ACKNOWLEDGEMENTS

I am grateful to many people who have assisted me in the completion of this work. First and foremost, I would like to thank my supervisor, Francesco Feri, for being a mentor, coach and inspiration in equal measure and without whose time and support this thesis would not have been possible. I would also like to thank my advisor, Sourav Bhattacharya, for very useful contributions.

The Department of Economics at Royal Holloway, University of London has provided an excellent, open and friendly research environment. I would like to thank Juan Pablo Rud, Dan Anderberg, Jonathan Wadsworth and Ija Trapeznikova for their support. And Bjoern Hartig, for help and guidance throughout the years.

I have benefitted greatly from co-authoring with Francesco Feri, Sourav Bhattacharya, Paolo Pin and Tiziano Rotesi. The exposure to the knowledge and thoughts of others in the creation of something new has been truly inspiring.

I would also like to place on record my thanks to the Social and Economic Research Council (ESRC) who provided funding for the research to be carried out. And to the ExpReSS Lab at Royal Holloway, University of London, for granting the budget to conduct experiments.
For Chris
This thesis consists of three experimental essays in the economics of networks. Chapters 1 and 3 study network intervention, where a central planner designs optimal targeted strategies. Agents have incomplete information about their location in a network and actions have complementarities. Strategies intend to increase actions in those who take a risky decision to encourage coordination on the efficient equilibrium. Chapter 1 investigates the effectiveness of a bonus that is conditional on a minimal participation rate in the network. A tough condition does not improve outcomes, but an easier condition is very effective. A budget-friendly alternative where the bonus is aimed at a subset of the network population is less effective, but a more sophisticated mechanism design could achieve the same result. Chapter 3 is a pilot study that examines the effectiveness of a temporary discount that is available only to a subset of the network population. We have two treatments that differ only in the network formation process. In each case the discount shifts coordination to the efficient equilibrium, but the effect does not persistent once the discount expires. Nevertheless, we observe significant differences between treatments. Chapter 2 compares rumour propagation and outcomes in two network structures. The networks have biased agents and unbiased agents. Biased agents want to enforce a specific decision and unbiased agents seek the true state. One agent learns the true state and can create a message to send to her neighbours. Agents in a line network decide whether to further transmit the message or not, and agents in a complete network receive the message simultaneously. Contrary to predictions, outcomes are the same in either treatment. The behaviours that underpin the outcomes differ between treatments. In the complete network message creation by biased agents matters, and in the line network transmission by unbiased agents is key.
Table of Contents

Page

10 Introduction

12 Chapter 1. Network Intervention: Incentives with Conditions

12 1.1 Introduction

15 1.2 Related literature

18 1.3 Experimental design

18 1.3.1 Game and treatments

21 1.3.2 Equilibrium predictions

27 1.3.3 Risk dominance

31 1.4 Research hypotheses

33 1.5 Experimental results

33 1.5.1 Data analysis

40 1.5.2 Equilibrium selection

41 1.5.3 Results

44 1.6 Conclusion

46 Chapter 2. Network Structure and the Efficacy of Rumours

46 2.1 Introduction

50 2.2 Experimental design

50 2.2.1 The model

64 2.2.2 Implementation

66 2.3 Research hypotheses

74 2.4 Experimental results
List of Tables

Chapter 1. Network Intervention - Incentives with Conditions
26   Table 1.1: Equilibrium action
35   Table 1.2: Statistics by treatment and degree
37   Table 1.3: Marginal effects of treatment by degree at period 20

Chapter 2. Network Structure and the Efficacy of Rumours
58   Table 2.1: Strategy profile and beliefs for each game
69   Table 2.2: Equilibrium predictions
76   Table 2.3: B2 message creation – two-tailed t-test
83   Table 2.4: Probabilities and frequencies for correct group decision
87   Table 2.5: Pearson correlation coefficient – broadcast treatment
89   Table 2.6: Marginal effects of beliefs – broadcast treatment
91   Table 2.7: Pearson correlation coefficient – network treatment

Chapter 3. NI: Temporary Incentives Persistence
99   Table 3.1: Connection probabilities by degree and treatment
104  Table 3.2: Equilibrium predictions
110  Table 3.3: Activity by treatment and degree
113  Table 3.4: Logit regression for the choice to be active or inactive
114  Table 3.5: Marginal effects of degree = 3
115  Table 3.6: Marginal effects of incentive
List of Figures

Page

**Chapter 1. Network Intervention: Incentives with Conditions**

18  Figure 1.1: The network and characteristics
39  Figure 1.2: Observed activity rates by treatment and degree

**Chapter 2. Network Structure and the Efficacy of Rumours**

53  Figure 2.1a: The line network
53  Figure 2.1b: The complete network
76  Figure 2.2: Messages creation frequencies of biased agents
78  Figure 2.3: Message creation frequencies of unbiased agents

**Chapter 3. NI: Temporary Incentives and Persistence**

112  Figure 3.1: Evolution of aggregate activity rates by degree
Appendices

Page

**Appendix A. Network Intervention: Incentives with Conditions**

118  A.1: Logistics regression with random effects

121  A.2: Marginal effects estimation

123  A.3: Evolution of play by treatment and session

127  A.4: Game instructions

**Appendix B. Network Structure and the Efficacy of Rumours**

132  B.1: Game instructions

143  B.2: Check Your Understanding questions

145  B.3: Beliefs and expectations questionnaire

147  B.4: Social preferences questionnaire

147  B.5: Lie aversion questionnaire

148  B.6: Risk test

148  B.7: Network and broadcast subject pool comparison

149  B.8: Network treatment marginal effects estimation

150  B.9: Posterior beliefs calculations (Table 2.1)

**Appendix C. NI: Temporary Incentives - Persistence**

153  C.1: App screenshots

154  C.2 Game instructions

155  C.3 Risk test

156  C.4 Activity rates by treatment and session
INTRODUCTION

This thesis consists of three experimental essays in the economics of social network interactions. Societies are increasingly connected and the behaviour of one person affects the welfare of others. We examine mechanisms and that shape behaviours and network structures that affect beliefs. The purpose of each essay is to understand how the factors under investigation affect the welfare of the system.

Chapters 1 and 3 study network intervention, which describes the process of using social network data to induce behavioural change. In each case a central planner designs optimal targeted intervention strategies to improve welfare in the system. Agents in a network have incomplete information about their location and actions have complementarities. We test mechanisms that intend to increase actions in those who take a risky decision to encourage coordination on the efficient equilibrium. Chapter 1 investigates the effectiveness of a bonus that is conditional on the participation of a minimum number of agents in the network. We find that a bonus indeed increases actions, but that choice of condition is critical to achieve efficient coordination. A tough condition does not improve outcomes, but an easier condition is highly effective. A budget-friendly alternative where the central planner aims the bonus at a subset of the network population is less effective, but a more sophisticated mechanism design could achieve the same result. Chapter 3 is a pilot study that examines the effectiveness of a temporary discount that is available only
to a subset of the network population. We have two treatments that differ only in the network formation process. In each case the discount successfully shifts coordination to the efficient equilibrium, but the effect does not persist after the discount expires. Nevertheless, we observe some significant differences between treatments.

Chapter 2 investigates the role of network structure in rumour propagation and the effect on outcomes. We study an environment with biased agents and unbiased agents. Biased agents want to enforce a specific decision and unbiased agents seek the truth. One agent learns the true state and has the option to create a message to send to her neighbours. We have a network treatment where agents decide whether to further transmit the message or not, and a broadcast treatment where all agents receive the message simultaneously. Contrary to predictions, outcomes are the same in either treatment. However, the behaviours that underpin the outcomes differ between treatments. In the broadcast treatment the message creation decisions of the biased agent are important and in the network treatment the transmission decisions of unbiased agents play a role.
CHAPTER 1:

Network Intervention: Incentives with Conditions

1.1 Introduction

Network intervention describes the process of using social network data to accelerate behavioural change or improve organisational performance (Valente, 2012). Corporations and institutions are particularly suitable environments for network intervention because of the detailed knowledge they have of their internal frameworks and the control they have over their rules. Employees are arranged in an organisational structure, and a socially interactional network of communication is formally defined. The actions of one employee affect those of other employees and can set off cascades of behaviour that ultimately lead to one outcome or another. Cascades that follow undesired actions lead to coordination failure which can cause organisations to settle into states that are inefficient and unsatisfactory for all involved, even though better outcomes are possible and would be stable if achieved. Even if the benefits of improved coordination are obvious, any mechanism designed to bring about a positive shift faces considerable obstacles.

In a modern take on a classic example, imagine a professional services firm introducing a new file sharing technology to improve efficiency. Workers add their files to a central repository so that colleagues can find
information that otherwise would need to be requested or perhaps would be undiscovered. If only one worker adds her files, she wastes her efforts because she does not receive the benefit of access to the files of others. If, however, a worker can be reasonably sure that her colleagues make the effort to add files, she should be willing to do the same. To decide whether to use the new technology, the worker needs to estimate the efforts of her colleagues and factor in the dynamic with her own effort. A worker who connects with many colleagues is more likely to benefit from adopting the new technology than one who engages with few, although the ultimate success of the repository depends also on the latter group. In this paper we study controlled laboratory experiments that simulate this type of dilemma in a network environment. Our goal is to explore the effects of conditional financial incentives on collective outcomes.

The coordination problem described above is of practical importance because it occurs in several economic settings. Our research is motivated by the problem of efficient coordination in organisations, but similar issues also play a role in political economics. Consider, for example, the problem of promoting a healthy lifestyle. Healthy behaviours are sustained more easily by those that interact with many others who engage in the same behaviour (see, e.g., Christakis and Fowler, 2009), but the aggregate level of health in society depends also on those on the periphery who connect with few. The policy question in this context is how to mobilise those on the periphery to participate in healthy lifestyles.
We study a repeated game of complements with incomplete information in a laboratory experiment. Subject are in a fixed network structure with 20 positions and each position has between one and four connections. A large proportion of positions is located on the periphery having only one connection. Knowing only their number of connections, players choose to participate or not. The choice to participate is a risky one for players on the periphery. We then systematically study the effect of a conditional premium incentive. The key variables are the condition and the premium. The condition is a minimum participation rate across the network and the premium is a fixed bonus that is paid to participating players when the condition is met. We have four treatments: a baseline treatment and three treatments with a bonus. The bonus is available either to all players (universal), or to players on the periphery only (targeted). Two treatments have a bonus that is available to all players and we vary the condition. One treatment requires full network activity and the other requires a minimum participation rate of 70 percent across the network. The latter condition also applies in the third treatment, where the bonus is available to periphery players only. We then focus on three questions: (1) Can a conditional bonus incentivise players on the periphery to take the risky decision to participate? (2) Does the choice of the condition affect collective outcomes? And, (3) Can a targeted bonus yield the same results as a universal bonus?

