support. This research was partly financed by the NOWETAS Foundation under the project title "Decision making among

collectives and the individual: the interdisciplinary analysis of perceptions of justice in collective decision processes and their feedback to the individual level” and by the German Research Foundation (DFG), research group FOR 2104: “Needs-based justice and distribution procedures” (Traub: TR 458/6-1). They would like to thank Nicola Maaser and the participants of several research seminars and the workshop “Positive and Normative Aspects of Distributive Justice: An Interdisciplinary Perspective” held at the University of Bremen, in particular, for their most helpful comments and suggestions.

## Endnotes

1. A similar effect of voting was reported by Frohlich and Oppenheimer [64].
2. For more general information on contest theory, we refer to the surveys by Corchón [65] and Konrad [42]. Experimental research on contests was recently surveyed by Dechenaux et al. [66]. Multiple-prize contests have intensively been investigated with respect to the optimality of the prize scheme in terms of maximum effort generation (see, e.g., [67]).
3. In general, the cost of effort include all monetary and nonmonetary costs such as cognitive effort and time. In our experiment, the only costs of solving questions in the intelligence test were cognitive. Time consumption was equal for all subjects and covered by a show-up fee.
4. The numbers listed in both tables are only point estimates for a single distribution of ability. One could extend the simulation by randomly drawing  $c$  from  $F(c)$  a sufficiently large number of times in order to obtain confidence intervals for mean effort. However, since the stylized facts would stay unchanged, we save the effort.
5. Note that the restriction  $s_i \geq 0$  implies that the subjects’ rank ordering according to their prizes cannot be reversed by the surplus sharing task. This parallels the condition stated by Meltzer and Richard [23] that the median voter’s tax rate is utility maximizing only if it does not exceed one and therefore preserves the rank ordering between productive and less productive individuals.
6. A loosely related study by Selten and Ockenfels [68] found a relative majority of subjects to comply with a fairness norm they called “fixed total sacrifice.” Subjects were willing to donate a fixed amount of money to the losers of a solidarity game irrespective of the number of losers. In a game studied by Okada and Riedl [69] proposers could form different coalitions of up to three players and then distribute the coalition surplus among the coalition members. In most cases, the proposer chose an inefficient two-person coalition and the responder reciprocated the choice of the proposer by not rejecting the unfair proposal that excluded a third of the population from payoff.
7. Goeree and Yariv [70] found that deliberation blurs differences between institutions (e.g., in terms of voting rules) and uniformly improves efficiency of group outcomes. We borrow from the work of these authors by considering two different voting rules under which the surplus-sharing game is conducted, namely, simple majority voting and unanimity rule. In so far, our study is also related to the literature on deliberation in collective decision making (see [70–72]). For an overview of the experimental literature on deliberation, see Karpowitz and Mendelberg [73].
8. For a transcript of the instructions see Appendix B.
9. We have also varied the chat possibilities in such a way that in half of the sessions participants could exchange full text messages and in the remaining sessions participants could only exchange numbers. Due to the fact that we do not find any differences between these sessions, we decided to pool these sessions.
10. The questions compassed mathematical, linguistical, and combinatorial tasks.
11. Note that the stranger matching protocol requires keeping the randomization of the ordering of the prize schemes constant in each session. Since we conducted four sessions per treatment, we used four predefined randomizations that stayed the same for all treatments in order to control for eventual path-dependencies.
12. Prize schemes 5 and 6 slightly violate the nonnegativity constraint for  $s_i^\alpha$  for large  $\alpha$ -values. This is due to the construction of the set of prize schemes, where half of the eight different prize schemes exhibit the same mean, variation coefficient, and skewness, respectively.
13. Note that the theoretical model of the contest outlined in Section 3 has to assume that subjects know their own abilities and the distribution of abilities in order to derive the result that those who are more able also expend more effort. We did not collect subjects’ expectations about their abilities and winning probabilities in the experiment, but it could be interesting to check whether their beliefs were correct and to compare their goodness of fit with the outcome of the contest.
14. As can be taken from the instructions, we let subjects enter  $\tau$  instead of  $\alpha$ , which turned out to be more intelligible in pilot experiments.
15. We also expected right-skewed prize schemes to increase within-group effort variation among contestants. Analyzing effort variation again speaks a slightly different language. Indeed, it seems to increase in BASELINE. However, the mean difference turns out to be insignificant (0.34, SE: 0.53,  $P = 0.259$ ). In VETO, the mean difference has the wrong sign, but it is significant (0.65, SE: 0.28,  $P = 0.012$ ).