The experimental data yield some interesting answers to these questions. Firstly, a conditional bonus increased the average probability of
participation for players on the periphery, as well as for other players in the network. Secondly, the choice of condition greatly affected collective outcomes when the bonus is available to all players. In the treatment that required full network participation, collective outcomes were not discernibly different from those in the baseline. By contrast, the easier condition of 70 percent participation led to near-perfect coordination on the efficient outcome, although not in every group. Thirdly, a bonus aimed at periphery players only was less effective than a bonus available to all player, but the mechanism has potential.

The rest of the paper is organised as follows. Section 2 reviews related literature. Section 3 describes the experimental design and sets out the theoretical analysis. Section 4 states the research hypotheses and section 5 presents the main experimental findings. Finally, section 6 concludes.

1.2 Related literature

A handful of papers study workplace incentive schemes that are conditional on collective efforts. Knez and Simester (2001) present evidence that a firm-wide incentive scheme introduced at Continental Airlines in 1995 raised employee performance after a sustained period of failure. The scheme promised a $65 bonus to every hourly employee in every month that Continental Airline’s on time-performance ranked in the industry’s top 5. Continental Airlines, like other major US airlines, has large numbers of autonomous work groups and this structure helped the
selection of higher efforts by employees due to mutual monitoring within work groups. Furthermore, interdependencies between work groups increased individual groups’ efforts because of their effect on overall firm performance, but it also risked reinforcement of low efforts. Ultimately, the scheme successfully contributed to restoring Continental Airlines to a position of profit. Motivated by the same problem, Che and Yoo (2001) explore theoretically how the design of explicit incentives interacts with the implicit incentives generated by repeated interactions. They show that the optimal incentive scheme uses low-powered group incentives, e.g. a bonus, to leverage interdependencies between agents and between teams.

The economic perspective on social network analysis contains a growing literature that is concerned with the externalities that underlie interdependencies (see Jackson, Rogers and Zanou (2016) for a comprehensive overview). An emerging literature focuses on targeted interventions in networks. Galeotti, Golub and Goyal (2019) and Demange (2017) are closest related to the current paper. Both papers consider a central planner who designs optimal targeted intervention strategies in network environments. Demange (2017) shows that when complementarities in actions are linear, the planner’s optimal strategy to increase actions is to allocate a bonus to agents with the maximal Katz-Bonacich index. Galeotti, Golub and Goyal (2019) show that in the case of strategic complements, a planner maximises welfare by targeting the top single principal component, which is determined by diagonalising the adjacency matrix of interactions. The current paper differs from these
papers in that it studies the effects of incentivising agents on the periphery rather than agents that are central by some definition.

Studies on networks that are relevant to the experimental design in this paper are Galeotti et al. (2010) and Charness, Feri, Meléndez-Jiménez and Sutter (2014). Galeotti et al. (2010) propose a theoretical framework to analyse strategic interactions in networks when local network structure affects payoffs. A key feature is the use of incomplete information: individuals know the structure of the network they are in, but not their exact location. The design limits the strategy space, which would be very large in a complete information setting. Charness et al. (2014) apply the framework and implement a range of specific examples of general network structures to test experimentally for the effects of network structure on equilibrium selection in strategic games. We use one of their game designs as a starting point for our experiment.

The current paper contributes to the experimental literature on network games. To the best of our knowledge, we are the first to use network characteristics to experimentally test for (i) the effect of a conditional universal bonus on periphery players and on collective outcomes, and (ii) the effect of a targeted bonus when compared to a universal bonus.
1.3 Experimental design

We describe the baseline game and the details of each treatment in section 1.3.1. We set out the equilibrium analysis in section 1.3.2 and finally we review the role of risk dominance in section 1.3.3.

1.3.1 Game and treatments

The game in the baseline treatment is described as follows (a copy of the written instructions provided to participants is included in Appendix A.4). In each session, a group of 20 subjects interact for 40 periods in the network shown in Figure 1.1. The network is exogenously generated and remains fixed over time. It has 20 positions (nodes) and each position is connected by a line (link) to one, two, three or four neighbours. Nodes that are linked affect each other’s payoffs. We refer to the number of neighbours that affect a player’s payoff as her degree. The network and its characteristics are common knowledge.

Figure 1.1: The network and characteristics

<table>
<thead>
<tr>
<th>Degree</th>
<th># Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>8</td>
</tr>
<tr>
<td>degree = 2</td>
<td>6</td>
</tr>
<tr>
<td>degree = 3</td>
<td>4</td>
</tr>
<tr>
<td>degree = 4</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>
At the start of each period, a player is assigned to one of the 20 nodes. She is privately informed of her degree but not of her exact position. We illustrate with an example. Say, she learns that she has four neighbours. She knows that she can be in position 4 or that she can be in position 16, but she does not know which one. She now faces a choice to be active (e.g. adopt the new technology) or to be inactive (e.g. not adopt the new technology). Her individual payoff depends on her own choice and on the choices of her neighbours. If she chooses to be active, she earns 33.33 points times the number of neighbours who are also active. If she chooses to be inactive, she earns 50 points regardless of the choices of her neighbours. After she submits her choice, she is privately informed of the position she was in, how many of her neighbours chose to be active, how many of her neighbours chose to be inactive, and her payoffs. Before the start of the next period, she is presented with a history table that shows for each previous period her degree, her position, her choice, her neighbours' choices and her payoffs. At the end of 40 periods, three periods are selected at random for payment. Points earned in those three periods are converted to cash at a rate of 40 points = 1 GBP. Finally, earnings from playing the game are added to the show-up fee of GBP 4,00 for the final payment.

The games with the conditional premium are identical except for the payoffs. In the universal premium treatments, any player who chooses to be active receives 33.33 bonus points in every period when at least \( \bar{x} \) players are active. In the treatment that requires full network activity \( x = 20 \) and in the treatment that requires 70 percent network activity \( x = 14 \).
In the targeted premium treatment, a player with degree 1 who chooses to be active receives 33.33 bonus points in every period when $x = 14$. Note that, to receive the bonus in any treatment, a player needs to be active herself and thus that the design does not allow free riding. At the end of each period, players also receive private information about the number of players that chose to be active in the network, whether the condition was met and if the bonus is paid.

In summary, we consider four treatments in the same 20-person network: (i) a baseline treatment ($BL$); (ii) a universal premium treatment with a condition of 20 active players ($TH20$); (iii) a universal premium treatment with a condition of 14 active players ($TH14$); and, (iv) a targeted premium treatment where the bonus is available only to players who have one connection and $x = 14$ ($TH14\_D1$).

We conducted four sessions of each treatment, summing to a total of 16 sessions involving 320 unique subjects. Participants were recruited from the student population of Royal Holloway, University of London. The sessions were held at the University’s ExpReSS Lab using z-Tree software for computerised economic experiments (Fishbacher, 2007). Subjects were not allowed to participate in more than one session of the experiment. Earnings averaged around £10 per subject for a 1-hour session.
1.3.2 Equilibrium predictions

The network is undirected. That is, a player is affected by her neighbours’ actions and her actions also affect her neighbours. The game has strategic complements because the marginal payoff from action 1 increases when more players take the same action. The baseline game has local interactions since a player’s payoffs are affected only by her neighbours. The universal premium games also have a global interaction as a player’s payoffs are affected by the choices of all players in the network. The global interaction exists only for players with degree 1 in the targeted premium game.

All games have multiple equilibria. Our equilibrium analysis considers only pure strategy Bayesian Nash Equilibria (BNE). That is, all players of the same degree take the same action. Charness et al. (2014) show that the baseline game has two equilibria: (i) a secure equilibrium, where all players choose to be inactive, and (ii) an efficient equilibrium, where players with degree 1 choose to be inactive and players with degrees 2, 3 and 4 choose to be active. The universal premium games have the same two equilibria, and a third equilibrium where all players are active. The third equilibrium is payoff dominant due to the bonus.

**Proposition 1:** The universal premium games have three equilibria, (i) a secure equilibrium, where players of all degrees choose the inactive action, (ii) an equilibrium where players with
degree 1 choose the inactive action and players with higher degree choose the active action, and (iii) a payoff dominant equilibrium, where players of all degrees choose the active action.

The targeted premium game has the same three equilibria as the universal premium games. But the payoff dominant equilibrium has lower payoffs for players with degrees 2, 3 and 4 since the bonus is available only to players with degree 1. Nevertheless, it remains the payoff dominant equilibrium for all players due to the activity of players with degree 1. Lastly, the targeted premium treatment has an inefficient equilibrium where players with degree 2 choose the inactive action and players with degrees 1, 3 and 4 choose the active action.

**Proposition 2:** The targeted premium treatment has four equilibria, (i) a secure equilibrium, where players of all degrees choose the inactive action, (ii) an equilibrium where players with degree 1 choose the inactive action and players with higher degree choose the active action, (iii) an equilibrium where players with degree 2 choose the inactive action and players with all other degrees choose the active action, and (iv) a payoff dominant equilibrium, where players of all degrees choose the active action.
The proof for Propositions 1 and 2 follows next.

**Proof.**

Definitions and Part 1 are replicated from Charness et al. (2014).

A player knows only her degree, so she can only condition her behaviour on this information when she chooses to be active (action 1) or inactive (action 0). Thus, a symmetric strategy profile is represented by the vector \( s = (s_1, s_2, s_3, s_4) \), where \( s_j \in \{0, 1\} \) is the action chosen by an agent with degree \( j \in \{1, 2, 3, 4\} \).

Let \( \pi_i^j(x_i, x_{-i}) \equiv \pi_i^j(x_i) \) be the payoff of an agent \( i \in N \equiv \{1, 2, \ldots, 20\} \) with degree \( j \in \{1, 2, 3, 4\} \) from action \( x_i \in \{0, 1\} \) when other players choose action \( x_{-i} \).

Note that strategy \((0, 0, 0, 0)\) is a strict Bayes-Nash equilibrium in any case because a deviation to action 1 yields a payoff of 0, against a payoff of 50 for action 0. In what follows, we consider the remaining strategy profiles separately (i) for the baseline game, (ii) for TH14, (iii) for TH14_D1, and (iv) for TH20.

Part (i). For the baseline game there are seven strategy profile candidates to be a pure-strategy Bayes-Nash equilibrium: all possible combinations of \( s_2, s_3, s_4 \) in \( s \) excluding \((0, 0, 0, 0)\). Strategy \((0, 1, 0, 0)\) is not an equilibrium because \( \pi_i^2(1, x_{-i}) = 22.22 < 50 = \pi_i^2(0, x_{-i}) \). Strategy
(0, 0, 1, 0) is not an equilibrium because \( \pi_i^3(1, x_{-i}) = 16.66 < 50 = \pi_i^3(0, x_{-i}) \). Strategy (0, 0, 0, 1) is not an equilibrium because \( \pi_i^4(1, x_{-i}) = 0 < 50 = \pi_i^4(0, x_{-i}) \). Strategy (0, 1, 1, 0) is not an equilibrium because \( \pi_i^4(1, x_{-i}) = 116.66 > 50 = \pi_i^4(0, x_{-i}) \). Strategy (0, 1, 0, 1) is not an equilibrium because \( \pi_i^3(1, x_{-i}) = 44.44 < 50 = \pi_i^3(0, x_{-i}) \). Strategy (0, 0, 1, 1) is not an equilibrium because \( \pi_i^3(1, x_{-i}) = 41.66 < 50 = \pi_i^3(0, x_{-i}) \). Finally, strategy (0, 1, 1, 1) is a strict equilibrium because \( \pi_i^4(1, x_{-i}) = 116.66 > 50 = \pi_i^4(0, x_{-i}) \), \( \pi_i^3(1, x_{-i}) = 58.33 > 50 = \pi_i^3(0, x_{-i}) \), \( \pi_i^2(1, x_{-i}) = 55.55 > 50 = \pi_i^2(0, x_{-i}) \) and \( \pi_i^1(0, x_{-i}) = 50 > 33.33 = \pi_i^1(1, x_{-i}) \).

Part (ii). Now consider TH14. There are five additional candidates to be a pure-strategy Bayes-Nash equilibrium: all possible combinations of \( s_1, s_2, s_3, s_4 \) in \( s \) where 14 of more agents choose action 1 in symmetric strategies. Strategy (1, 1, 0, 0) is not an equilibrium because \( \pi_i^4(1, x_{-i}) = 116.66 > 50 = \pi_i^4(0, x_{-i}) \). Strategy (1, 1, 1, 0) is not an equilibrium because \( \pi_i^4(1, x_{-i}) = 166.66 > 50 = \pi_i^4(0, x_{-i}) \). Strategy (1, 1, 0, 1) is not an equilibrium because \( \pi_i^3(1, x_{-i}) = 116.66 > 50 = \pi_i^3(0, x_{-i}) \). Strategy (1, 0, 1, 1) is not an equilibrium because \( \pi_i^2(1, x_{-i}) = 77.77 > 50 = \pi_i^2(0, x_{-i}) \). Finally, strategy (1, 1, 1, 1) is a strict equilibrium because \( \pi_i^4(1, x_{-i}) = 166.66 > 50 = \pi_i^4(0, x_{-i}) \), \( \pi_i^3(1, x_{-i}) = 133.33 > 50 = \pi_i^3(0, x_{-i}) \), \( \pi_i^2(1, x_{-i}) = 99.99 > 50 = \pi_i^2(0, x_{-i}) \) and \( \pi_i^1(1, x_{-i}) = 66.66 > 50 = \pi_i^1(0, x_{-i}) \).
Part (iii). Now consider TH14_D1. There are five additional candidates to be a pure-strategy Bayes-Nash equilibrium: all possible combinations of \(s_1, s_2, s_3, s_4\) in \(s\) where 14 of more agents choose action 1 in symmetric strategies. Strategy \((1,1,0,0)\) is not an equilibrium because \(\pi_1^4(1,x_{-i}) = 83.33 > 50 = \pi_1^4(0,x_{-i})\). Strategy \((1,1,1,0)\) is not an equilibrium because \(\pi_1^4(1,x_{-i}) = 133.33 > 50 = \pi_1^4(0,x_{-i})\). Strategy \((1,1,0,1)\) is not an equilibrium because \(\pi_1^3(1,x_{-i}) = 83.33 > 50 = \pi_1^3(0,x_{-i})\). Finally, strategy \((1,0,1,1)\) is a strict equilibrium because \(\pi_1^4(1,x_{-i}) = 66.66 > 50 = \pi_1^4(0,x_{-i}), \pi_1^3(1,x_{-i}) = 83.33 > 50 = \pi_1^3(0,x_{-i}), \pi_1^2(1,x_{-i}) = 44.44 < 50 = \pi_1^2(0,x_{-i}), \pi_1^1(1,x_{-i}) = 58.33 > 50 = \pi_1^1(0,x_{-i}), \) and strategy \((1,1,1,1)\) is a strict equilibrium because \(\pi_1^4(1,x_{-i}) = 133.33 > 50 = \pi_1^4(0,x_{-i}), \pi_1^3(1,x_{-i}) = 99.99 > 50 = \pi_1^3(0,x_{-i}), \pi_1^2(1,x_{-i}) = 66.66 > 50 = \pi_1^2(0,x_{-i}), \pi_1^1(1,x_{-i}) = 66.66 > 50 = \pi_1^1(0,x_{-i})\).

Part (iv). Finally consider TH20. There is one additional candidate to be a pure-strategy Bayes-Nash equilibrium in TH20: the only combination of where 20 agents in \(s = (s_1,s_2,s_3,s_4)\) choose action 1 in symmetric strategies. Strategy \((1,1,1,1)\) is a strict equilibrium because \(\pi_1^4(1,x_{-i}) = 166.66 > 50 = \pi_1^4(0,x_{-i}), \pi_1^3(1,x_{-i}) = 133.33 > 50 = \pi_1^3(0,x_{-i}), \pi_1^2(1,x_{-i}) = 99.99 > 50 = \pi_1^2(0,x_{-i}), \) and \(\pi_1^1(1,x_{-i}) = 66.66 > 50 = \pi_1^1(0,x_{-i})\).

QED.
Note that the efficient equilibrium varies between games. In the baseline game the efficient equilibrium is \((0, 1, 1, 1)\) and in the conditional premium games the efficient equilibrium is \((1, 1, 1, 1)\). In what follows we refer to \((0, 0, 0, 0)\) as the *secure equilibrium* and to \((1, 1, 1, 1)\) as the *payoff dominant equilibrium*. We call \((0, 1, 1, 1)\) intermediate 1 and \((1, 0, 1, 1)\) intermediate 2. Table 1.1 summarises in the top panel the equilibrium actions and in the bottom panel the games where the equilibrium exists. An asterisk indicates that the equilibrium is the efficient equilibrium in that game.

Table 1.1: Equilibrium actions

<table>
<thead>
<tr>
<th>Equilibrium</th>
<th>Secure Action</th>
<th>Intermediate 1 Action</th>
<th>Intermediate 2 Action</th>
<th>Payoff Dominant Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>degree = 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>degree = 3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>degree = 4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

In game

<table>
<thead>
<tr>
<th></th>
<th>BL</th>
<th>BL*</th>
<th>TH14</th>
<th>TH14*</th>
<th>TH20</th>
<th>TH20*</th>
<th>TH14_D1</th>
<th>TH14_D1*</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH14</td>
<td></td>
<td></td>
<td>-</td>
<td>TH14*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH20</td>
<td></td>
<td></td>
<td>-</td>
<td>TH20*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH14_D1</td>
<td></td>
<td></td>
<td>TH14_D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH14_D1*</td>
<td></td>
<td></td>
<td>TH14_D1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1.3.3 Risk dominance

Charness et al. (2014) apply the concept of generalised-risk dominance proposed by Peski (2010) to proof that none of the equilibria in the baseline game are ordinal generalised-risk dominant (ordinal GR-dominant). We follow their approach to proof that equilibria in the conditional premium games are also not ordinal GR-dominant.

**Proposition 3:** There is no pure strategy Bayesian Nash equilibrium that is ordinal generalised-risk dominant.

**Proof.**
Definitions, Part 1 and Part 2 are provided by Charness et al. (2014).

Let $N$ be the set of players and let $a = (a_1, \ldots, a_n)$ be an action (strategy) profile. Players can have maximum degree 4, and for each player $i \in N$, $a_i = (a_{i,1}, a_{i,2}, a_{i,3}, a_{i,4}) \in \{0,1\}$. For each $k \in \{1, 2, 3, 4\}$, $a_{i,k} = 0$ ($a_{i,k} = 1$) represents the choice of inactive (active) in the event in which player $i$ has degree $k$.

Definition 1: Given an action profile $a$, two action profiles, $\eta$ and $\bar{\eta}$, are $a$-associated if, for each $i \in N$, either $\eta_i = a_i$ or $\bar{\eta}_i = a_i$.
Definition 2: An action profile \( a \) is ordinal \( GR \)-dominant if, for each player \( i \), and for each pair of \( a \)-associated action profiles, \( \eta \) and \( \bar{\eta} \), \( a_i \) is a best response of player \( i \) to either \( \eta \) or \( \bar{\eta} \).

Part 1. We first show that the action profile \( a \) such that for each \( i \in N \), \( a_i = (0, 1, 1, 1) \) is not ordinal \( GR \)-dominant. Consider, without loss of generality, player 1. There exist two \( a \)-associated action profiles, \( \eta \) and \( \bar{\eta} \), such that for each \( i \in \{2, \ldots, 10\} \), \( \eta_i = (0, 0, 0, 0) \) and \( \bar{\eta}_i = (0, 1, 1, 1) \), and for each \( j \in \{11, \ldots, 20\} \), \( \eta_j = (0, 1, 1, 1) \) and \( \bar{\eta}_j = (0, 0, 0, 0) \). Consider the event in which \( n_1 = 2 \). Clearly, if \( a_{1,2} = 1 \) is not a best response to profile \( \eta \), it will neither be to profile \( \bar{\eta} \), since the latter has one more player choosing the full-inactivity strategy (10 vs. 9 players) and all the remaining players choose the same strategy \( (0, 1, 1, 1) \) (recall that all players are randomly allocated in the network with uniform probability. Thus, to prove the result is suffices to show that \( a_{1,2} = 1 \) is not a best response to profile \( \eta \).