16. Significance level of a one-tailed  $t$ -test. Tests regarding directional hypothesis are one-tailed unless otherwise specified. SE: standard error.

## References

- [1] T. Piketty, *Capital in the Twenty-First Century*, Belknap Press, Cambridge, Mass, USA, 2014.
- [2] OECD, *Growing Unequal? Income Distribution and Poverty in OECD Countries*, OECD, Paris, France, 2008.
- [3] S. K. Singh and G. S. Maddala, "A function for size distribution of incomes," *Econometrica*, vol. 44, no. 5, pp. 963–970, 1976.
- [4] C. Seidl, K. Pogorelskiy, and S. Traub, *Tax Progression in OECD Countries. An Integrative Analysis of Tax Schedules and Income Distributions*, Springer, Berlin, Germany, 2013.
- [5] R. H. Frank and P. Cook, *The Winner-Take-All Society*, Free Press, New York, NY, USA, 1995.
- [6] R. H. Frank, *How Rising Inequality Harms the Middle Class*, University of California Press, Berkely, Calif, USA, 2007.
- [7] J. S. Hacker and P. Pierson, *Winner-Take-All Politics. How Washington Made the Rich Richer—and Turned Its Back on the Middle Class*, Simon & Schuster, New York, NY, USA, 2010.
- [8] S. Rosen, "The economics of superstars," *The American Economic Review*, vol. 71, no. 5, pp. 845–858, 1981.
- [9] T. A. DiPrete and G. M. Eirich, "Cumulative advantage as a mechanism for inequality: a review of theoretical and empirical developments," *Annual Review of Sociology*, vol. 32, pp. 271–297, 2006.
- [10] E. Franck and S. Nüesch, "Talent and/or popularity: what does it take to be a superstar?" *Economic Inquiry*, vol. 50, no. 1, pp. 202–216, 2012.
- [11] J. E. Stiglitz, *The Price of Inequality: How Today's Divided Society Endangers Our Future*, W. W. Norton & Company, New York, NY, USA, 2012.
- [12] F. Fukuyama, *The End of History and the Last Man*, Free Press, New York, NY, USA, 1992.
- [13] A. Alesina and G.-M. Angeletos, "Fairness and redistribution," *American Economic Review*, vol. 95, no. 4, pp. 960–980, 2005.
- [14] J. M. Buchanan and G. Tullock, *The Calculus of Consent. Logical Foundations of Constitutional Democracy*, University of Michigan Press, Ann Arbor, Mich, USA, 1962.
- [15] L. Miller and C. Vanberg, "Decision costs in legislative bargaining: an experimental analysis," *Public Choice*, vol. 155, no. 3–4, pp. 373–394, 2013.
- [16] F. W. Scharpf, "A game-theoretical interpretation of inflation and unemployment in Western Europe," *Journal of Public Policy*, vol. 7, no. 3, pp. 227–257, 1987.
- [17] W. Easterly, "The middle class consensus and economic development," *Journal of Economic Growth*, vol. 6, no. 4, pp. 317–335, 2001.
- [18] E. Glaeser, J. Scheinkman, and A. Shleifer, "The injustice of inequality," *Journal of Monetary Economics*, vol. 50, no. 1, pp. 199–222, 2003.
- [19] W. Arts and J. Gelissen, "Welfare states, solidarity and justice principles: does the type really matter?" *Acta Sociologica*, vol. 44, no. 4, pp. 283–299, 2001.
- [20] L. d'Anjou, A. Steijn, and D. van Aarsen, "Social position, ideology, and distributive justice," *Social Justice Research*, vol. 8, no. 4, pp. 351–384, 1995.
- [21] M. M. Jæger, "What makes people support public responsibility for welfare provision: self-interest or political ideology? A longitudinal approach," *Acta Sociologica*, vol. 49, no. 3, pp. 321–338, 2006.
- [22] B. Moldovanu and A. Sela, "The optimal allocation of prizes in contests," *American Economic Review*, vol. 91, no. 3, pp. 542–558, 2001.
- [23] A. H. Meltzer and S. F. Richard, "A rational theory of the size of government," *Journal of Political Economy*, vol. 89, no. 5, pp. 914–927, 1981.
- [24] R. Perotti, "Growth, income distribution, and democracy: what the data say," *Journal of Economic Growth*, vol. 1, no. 2, pp. 149–187, 1996.
- [25] A. Alesina and R. Perotti, "Income distribution, political instability, and investment," *European Economic Review*, vol. 40, no. 6, pp. 1203–1228, 1996.
- [26] R. Benabou and E. A. Ok, "Social mobility and the demand for redistribution: the POUM hypothesis," *Quarterly Journal of Economics*, vol. 116, no. 2, pp. 447–487, 2001.