To this aim, consider a profile \( \eta' \) such that, for each \( i \in \{2, \ldots, 10\} \), \( \eta'_i = (0, 0, 0, 0) \), and for each \( j \in \{11, \ldots, 20\} \), \( \eta'_j = (1, 1, 1, 1) \). In profile \( \eta' \) there are ten players in \( N \setminus \{1\} \) that are always active (regardless of their degree), and nine players in \( N \setminus \{1\} \) that are always inactive. It is straightforward to see that if \( a_{1,2} = 1 \) is not a best response to profile \( \eta' \), it cannot be a best response to profile \( \eta \). Hence, it suffices to prove that \( a_{1,2} = 1 \) is not a best response to profile \( \eta' \). Under profile \( \eta' \), when \( n_1 = 2 \), the probability that player 1 has \( k \in \{0, 1, 2\} \) active neighbours is \( p'_k = \frac{\binom{10}{k} \cdot \binom{9-k}{2}}{\binom{19}{2}} \). Thus, in such a case, the expected payoff to player 1 by choosing \( a_{1,2} = 1 \) is \( \sum_{k=0}^{2} p'_k \cdot k \cdot \frac{100}{3} = \)
35.1 < 50. Thus, \( a_{1,2} = 1 \) is not a best response to profile \( \eta' \). It follows that \( a_{1,2} = 1 \) (and, therefore, \( a_1 \)) is neither a best response to \( \eta \) nor to \( \tilde{\eta} \) and, thus, \( a \) is not ordinal GR-dominant.

Part 2. We now show that the action profile \( a \) such that for each \( i \in N \), \( a_i = (0, 0, 0, 0) \) is not ordinal GR-dominant. Consider, without loss of generality, player 1. There exist two \( a \)-associated action profiles, \( \eta \) and \( \tilde{\eta} \), such that for each \( i \in \{2, \ldots, 10\} \), \( \eta_i = (1, 1, 1, 1) \) and \( \tilde{\eta}_i = (0, 0, 0, 0) \), and for each \( j \in \{11, \ldots, 20\} \), \( \eta_j = (0, 0, 0, 0) \) and \( \tilde{\eta}_j = (1, 1, 1, 1) \). In profile \( \eta \) (profile \( \tilde{\eta} \)) there are nine (ten) players in \( N \setminus \{1\} \) that are always active (regardless of their degree), and ten (nine) players in \( N \setminus \{1\} \) that are always inactive. Consider the event in which \( n_1 = 4 \). Clearly, if \( a_{1,4} = 0 \) is not a best response to profile \( \eta \), it will neither be to profile \( \tilde{\eta} \) (recall that all players are randomly allocated in the network with uniform probability. Thus, to prove the result is suffices to show that \( a_{1,4} = 0 \) is not a best response to profile \( \eta \).

Under profile \( \eta \), when \( n_1 = 4 \), the probability that player 1 has \( k \in \{0, 1, 2, 3, 4\} \) active neighbours is \( q'_k = \frac{\binom{2}{k} \binom{10}{4-k}}{\binom{12}{4}} \). Thus, in such a case, the expected payoff to player 1 by choosing \( a_{1,4} = 0 \) (i.e. 50) is lower than the expected value he would get by choosing action 1, i.e. \( \sum_{k=0}^{4} q'_k \cdot k \cdot \frac{100}{3} = 63.2 \).

It follows that \( a_{1,4} = 0 \) (and, therefore, \( a_1 \)) is neither a best response to \( \eta \) nor to \( \tilde{\eta} \) and, thus, \( a \) is not ordinal GR-dominant.

Part 3. We now show that the action profile \( a \) such that for each \( i \in N \), \( a_i = (1, 1, 1, 1) \) is not ordinal generalised-risk dominant. Consider, without
loss of generality, player 1. There exist two \( \alpha \)-associated action profiles, \( \eta \) and \( \bar{\eta} \), such that for each \( i \in \{2, \ldots, 10\} \), \( \eta_i = (0, 0, 0, 0) \) and \( \bar{\eta}_i = (1, 1, 1, 1) \), and for each \( j \in \{11, \ldots, 20\} \), \( \eta_i = (1, 1, 1, 1) \) and \( \bar{\eta}_i = (0, 0, 0, 0) \). In profile \( \eta \) (profile \( \bar{\eta} \)) there are ten (nine) players in \( N\setminus\{1\} \) that are always active (regardless of their degree), and nine (ten) players in \( N\setminus\{1\} \) that are always inactive. The condition of 14 active players is not met in either profile \( \eta \) or profile \( \bar{\eta} \) and therefore the bonus is not received. Consider the event in which \( \eta_1 = 1 \). Clearly, if \( a_{1,1} = 1 \) is not a best response to profile \( \eta \), it will neither be to profile \( \bar{\eta} \) since profile \( \eta \) has a higher average activity rate than profile \( \bar{\eta} \) and therefore profile \( \eta \) has a higher expected payoff than profile \( \bar{\eta} \) (recall that all players are randomly allocated in the network with uniform probability). Thus, to prove the result it suffices to show that \( a_{1,1} = 1 \) is not a best response to profile \( \eta \). Under profile \( \eta \), when \( \eta_1 = 1 \), the probability that player 1 has \( k \in \{0, 1\} \) active neighbours is \( r_k = \binom{10}{k} \cdot \frac{1}{9} \cdot \left(1 - \frac{1}{9}\right)^{10-k} \). Thus, in such a case, the expected payoff to player 1 by choosing \( a_{1,1} = 1 \) is

\[
\sum_{k=0}^{1} r_k \cdot k \cdot \frac{100}{3} = 17.54 < 50.
\]

Thus, \( a_{1,1} = 1 \) is not a best response to \( \eta \). It follows that \( a_{1,1} = 1 \) (and therefore \( a_i \)) is neither a best response to \( \eta \) nor to \( \bar{\eta} \) and, thus, \( a \) is not ordinal GR-dominant.

Part 4. We lastly show that the action profile \( a \) such that for each \( i \in N \), \( a_i = (1, 0, 1, 1) \) is not ordinal GR-dominant. Consider, again, the event in which \( \eta_1 = 1 \) and \( a_{1,1} = 1 \). The case is the same as in Part 3 and the proof is identical.

QED.
1.4 Research hypotheses

Each of our games has multiple equilibria and theory is silent on the issue of equilibrium selection. We form three behavioural research hypotheses concerning the effects of a conditional premium on activity rates and on coordination.

Firstly, it is never profitable for a player with degree 1 to choose the active action in the baseline game. Recall that the choice to be active produces a maximal payoff potential of 33.33 points, while the choice to be inactive results in 50 points for sure. In the conditional premium games, the choice to be active has a payoff of 66.66 if the condition is met. But if the condition is not met, the payoff is once again at most 33.33 points. Therefore, the choice to be active is risky for players with degree 1. Nevertheless, we predict that the conditional premium positively affects activity rates of players with degree 1 in our first hypothesis:

H1: The premium effect. The conditional premium increases the rate of active choices for players with degree 1.

Secondly, the universal premium games have the same set of three equilibria and theory does not provide guidance on which one will be selected. We resort to behavioural notions. On the one hand, the we expect that the premium effect [H1] increases the activity rates of players with degree 1, which could cascade to players with higher degrees and result in
full network activity and thus coordination on the payoff dominant equilibrium. On the other hand, a condition could present a challenging environment. The increase in the rate of active choices may not be enough to induce full network activity. But even in this case the universal premium could support coordination on the intermediate equilibrium where all players with degrees higher than 1 are active. We predict that the universal premium improves coordination, but that the condition plays a critical role in the realised efficiency.

**H2: The universal premium threshold effect.** The universal premium improves efficient coordination, but condition $x = 14$ leads to greater efficiency than condition $x = 20$.

Thirdly, we predict that the conditional premium results in coordination on the payoff dominant equilibrium when $x = 14$. Note that the active choice is profitable for players with degrees 2, 3 and 4 in any game, and that it is players with degree 1 who take a risky decision. Therefore, we expect that the conditional premium that targets only players with degree 1 has the same effect as the universal premium with the same condition $x = 14$. We capture this in our third and last hypothesis:

**H3: The conditional premium targeted at degree 1 has the same effect as the universal conditional premium with $x = 14$.**
1.5 Experimental results

We discuss our experimental results in terms of the choice to be active. First, we use non-parametric tests to compare activity levels by degree between treatments and second, we perform an econometric analysis that estimates the marginal effects of each treatment on the probability of being active by the degree at the midpoint of the game. The results of both analyses are reported in section 1.5.1. We discuss equilibrium selection in section 1.5.2 and we summarise our results in the context of our hypotheses in section 1.5.3.

1.5.1 Data analysis

The top panel of Table 1.2 reports the observed mean frequencies of being active by treatment and by degree across sessions. The data show that the mean activity rate for players with degree 1 is 5.6 percent in the baseline game, and that it is higher in each of the conditional premium games. In the universal premium games, the mean activity rate increases modestly to 10.4 percent when $x = 20$, while it is boosted to 74.2 percent when $x = 14$. In the targeted premium game, the activity rate also increases but to a lower 31.3 percent when compared to the universal premium game with $x = 14$. Activity rates for players with degrees 2 and 3 follow a similar pattern. In the universal premium games, the mean activity rate increases for players with degree 2 (degree 3) from 34.1 (60.5)
percent in the baseline game to 44.1 (73.3) percent when $x = 20$ and to 84.3 (92.8) percent when $x = 14$. In the targeted premium treatment, the mean activity rate increases for degree 2 (degree 3) to 50.4 (74.5) percent when compared with the baseline game, so we observe positive spill-over effects from greater activity by players with degree 1. Again, the increases in activity rates are largest in the universal premium game when $x = 14$ and smallest when $x = 20$. The picture is different for players with degree 4. The mean activity rate is 92.2 percent in the baseline game and with the universal premium this increases to 94.4 percent when $x = 20$ and to 99.4 percent when $x = 14$. In the targeted premium treatment, however, the mean activity rate falls to 85.6 percent. Thus, we find that a conditional premium increases the frequency of active choices for players of each degree in every treatment, except for players with degree 4 in the targeted premium treatment.
Table 1.2: Statistics by treatment and degree