- [27] J. E. Roemer, "Why the poor do not expropriate the rich: an old argument in new garb," *Journal of Public Economics*, vol. 70, no. 3, pp. 399–424, 1998.
- [28] J. E. Roemer and W. Lee, "Racism and redistribution in the United States: a solution to the problem of American exceptionalism," *Journal of Public Economics*, vol. 90, no. 6–7, pp. 1027–1052, 2006.
- [29] C. Fong, "Social preferences, self-interest, and the demand for redistribution," *Journal of Public Economics*, vol. 82, no. 2, pp. 225–246, 2001.
- [30] N. Frohlich and J. A. Oppenheimer, *Choosing Justice: An Experimental Approach to Ethical Theory*, University of California Press, Ewing, NJ, USA, 1993.
- [31] D. Miller, *Principles of Social Justice*, Harvard University Press, Cambridge, Mass, USA, 1999.
- [32] J. Konow, "Which is the fairest one of all? A positive analysis of justice theories," *Journal of Economic Literature*, vol. 41, no. 4, pp. 1188–1239, 2003.
- [33] S. Traub, C. Seidl, U. Schmidt, and M. V. Levati, "Friedman, Harsanyi, Rawls, Boulding—or somebody else? An experimental investigation of distributive justice," *Social Choice and Welfare*, vol. 24, no. 2, pp. 283–309, 2005.
- [34] A. Sen, *The Idea of Justice*, Belknap Press, Cambridge, Mass, USA, 2009.
- [35] W. Gaertner and E. Schokkaert, *Empirical Social Choice. Questionnaire-Experimental Studies on Distributive Justice*, Cambridge University Press, Cambridge, UK, 2012.
- [36] E. F. Klor and M. Shayo, "Social identity and preferences over redistribution," *Journal of Public Economics*, vol. 94, no. 3–4, pp. 269–278, 2010.
- [37] J.-R. Tyran and R. Sausgruber, "A little fairness may induce a lot of redistribution in democracy," *European Economic Review*, vol. 50, no. 2, pp. 469–485, 2006.
- [38] E. Fehr and K. M. Schmidt, "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.
- [39] W. Höchtl, R. Sausgruber, and J.-R. Tyran, "Inequality aversion and voting on redistribution," *European Economic Review*, vol. 56, no. 7, pp. 1406–1421, 2012.
- [40] G. E. Bolton and A. Ockenfels, "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: comment," *American Economic Review*, vol. 96, no. 5, pp. 1906–1911, 2006.
- [41] K. D. Messer, G. L. Poe, D. Rondeau, W. D. Schulze, and C. Vossler, "Exploring voting anomalies using a demand revealing



- random price voting mechanism,” *Journal of Public Economics*, vol. 94, pp. 308–317, 2010.
- [42] F. Paetzl, R. Sausgruber, and S. Traub, “Social preferences and voting on reform: an experimental study,” *European Economic Review*, vol. 70, pp. 36–55, 2014.
- [43] L. Balafoutas, M. G. Kocher, L. Putterman, and M. Sutter, “Equality, equity and incentives: an experiment,” *European Economic Review*, vol. 60, pp. 32–51, 2013.
- [44] J. Esarey, T. C. Salmon, and C. Barrilleaux, “What motivates political preferences? Self-interest, ideology, and fairness in a laboratory democracy,” *Economic Inquiry*, vol. 50, no. 3, pp. 604–624, 2012.
- [45] R. Durante, L. Putterman, and J. van der Weele, “Preferences for redistribution and perception of fairness: an experimental study,” *Journal of the European Economic Association*, vol. 12, no. 4, pp. 1059–1086, 2014.
- [46] R. M. Sheremeta, W. A. Masters, and T. N. Cason, “Winner-take-all and proportional-prize contests: theory and experimental results,” Working Paper 12-04, Economic Science Institute Chapman University, 2012.
- [47] J. Farrell and M. Rabin, “Cheap talk,” *Journal of Economic Perspectives*, vol. 10, no. 3, pp. 103–118, 1996.
- [48] V. Crawford, “A survey of experiments on communication via cheap talk,” *Journal of Economic Theory*, vol. 78, no. 2, pp. 286–298, 1998.
- [49] R. Croson, T. Boles, and J. K. Murnighan, “Cheap talk in bargaining experiments: lying and threats in ultimatum games,” *Journal of Economic Behavior and Organization*, vol. 51, no. 2, pp. 143–159, 2003.