Frequencies of activity by treatment and degree

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
<td>%</td>
<td>Total</td>
<td>Activity</td>
<td>%</td>
<td>Total</td>
<td>Activity</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>degree = 1</td>
<td>72</td>
<td>5.63%</td>
<td>1280</td>
<td>133</td>
<td>10.39%</td>
<td>1280</td>
<td>950</td>
<td>74.22%</td>
</tr>
<tr>
<td>degree = 2</td>
<td>327</td>
<td>34.06%</td>
<td>960</td>
<td>423</td>
<td>44.06%</td>
<td>960</td>
<td>809</td>
<td>84.27%</td>
</tr>
<tr>
<td>degree = 3</td>
<td>387</td>
<td>60.47%</td>
<td>640</td>
<td>469</td>
<td>73.28%</td>
<td>640</td>
<td>594</td>
<td>92.81%</td>
</tr>
<tr>
<td>degree = 4</td>
<td>295</td>
<td>92.19%</td>
<td>320</td>
<td>302</td>
<td>94.38%</td>
<td>320</td>
<td>318</td>
<td>99.38%</td>
</tr>
</tbody>
</table>

Wilcoxon Mann–Whitney test by degree: \( p \)-value

<table>
<thead>
<tr>
<th></th>
<th>TH14 vs Baseline</th>
<th>TH20 vs Baseline</th>
<th>TH14_D1 vs Baseline</th>
<th>TH20 vs TH14</th>
<th>TH20 vs TH14_D1</th>
<th>TH14 vs TH14_D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>0.0286**</td>
<td>0.1714</td>
<td>0.0286**</td>
<td>0.0571*</td>
<td>0.0571*</td>
<td>0.2000</td>
</tr>
<tr>
<td>degree = 2</td>
<td>0.0571*</td>
<td>0.4286</td>
<td>0.4286</td>
<td>0.2000</td>
<td>0.1000</td>
<td>0.1143</td>
</tr>
<tr>
<td>degree = 3</td>
<td>0.0286**</td>
<td>0.3429</td>
<td>0.4857</td>
<td>0.1143</td>
<td>0.8857</td>
<td>0.1143</td>
</tr>
<tr>
<td>degree = 4</td>
<td>0.0857*</td>
<td>0.4000</td>
<td>0.5429</td>
<td>0.2571</td>
<td>0.4000</td>
<td>0.0571*</td>
</tr>
</tbody>
</table>

***, **, * denote significance at 1%, 5% and 10% levels, respectively, in a two-tailed test
Next, we compare activity rates between treatments using the Wilcoxon rank sum test, where an independent observation is the average decision by degree within a twenty-player group over 40 periods of the game. We have four observations for each degree in each treatment. Since our sample size is small, we calculate exact statistics (Harris and Hardin, 2013) and results can be measured at a maximum critical value of 5 percent. The bottom panel of Table 1.2 displays outcomes in terms of $p$-values. We first consider the universal premium games. Compared with the baseline game, we do not observe a treatment effect when $x = 20$, but when $x = 14$ we find significant effects at the 5% level for degrees 1 and 3, and at the 10% level for degrees 2 and 4. Between the two universal premium games we find a statistically significant difference for players with degree 1 at the 10% level. Clearly, the choice of condition matters and affects degree 1 most. We now turn to the targeted premium game. Compared with the baseline game we observe a significant treatment effect at the 5% level for players with degree 1. In a comparison between the targeted premium game and the universal premium game with $x = 14$, we find a significant treatment effect for degree 4 at the 10% level. The targeted premium does not have the same effect as the universal premium. Mean activity rates are lower for each degree when the premium is targeted, but, surprisingly, mostly affect degree 4.
Our non-parametric tests consider only the distribution of session-level mean activity rates. A separate econometric analysis estimates the probability of being active using a logistic function with random effects. The data are arranged as a panel where a unit of observation is a single participant who is observed for 40 periods. The explanatory variables are period, dummies for treatment, dummies for degrees and all interactions between period and these dummies. We subsequently estimate the effect of each treatment on activity by calculating the marginal effects for each treatment by degree in relation to the probability measured at period 20. The marginal effects are interpreted as the change in the probability to be active. Table 1.3 reports the results (details are included in Appendix A.1 and A.2).

<table>
<thead>
<tr>
<th></th>
<th>TH14 vs Baseline</th>
<th>TH20 vs Baseline</th>
<th>TH14_D1 vs Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>0.896***</td>
<td>0.032**</td>
<td>0.214***</td>
</tr>
<tr>
<td>degree = 2</td>
<td>0.728***</td>
<td>0.177**</td>
<td>0.329***</td>
</tr>
<tr>
<td>degree = 3</td>
<td>0.352***</td>
<td>0.166**</td>
<td>0.231***</td>
</tr>
<tr>
<td>degree = 4</td>
<td>0.019**</td>
<td>0.003</td>
<td>0.022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TH20 vs TH14</th>
<th>TH20 vs TH14_D1</th>
<th>TH14 vs TH14_D1</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>-0.864***</td>
<td>-0.182***</td>
<td>0.682***</td>
</tr>
<tr>
<td>degree = 2</td>
<td>-0.551***</td>
<td>-0.152*</td>
<td>0.389***</td>
</tr>
<tr>
<td>degree = 3</td>
<td>-0.186***</td>
<td>-0.066</td>
<td>0.121***</td>
</tr>
<tr>
<td>degree = 4</td>
<td>-0.016**</td>
<td>0.025</td>
<td>0.040***</td>
</tr>
</tbody>
</table>

***, **, * denote significance at 1%, 5% and 10% levels, respectively
The top panel of Table 1.3 shows the marginal effects of each conditional premium treatment when compared with the baseline, while the bottom panel displays the marginal effects when conditional premium treatments are compared with one another. Relative to the baseline, the conditional premium with either condition increases the probability to be active for every degree. The effects for degrees 1, 2 and 3 are significant at the 1% level for both the universal premium treatment with $x = 14$ and the targeted premium treatment, and at the 5% level when $x = 20$. For degree 4, the effect is significant only for the universal premium treatment when $x = 14$. The universal premium treatment with $x = 20$ reduces the probability to be active relative to the other conditional premium treatments, except for degree 4 in the targeted premium treatment. Finally, the universal premium with $x = 14$ increases the probability to be active for all degrees relative to the targeted premium treatment. Moreover, the effects are significant at the 1% level for all degrees.

The results of our econometric analysis at period 20 are broadly aligned with our non-parametric findings. However, in terms of statistical significance, our findings appear paradoxical for the comparison between the baseline and the universal premium with $x = 20$, and for the comparison between the baseline and targeted premium. At this stage it is instructive to look Figure 1.2 below.
Figure 1.2: Observed activity rates by treatment and degree

Figure 2.1 displays for each game the evolution of decisions over the 40 periods of play at the aggregate level. In the baseline game we observe decay in the activity rates for each degree over time. The patterns are similar for the universal premium game when $x = 20$ and for the targeted premium game, except for degree 1 in the last case. Note, however, that initial activity rates are higher in the conditional premium games than in the baseline. Hence, the distributions of activity rates are similar between the conditional premium games and the baseline, and merely shifted upwards at the midpoint of the game.
1.5.2 Equilibrium selection

After 40 periods of play decisions may or may not have reached a steady state. Figure 1.2 is informative for the trajectory of choices and therefore indicative of equilibrium selection. In the baseline game we observe little activity for degree 1 from the start and declining activity rates for degrees 2, 3 and 4. Ultimately, play converges to the secure equilibrium without activity. The result matches that obtained by Charness et al. (2014) for the same game. Outcomes contrast between the two universal premium games. When $x = 20$, activity rates decline in a similar way to those the baseline although they start from a higher base. Conversely, when $x = 14$ activity rates are high from the outset and maintained throughout the game. Thus, play converges to the secure equilibrium when $x = 20$, but to the payoff dominant equilibrium when $x = 14$. Activity rates in the targeted premium game do not resemble those in the universal premium game with $x = 14$, which is contrary to our predictions. We note that the activity rate of degree 4 declines earlier than in any other game. The activity rate of degree 1 seems to stabilise around 25 percent in the second half of the game, but we expect that ultimately play converges to the secure equilibrium.

We observe heterogeneity across groups in the universal premium game with $x = 14$ and in the targeted premium game. We do not identify the cause of the heterogeneity, but the literature offers several explanations for different outcomes in groups even when individual preferences are the
same in both groups (e.g. see Schelling, 1971; Granovetter, 1978). In the universal premium game with $x = 14$, the payoff dominant equilibrium was played 83.5% of the time on aggregate. Three groups coordinated on the payoff dominant equilibrium, but in the fourth group play converged to the secure equilibrium (see appendix A.3 for details at the group level). In the last (first) 10 periods, the mean activity rate across the three efficient groups was 98.2 (94.8) percent and for the fourth groups it was only 29.5 (57.5) percent. In the targeted premium game, play converged to the secure equilibrium in three groups, and the fourth group selected the payoff dominant equilibrium. In the last (first) 10 periods, the mean activity rate across the three inefficient groups was 22.7 (55) percent, but in the fourth group it was 88.5 (79) percent. The activity rates in the first 10 periods are key for equilibrium selection in both games. Moreover, groups coordinate on the payoff dominant equilibrium when players with degree 1 maintain a mean activity rate above 50 percent across the first 10 periods.

1.5.3 Results

We predict that the conditional premium increases the rate of active choices for players with degree 1 when compared to the baseline. Our data show that activity rates increase in each premium game and for all degrees, except for degree 4 in the targeted premium game. In a Wilcoxon rank sum test, the effects are significant for degree 1 at the 5% level with
the universal premium when $x = 14$ and with the targeted premium. Our econometric analysis shows that, relative to the baseline game, the probability for degree 1 to be active increases by 89.6 percent with the universal premium when $x = 14$, by 21.4 percent with the targeted premium, and by 3.2 percent with the universal premium when $x = 20$. Our data provide support for the premium effect [H1].

**Result 1:** The conditional premium increases the rate of active choices for players with degree 1. The effect is significant for the universal premium when $x = 14$ and for the targeted premium. The effect is larger for the universal premium than for the targeted premium.

We predict that players coordinate more efficiently in the universal premium games relative to the baseline games, but that the condition $x = 14$ leads to a more efficient outcome relative to the condition $x = 20$. We find that play converges to the payoff dominant equilibrium when $x = 14$ and that the secure equilibrium is selected when $x = 20$. Indeed, the universal premium with $x = 14$ leads to a more efficient outcome than with $x = 20$. It is, however, somewhat surprising that the latter condition did not in any group aid discovery of the intermediate equilibrium where degrees 2, 3 and 4 are active. Nevertheless, we find support for the universal premium threshold effect [H2].
**Result 2:** The universal premium leads to efficient coordination only when \( x = 14 \). When \( x = 20 \) play converges to the same outcome as without a premium.

Finally, we predict that the targeted premium has the same effect as the universal premium with \( x = 14 \). Activity rates are lower for each degree with the targeted premium compared to the universal premium with \( x = 14 \). The effect is significant in a Wilcoxon rank sum test (10% level) for, surprisingly, players with degree 4. The probabilities to be active at the midpoint of the game are also lower for each degree. At the aggregate level, play converges to the secure equilibrium with the targeted premium, and to the payoff dominant equilibrium with the universal premium. We note that, in either treatment, the payoff dominant equilibrium is selected if players with degree 1 choose to be active for at least half of the time in the early stage of the game. We observe heterogeneity across groups in either game but, overall, we do not find evidence to support our hypothesis that the targeted premium has the same effect as the universal premium [H3].

**Result 3:** Activity rates for all degrees are lower with the targeted premium than with the universal premium when \( x = 14 \). Play converges to the secure equilibrium with the targeted premium and to the payoff dominant equilibrium with the universal premium when \( x = 14 \).
**Result 4:** Play converges to the payoff dominant equilibrium in either treatment if players with degree 1 choose to be active at least half of the time in the early stage of the game.

### 1.6 Conclusion

Networks are ubiquitous in the social and economic landscape. The question of efficient network intervention is a vital one for central planners involved with business decisions and government policy alike. We conduct a series of experiments to test incentive mechanisms that take account of the network structure and characteristics. We find that a conditional bonus can motivate agents on the periphery to take a risky decision, but that the choice of the condition is key if efficient coordination is to be achieved. If the condition is practicable, play converges to the payoff dominant equilibrium where all participate. But if the condition is too difficult, play settles on the secure equilibrium without activity. A result that is the same as in the absence of a bonus. Theoretically, the condition $x = 14$ perhaps appears easy. To move from the intermediate equilibrium where all players with more than one connection are active to the payoff dominant equilibrium requires just two periphery players to change from being inactive to being active. But we do not observe coordination on this intermediate equilibrium in any treatment or in any group. Behaviourally,
then, a shift to the payoff dominant equilibrium is challenging and unlikely to be the result of some random perturbation.

We find that the targeted bonus aimed at agents on the periphery alone does not lead to the same outcome as the universal bonus with the same condition. But it is evident that participation from agents on the periphery is key for selection of the payoff dominant equilibrium. An adjusted design that better mitigates the risk of incurring costs for these agents could prove successful. For example, agents on the periphery could be compensated if the choice to be active resulted in a loss. In this case, the budget required for a successful targeted intervention would still be lower than that for a universal intervention.

The existing literature on network intervention focuses attention on positions that are central by some definition (e.g. Galeotti, Golub and Goyal, 2019; Demange, 2017; Bramoullé, Kranton and D’Amours, 2014). In our games, central positions tend to be active at the start, but activity does not spread to agents on the periphery and subsequently declines in central nodes. Instead, we observe that it is activity of agents on the periphery that cascades towards the centre and stimulates continued activity in central positions. We hope that our results inspire others to also research the role of the periphery on coordination and outcomes.
CHAPTER 2:

Network Structure and the Efficacy of Rumours

2.1 Introduction

Social media has changed the news we consume. Traditional media technologies, such as newspapers, television and radio, broadcast news to mass audiences in one-way communication via a centralised point of editorial control. This structure contrasts sharply with social media, where news stories are shared on paths of social networks between users with just a click and without significant fact-checking or source verification. The structure of social media facilitates dissemination of “fake news”, false news stories that are intended to influence public opinion. Facebook, the largest social media platform, has 1.58 billion daily active users (Facebook, 2019) and the sheer size of the audience creates cause for concern about the social costs that fake news can impose on society.

Apprehension about the effects of fake news first spiked after the 2016 U.S. presidential election. Traditional media reported that Donald Trump owed his victory to the influence of fake news on social media (e.g. see Dewey, 2016; Parkinson, 2016; Read, 2016; Allcot and Gentzkow, 2017). Post-election day analyses show that top fake election news stories were overtly pro-Trump and generated more engagement on Facebook than the most popular news stories in traditional media (Silverman, 2016).
Facebook has 187 million users in the U.S. and Canada (Facebook, 2019) and an individual user with no credentials can reach an audience as large as Fox News or CNN.

In Europe, Facebook has 286 million in users (Facebook, 2019) and concerns are the same. During the U.K.’s Brexit referendum campaign in 2016, the Vote Leave movement operated a targeted release of fake news adverts on Facebook to win votes, and ultimately the referendum (BBC, 2018; Merrick, 2019; Hern, 2019). Throughout the 2017 French presidential election, voters were exposed to fake news stories that were explicitly anti-Macron and that condemned the liberal agenda in general (BBC, 2017; Hosenball and Menn, 2017; Scott, 2017). Additionally, Ferrara (2017) finds evidence of computer-controlled Twitter accounts, “bots”, that propagated unverified ‘leaked’ documents in a coordinated effort to discredit Macron’s campaign. The same bots supported Donald Trump during the U.S. presidential election a year earlier. Evidence suggests the possible existence of a black market for reusable fake news bots to influence political results (Ferrara, 2017).

Fake news stories are not limited to political events. Muller and Schwarz (2018) use Facebook data from the German right-wing AfD political party to show that anti-immigration sentiment combined with regional social media activity predicts violent crimes against refugees. Their results suggest that social media can act as a propagation mechanism between online hate speech and violent crime. The lines between online hate speech and violence were blurred in March 2019,
when a terrorist attack on a mosque in Christchurch, New Zealand, was broadcast live on Facebook by the gunman (BBC, 2019). In Sri Lanka, the Department of Government Information temporarily blocked social media after a series of suicides bombings in 2019 to prevent “social unrest via hate messages and false information” (Iyengar, 2019). Fake news stories are clearly pervasive and harmful to society. Yet people continue to use social media as a source of information.

Most social media news consumers in the U.S. expect reporting to be largely inaccurate, but convenience and interactions with people outweigh concerns (Matsa and Shearer, 2018). People also rely on others to judge what is true. Nearly half of Americans think that most people are capable of separating fact from fiction on social media (48%), whilst confidence is slightly lower in Europe (~45%) (Ipsos, 2019). The average social media user prefers to not spread false information. Friggeri, Adamic, Eckles and Chen (2014) show that re-shares of rumours on Facebook are significantly more likely to be deleted by the re-poster once a link to the fact-checked story is placed in the comments. However, large cascades can accumulate hundreds of these comments and continue to propagate. The key question is under which conditions rumours continue to spread.

The goal of this paper is to shed light on the role of social network structure in this debate. Our starting point is a model by Bloch, Demange and Kranton (2018), henceforth BDK, which considers rumour transmission in a network environment. The network environment has biased agents who potentially create rumours, and unbiased agents who
seek the truth. BDK show that unbiased agents believe that rumours are true when biased agents are few and far in between. If biased agents congregate, however, the network can act as a filter. Unbiased agents located near such a group prevent transmission of false messages into the wider community. We implement the model in a laboratory experiment with two treatments: a network treatment and a broadcast treatment. In the network treatment messages travel along the structure of the network if players decide to pass on the messages they receive. In the broadcast treatment messages are received by all players simultaneously. The key difference between the treatments is message transmission between players, which takes place in the network treatment only. We then consider two questions: (1) Can unbiased agents take into account the credibility of messages coming from different parts of the network? And, (2) Does message transmission along the structure of the network affect collective outcomes?

Our experimental data reveal some interesting insights into these questions. Firstly, we find some evidence to suggest that unbiased agents can discriminate between more and less truthful sources of information. Secondly, and contrary to theoretical predictions, either environment leads to the same set of outcomes. But the behaviours that underpin the outcomes differ. The main factor in the broadcast treatment is the message creation strategy of the biased agent, and in the network treatment it is the transmission strategy of the unbiased agent.
The remainder of this paper is organised as follows. We describe our experimental design in section 2.2 and we outline our research hypotheses in section 2.3. We discuss our results in section 2.4 and we present our conclusions in section 2.5.

2.2 Experimental design

We formally define the model in section 2.2.1 and we outline the experiment implementation in section 2.2.2.

2.2.1 The model

We use the general model of BDK to specify one applied model that forms the basis for our two experimental treatments. Our applied model is described as follows. A group of four agents, denoted by $i \in N \equiv \{1, 2, 3, 4\}$, play a game where individual payoffs depend on a group decision, $x \in \{0, 1\}$, and on a state of the world, $\theta \in \{0, 1\}$. There are two types of agent in the group, each with different preferences: one biased agent, $i = 2$, and three unbiased agents, $i \in \{1, 3, 4\}$. Unbiased agents prefer the group decision to match the true state of world, i.e. $x = \theta$, and have utility:

$$w(x, \theta) = 10 \left(- (x - \theta)^2 + 1\right)$$  \hspace{1cm} (1)

The biased agent prefers group decision $x = 1$, regardless of the true state of the world, and has utility:

$$v(x, \theta) = 10 \left(- (x - 1)^2 + 1\right)$$  \hspace{1cm} (2)

The preferences of each type are common knowledge.
Agents have a common prior belief that $\theta = 1$ with probability $\pi = 0.3$. Hence, agents initially believe that $\theta = 0$ with higher probability and look for credible information that $\theta = 1$. Agents are arranged in an exogenously generated undirected network with four nodes, where each node represents an agent and each link represents the possibility to send and receive messages. A link between agents $i$ and $j$ is denoted by $ij$ and the network is represented by the set $G$ of all existing links. We denote by $N_i$ the set of agents that have a link with agent $i$. That is, $N_i$ represents an agent $i$’s set of neighbours.

Agents interact in three stages to reach a group decision: (i) a message creation stage, (ii) a message transmission stage, and (iii) a voting stage.

(i) **Message creation.** Nature generates a perfect signal of the true state of the world, $s = \theta$, with probability $p = 0.9$. If generated, the signal is received by one agent in the group (the *signal recipient*), and each agent is selected to be the signal recipient with equal probability. An agent $i$ who receives the signal can either create a message $m_i \in \{0, 1\}$, independent of the information that is contained in the signal, or create no message, $\emptyset$. That is, the message creation strategy of agent $i$ is a mapping from the set of states of the world $\{0, 1\}$ to the set $\{0, 1, \emptyset\}$ of possible messages. If an agent $i$ creates a message, it is received by all neighbours, $j \in N_i$.