- [50] A. Roth, “Toward a focal-point theory of bargaining,” in *Game-Theoretic Models of Bargaining*, A. Roth, Ed., pp. 259–268, Cambridge University Press, Cambridge, UK, 1985.
- [51] A. Roth, “Bargaining phenomena and bargaining theory,” in *Laboratory Experimentation in Economics: Six Points of View*, A. Roth, Ed., pp. 14–41, Cambridge University Press, Cambridge, UK, 1987.
- [52] R. Forsythe, J. Kennan, and B. Sopher, “An experimental analysis of strikes in bargaining games with one-sided private information,” *The American Economic Review*, vol. 81, no. 1, pp. 253–278, 1991.
- [53] M. Shaked and G. J. Shanthikumar, *Stochastic Orders and Their Applications*, Academic Press, San Diego, Calif, USA, 1994.
- [54] J. Nash, “Two-person cooperative games,” *Econometrica*, vol. 21, pp. 128–140, 1953.
- [55] A. E. Roth, “Bargaining experiment,” in *The Handbook of Experimental Economics*, J. H. Kagel and A. E. Roth, Eds., Princeton University Press, Princeton, NJ, USA, 1995.
- [56] A. Rubinstein, “Perfect equilibrium in a bargaining model,” *Econometrica*, vol. 50, no. 1, pp. 97–109, 1982.
- [57] D. P. Baron and J. A. Ferejohn, “Bargaining in legislatures,” *The American Political Science Review*, vol. 83, no. 4, pp. 1181–1206, 1989.
- [58] H. Moulin, “Equal or proportional division of a surplus, and other methods,” *International Journal of Game Theory*, vol. 16, no. 3, pp. 161–186, 1987.
- [59] L. J. Mirman and Y. Tauman, “Demand compatible equitable cost sharing prices,” *Mathematics of Operations Research*, vol. 7, no. 1, pp. 40–56, 1982.
- [60] J. Rawls, *A Theory of Justice*, Harvard University Press, Cambridge, Mass, USA, 1971.
- [61] N. Anbarci and I. Boyd, “Nash demand game and the Kalai-Smorodinsky solution,” *Games and Economic Behavior*, vol. 71, no. 1, pp. 14–22, 2011.
- [62] U. Fischbacher, “Z-Tree: zurich toolbox for ready-made economic experiments,” *Experimental Economics*, vol. 10, no. 2, pp. 171–178, 2007.
- [63] B. Greiner, “The online recruitment system ORSEE 2.0—a guide for the organization of experiments in economics,” Tech. Rep., University of Cologne, Cologne, Germany, 2004.
- [64] N. Frohlich and J. A. Oppenheimer, “Choosing justice in experimental democracies with production,” *The American Political Science Review*, vol. 84, no. 2, pp. 461–477, 1990.
- [65] L. C. Corchón, “The theory of contests: a survey,” *Review of Economic Design*, vol. 11, no. 2, pp. 69–100, 2007.
- [66] E. Dechenaux, D. Kovenock, and R. M. Sheremeta, “A survey of experimental research on contests, all-pay auctions and tournaments,” Discussion Paper No. SP II 2012-109, Wissenschaftszentrum für Sozialforschung, Berlin, Germany, 2012.
- [67] C. Harbring and B. Irlenbusch, “An experimental study on tournament design,” *Labour Economics*, vol. 10, no. 4, pp. 443–464, 2003.
- [68] R. Selten and A. Ockenfels, “An experimental solidarity game,” *Journal of Economic Behavior and Organization*, vol. 34, no. 4, pp. 517–539, 1998.
- [69] A. Okada and A. Riedl, “Inefficiency and social exclusion in a coalition formation game: experiment evidence,” *Games and Economic Behavior*, vol. 50, no. 2, pp. 278–311, 2005.
- [70] J. K. Goeree and L. Yariv, “An experimental study of collective deliberation,” *Econometrica*, vol. 79, no. 3, pp. 893–921, 2011.
- [71] D. Austen-Smith and T. J. Feddersen, “Deliberation, preference uncertainty, and voting rules,” *American Political Science Review*, vol. 100, no. 2, pp. 209–217, 2006.
- [72] D. Gerardi and L. Yariv, “Deliberative voting,” *Journal of Economic Theory*, vol. 134, no. 1, pp. 317–338, 2007.
- [73] C. F. Karpowitz and T. Mendelberg, “An experimental approach to citizen deliberation,” in *Cambridge Handbook of Experimental Political Science*, J. N. Druckman, D. P. Green, J. H. Kuklinski, and A. Lupia, Eds., pp. 258–272, Cambridge University Press, Cambridge, Mass, USA, 2011.