(ii) **Message transmission.** An agent $i$ who receives a message $m(j)$ from a neighbour $j \in N_i$ chooses to transmit the message to her
neighbours who have not yet received it, \( t_i(m(j)) = m(j) \), or to block the message, \( t_i(m(j)) = \emptyset \). That is, the transmission strategy of agent \( i \) is a mapping from the set of all possible messages \( m(j), \{0, 1\} \times N_j \), into the set \( \{m(j), \emptyset\} \), of all possible transmission decisions. Note that agents cannot transform any message \( m(j) \), and that at most one message can circulate in the population. The message transmission stage is complete when all transmission decisions have been taken, when no message was created, or when no signal was generated.

(iii) Voting. Once the message transmission stage is complete, each agent \( i \) has a posterior belief \( \rho_i \in [0, 1] \) for the event that \( \theta = 1 \) and casts a vote either for outcome 0 or for outcome 1, i.e. \( v_i \in \{0, 1\} \). One vote is selected, each with equal probability, and implemented as the group decision \( x \in \{0, 1\} \). That is, we implement a probabilistic voting model where the probability that an alternative is chosen increases with the number of votes it receives. Let \( z \) be the number of other agents who vote for outcome 1. Then the probability that group decision 1 is implemented is \( \frac{z}{4} \) if agent \( i \) votes for outcome 0, or \( \frac{z+1}{4} \) if agent \( i \) votes for outcome 1. It follows directly that the expected utility of agent \( i \) is:

\[
E_i[w(x, \theta)] = 10 \frac{z}{4} \rho_i + 10 \frac{4-z}{4} (1 - \rho_i) \quad \text{if } v_i = 0 \tag{3}
\]

\[
E_i[w(x, \theta)] = 10 \frac{z+1}{4} \rho_i + 10 \frac{3-z}{4} (1 - \rho_i) \quad \text{if } v_i = 1 \tag{4}
\]

We implement two network configurations: a line network and a complete network. In the line network, agents \( i \) and \( i+1, i \in \{1, 2, 3\} \) are connected, and in the complete network each agent is connected with all
other agents. In both networks the biased agent is $i = 2$. In what follows we assign the label B2 to the biased agent and the labels U1, U3 and U4 to the unbiased agents. Figure 1 depicts the two networks.

![Figure 2.1a: The line network](image)

![Figure 2.1b: The complete network](image)

The key difference between the networks is the message transmission stage. When the signal recipient creates a message, it is received by all her neighbours. Hence, in the complete network all agents receive the message simultaneously and therefore the message transmission stage is not played. Furthermore, when an agent receives a message $m(j)$ from a neighbour $j \in N_i$, she is informed of the identity of the sender. In the complete network, the sender must also be the creator of the message and players know whether a message was created by the biased agent or by an unbiased agent. In the line network this is not necessarily the case. For example, when agent U1 receives a message from agent B2, she is unsure whether the message was created by agent B2, U3 or U4. Therefore, a
comparison between the two networks allows us to study the effect of the transmission stage, as a proxy for social network structure, on collective outcomes. We now refer to the line network environment as the *network* treatment and to the complete network environment as the *broadcast* treatment.

There exist many equilibria, including equilibria where agents do not send messages and babbling equilibria where messages do not contain useful information and wherefore agents do not update their priors. We are concerned with equilibria where agents send messages and where welfare is maximised. Our analysis follows BDK, who characterise a *full communication equilibrium* (FCE) and a *maximal communication equilibrium* (MCE). In an FCE unbiased agents create $m_i = s$ and transmit any message $m(j)$, that is $t_i(m(j)) = m(j)$. The biased agent creates message $m_{B2} = 1$, independent of $\theta$, transmits message $m(Uj) = 1$, $j \in \{1, 3\}$ and blocks message $m(Uj) = 0$, $j \in \{1, 3\}$. Finally, the biased agent votes for outcome 1, i.e. $v_{B2} = 1$, and unbiased agents vote according the message if received, $v_i = m(j)$, or vote for outcome 0, $v_i = 0$, if no message was received. We now can state our first theoretical prediction.

**Proposition 1:** An FCE does not exist in either the network game or the broadcast game.
Proof.

We prove the proposition by contradiction. Assume that an FCE exists. First consider the network game when agent U3 receives message $m(B2) = 1$ and computes her posterior belief $\rho_{U3}$ for the event that $\theta = 1$. U3 believes that state 1 is the most probable state if $\rho_{U3} > \frac{1}{2}$.

Applying Bayes’ rule, we have:

$$\rho_{U3} = Pr (\theta = 1 | m(B2) = 1)$$

$$= \frac{Pr(m(B2) = 1 | \theta = 1) Pr(\theta = 1)}{Pr(m(B2) = 1 | \theta = 1) Pr(\theta = 1) + Pr(m(B2) = 1 | \theta = 0) Pr(\theta = 0)}$$

Note that $m(B2) = 1$ is created by B2 or by U1 with equal probability. If $\theta = 1$, both agents create truthful messages and $Pr(m(B2) = 1 | \theta = 1) = 1$. If $\theta = 0$, only U1 creates a truthful message and $Pr(m(B2) = 1 | \theta = 0) = 0.5$. We substitute these values into (5) to obtain $\rho_{U3} = \frac{0.3}{0.3 + 0.5 \cdot 0.7} = 0.46$. U3 does not believe that state 1 is the most probable state because $\rho_{U3} < \frac{1}{2}$. It follows directly that transmission of the message decreases her expected utility since U4 will vote according to the message received. Therefore, transmission of $m(B2) = 1$ is not optimal and it is in contradiction with the existence of FCE

Now consider the broadcast game when unbiased agents receive $m(B2) = 1$. We compute the posterior belief of one them, say $\rho_{U3}$, without loss of generality. Applying equation (5) we obtain $\rho_{U3} = \frac{0.3}{0.3 + 0.7} = 0.3 = \pi$. U3 does not believe that state 1 is the most probable state because $\rho_{U3} < \frac{1}{2}$. 

Therefore, it is not optimal to vote for outcome 1 and this is in contradiction with the existence of an FCE.

QED

Given that an FCE does not exist, we focus our attention on the MCE.

**Proposition 2:** The following strategy profile and beliefs form a weak perfect Bayesian equilibrium and, hence, constitute an MCE.

1. **Message creation strategy:** Biased agent B2 creates message \( m_{B2} = 1 \) regardless of the signal that is received. Unbiased agents create message \( m_i = s \).

2. **Message transmission strategy (network game only):** Biased agent B2 transmits message \( m(U1) = 1 \) and \( m(U3) = 1 \) and blocks message \( m(U1) = 0 \) and \( m(U3) = 0 \). Unbiased agent \( U3 \) transmit message \( m(B2) = 0, m(U4) = 0 \) and \( m(U4) = 1 \), and blocks message \( m(B2) = 1 \).

3. **Voting strategy:** Biased agent B2 votes \( v_{B2} = 1 \). Unbiased agents vote \( v_i = 0 \) when no message is received and vote \( v_i = m(j) \) when the message is received from another unbiased agent. When \( m(B2) = 1 \) is received in the network game, U1 votes \( v_{U1} = 1 \) and U3
votes \( v_{U3} = 0 \). When \( m(B2) = 1 \) is received in the broadcast game, unbiased agents vote \( v_i = 0 \).

4. **Beliefs.** Beliefs apply only to unbiased agents and are therefore described only for this type. For information set that are reached with positive probability, beliefs follow Bayes’ rule. We can state the following: After \( m(j) = \emptyset \), \( \rho_i \leq 0.3 \) because with some probability message \( m_i = 0 \) was blocked by B2. After \( m(j) = 0, j \neq B2, \rho_i = 0 \) and after \( m(B2) = 0, \rho_i = 0.3 \) (the last case is off the equilibrium path). After \( m(j) = 1, j \neq B2, \rho_i > \frac{1}{2} \). After \( m(B2) = 1 \), in the broadcast game \( \rho_i = 0.3 \) and in the network game \( \rho_{U1} > \frac{1}{2} \) and \( \rho_{U3} < \frac{1}{2} \).

The strategy profile and beliefs are summarised by treatment in table 2.1 and computations are included in Appendix B.9.
<table>
<thead>
<tr>
<th>Agent</th>
<th>Message creation</th>
<th>Message received</th>
<th>Message transmission</th>
<th>Voting</th>
<th>Posterior belief</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( i )</td>
<td>( m_i )</td>
<td>( m(j) )</td>
<td>( t_i(m(j)) )</td>
<td>( v_i )</td>
</tr>
<tr>
<td>Network</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>( m_{U1} = s )</td>
<td>-</td>
<td>-</td>
<td>( s )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(B2) = 0 )</td>
<td>-</td>
<td>0</td>
<td>( 0.3 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(B2) = 1 )</td>
<td>-</td>
<td>1</td>
<td>( 0.5625 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>0</td>
<td>( 0.0723 )</td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>( m_{B2} = 1 )</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U1) = 0 )</td>
<td>( \emptyset )</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U1) = 1 )</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U3) = 0 )</td>
<td>( \emptyset )</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U3) = 1 )</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>( m_{U3} = s )</td>
<td>-</td>
<td>-</td>
<td>( s )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(B2) = 0 )</td>
<td>0</td>
<td>0</td>
<td>( 0.3 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(B2) = 1 )</td>
<td>( \emptyset )</td>
<td>0</td>
<td>( 0.4615 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U4) = 0 )</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U4) = 1 )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>0</td>
<td>( 0.1165 )</td>
<td></td>
</tr>
<tr>
<td>U4</td>
<td>( m_{U4} = s )</td>
<td>-</td>
<td>-</td>
<td>( s )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U3) = 0 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(U3) = 1 )</td>
<td>1</td>
<td>1</td>
<td>( 0.5625 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>0</td>
<td>( 0.4204 )</td>
<td></td>
</tr>
<tr>
<td>Broadcast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1, ( m_i = s )</td>
<td>-</td>
<td>-</td>
<td>( s )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3, ( m(B2) = 0 )</td>
<td>-</td>
<td>0</td>
<td>( 0.3 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4 ( m(B2) = 1 )</td>
<td>-</td>
<td>0</td>
<td>( 0.3 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(j) = 0, j \neq B2 )</td>
<td>-</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(j) = 1, j \neq B2 )</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>0</td>
<td>( 0.3 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2 ( m_{B2} = s )</td>
<td>-</td>
<td>-</td>
<td>( 1 )</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>( m(j) = 0 )</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(j) = 1 )</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( m(j) = \emptyset )</td>
<td>-</td>
<td>1</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

58
Proof.

Voting stage. The voting strategy described is optimal in either game. Given beliefs and expected payoffs in equations (3) and (4), it is straightforward to see that unbiased agent $i$ votes for outcome 1 if $\rho_i > \frac{1}{2}$ and for outcome 0 if $\rho_i < \frac{1}{2}$. If $\rho_i = \frac{1}{2}$, unbiased agent $i$ is indifferent between voting for outcome 1 or 0.

Transmission stage (network game only). First consider biased agent B2. A deviation where B2 blocks $m(U1) = 1$ does not change her expected payoff. Either action to transmit or to block the message induces posterior belief $\rho_i < \frac{1}{2}$ in U3 and in U4 and both agents vote for outcome 0 either case. A deviation where B2 blocks $m(U3) = 1$ decreases her expected payoff because the action reduces the posterior belief of U1 to $\rho_{U1} < \frac{1}{2}$. Consequently, U1 votes for outcome 0 instead of outcome 1 and reduces the probability that preferred outcome 1 is implemented. A deviation where B2 transmits either $m(U1) = 0$ or $m(U3) = 0$ does not change her expected payoff. Either action to transmit or to block induces posterior belief $\rho_i < \frac{1}{2}$ in U1 and in U3 and all unbiased agents vote for outcome 0 either case.

Now consider unbiased agent U3. A deviation where U3 transmits $m(B2) = 1$ decreases her expected payoff. After $m(B2) = 1$, U3 has posterior belief $\rho_{U3} < \frac{1}{2}$ and her expected payoff for outcome 0 is higher than for outcome 1. Transmitting the message induces in U4 posterior belief $\rho_{U4} > \frac{1}{2}$. U4 votes for outcome 1 and reduces the probability that preferred
outcome 0 is implemented. A deviation where U3 blocks \( m(B2) = 0 \) does not change her expected payoff. Either action to transmit or to block the message induces posterior belief \( \rho_i < \frac{1}{2} \) in U4 and U4 votes for outcome 0 either case. A deviation where U3 blocks \( m(U4) = 0 \) does not change her expected payoff. Either action by B2 to transmit or to block the message induces in U1 posterior belief \( \rho_{U1} < \frac{1}{2} \) and U1 votes outcome 0, while B2 votes for outcome 1 in any case. A deviation where U3 blocks \( m(U4) = 1 \) decreases her expected payoff. After \( m(U4) = 1 \), U3 has posterior belief \( \rho_{U3} = 1 \) and her expected payoff for outcome 1 is higher than for outcome 0. If U3 blocks the message, she induces in U1 posterior belief \( \rho_{U1} < \frac{1}{2} \). U1 votes for outcome 0 and reduces the probability that preferred outcome 1 is implemented.

*Message creation stage.* Recall that nature generates a perfect signal of the true state of the world. An unbiased who receives signal \( s \) is therefore perfectly informed of the state of the world \( \theta \). She votes \( v_i = s \) because outcome \( s \) has a higher expected payoff than the alternative outcome.

We start with the network game. Consider unbiased agent U1 and suppose \( m_{U1} = 0 \). B2 blocks the message and induces \( \rho_{U3} < \frac{1}{2} \) and \( \rho_{U4} < \frac{1}{2} \). Suppose \( m_{U1} = 1 \). B2 transmits the message but U3 blocks the message and \( \rho_{U3} < \frac{1}{2} \) and \( \rho_{U4} < \frac{1}{2} \). Suppose \( m_{U1} = \emptyset \). No-one receives a message and, again, \( \rho_{U3} < \frac{1}{2} \) and \( \rho_{U4} < \frac{1}{2} \). Given optimal strategies in the voting stage,
agents U3 and U4 vote for outcome 0 regardless of U1’s message creation strategy. It follows directly that the probability that U1’s preferred outcome is implemented, as well as her expected payoff, is the same in each case. Therefore, there are no profitable deviations.

Consider unbiased agent U3 and suppose \( m_{U3} = 0 \). U4 receives the message but B2 blocks transmission to U1, so the action induces \( \rho_{U1} < \frac{1}{2} \) and \( \rho_{U4} < \frac{1}{2} \) and all unbiased agents vote for outcome 0. Suppose \( m_{U3} = 1 \). All agents receive the message, which induces posterior belief \( \rho_i > \frac{1}{2} \) in all unbiased agents. All agents in the network vote for outcome 1. Suppose \( m_{U3} = \emptyset \). No-one receives a message and \( \rho_{U1} < \frac{1}{2} \) and \( \rho_{U4} < \frac{1}{2} \). Given optimal strategies in the voting stage, agents U1 and U4 vote for outcome 0 when \( m_{U3} = 0 \) and when \( m_{U3} = \emptyset \), while B2 votes for outcome 1. It directly follows that creating a false message or no message reduces the probability that U3’s preferred outcome is implemented, as well as her expected payoff. Therefore, there are no profitable deviations.

Consider unbiased agent U4 and suppose \( m_{U4} = 0 \). U3 receives the message but B2 blocks transmission to U1, so the action induces \( \rho_{U1} < \frac{1}{2} \) and \( \rho_{U3} < \frac{1}{2} \) and all unbiased agents vote for outcome 0. Suppose \( m_{U4} = 1 \). All agents receive the message, which induces posterior belief \( \rho_i > \frac{1}{2} \) in all unbiased agents. All agents in the network vote for outcome 1. Suppose \( m_{U4} = \emptyset \). No-one receives a message and \( \rho_{U1} < \frac{1}{2} \) and \( \rho_{U3} < \frac{1}{2} \). Given optimal strategies in the voting stage, agents U1 and U3 vote for outcome 0 when
$m_{U4} = 0$ and when $m_{U4} = \emptyset$, while B2 votes for outcome 1. It directly follows that creating a false message or no message reduces the probability that U4’s preferred outcome is implemented, as well as her expected payoff. Therefore, there are no profitable deviations.

Consider biased agent B2 and suppose $m_{B2} = 1$. The message is received by U1 and U3, but U3 blocks transmission to U4. The action induces $\rho_{U1} > \frac{1}{2}, \rho_{U3} < \frac{1}{2}$ and $\rho_{U4} < \frac{1}{2}$. Consequently, U1 votes for outcome 1, while U3 and U4 vote for outcome 0. Suppose $m_{B2} = 0$. All unbiased agents receive the message, which induces posterior belief $\rho_i < \frac{1}{2}$ and all vote for outcome 0. Suppose $m_{B2} = \emptyset$. No-one receives a message. All unbiased agents have posterior belief $\rho_i < \frac{1}{2}$ and all vote for outcome 0. Any deviation from $m_{B2} = 1$ decreases the probability that B2’s preferred outcome 1 is implemented, as well as her expected payoff. Therefore, there are no profitable deviations.

Now we look at the broadcast game. Consider biased agent B2. Any message created by B2 induces posterior belief $\rho_i = 0.3$ in all unbiased agents, and all unbiased agents vote for outcome 0. Therefore, there are no profitable deviations.

Consider any unbiased agent, for example U1, and suppose $m_{U1} = 1$. The message is received by all agents and induces posterior belief $\rho_i > \frac{1}{2}$ in all unbiased agents. All agents vote for outcome 1. Suppose $m_{U1} = 0$. The message is received by all agents. It induces posterior belief $\rho_i < \frac{1}{2}$ in all
unbiased agents, who vote for outcome 0. Suppose \( m_{U1} = \emptyset \). No-one receives a message. All unbiased agents have posterior belief \( \rho_i < \frac{1}{2} \) and all vote for outcome 0. Given optimal strategies in the voting stage, agents U3 and U4 vote for outcome 0 when \( m_{U1} = 0 \) and when \( m_{U1} = \emptyset \), while B2 votes for outcome 1. It directly follows that creating a false message or no message reduces the probability that U1’s preferred outcome is implemented, as well as her expected payoff. Therefore, there are no profitable deviations.

QED

The proofs demonstrate that the MCE under consideration is not strict in many information sets. This leaves open the possibility of different but almost equivalent equilibria. Suppose, for example, the network game where the MCE strategy profile is modified so that \( m_{U1} \) is created randomly instead of fully informatively. It is straightforward to see that, although the posterior beliefs are different from the MCE, this profile is an equilibrium. Furthermore, any modifications where unbiased agents do not create a message after receiving \( s = 0 \) are equilibria of the game. Similarly, in the broadcast game, a modified strategy profile where \( m_{B2} \) is created randomly and uninformatively remains an equilibrium. The multiplicity of equilibria makes coordination on the MCE a difficult and challenging task to explore in the laboratory.
2.2.2 Implementation

The experiments took place in an economics laboratory in the form of a computerised abstract decision-making task using neutral language that avoids direct reference to terminology that may trigger personal motivations (e.g., biased, rumour, etc.). On entering the laboratory, subjects were randomly allocated to a seat at a computer and given a copy of the written instructions provided in Appendix B.1. To make the instructions easier to follow, the less likely state 1 was labelled ‘red’ and the more likely state 0 was labelled ‘blue’. We adopt this convention in the remainder of the paper.

After reading the instructions, subjects answered a set of questions to check their understanding (included in Appendix B.2) and became familiar with the game during three unpaid trial periods, which were followed by 25 paid periods. At the start of the first paid period, participants were randomly and anonymously matched into groups of four players, and each player was subsequently assigned to one of the four positions in their group. Groups and assigned positions remained fixed throughout the game.

At the end of each period, players receive private information about earnings and a full account of events, including the true state, the position that received the signal if there was a signal, the message that was created, transmission decisions, votes cast by each position, the position whose vote was implemented, and the group decision.
At the end of 25 paid periods, participants answered three sets of questions. The first set of questions collected information regarding participants’ beliefs and expectations about the behaviour of the other players. The second set was a social preferences elicitation questionnaire to measure social preferences in terms of pro-sociality and envy following Bartling, Fehr, Marechal and Schunk (2009). The third set was a lie aversion elicitation questionnaire to measure an individual’s attitude about deceptive communication (Oliveira and Levine, 2008). Finally, to assess risk attitudes, participants played an investment game in which they were given the option to invest all, part, or none of their show-up fee in a risky investment (Charness and Gneezy, 2010). Details of the questionnaires and test are included in Appendices B.3 to B.6.

We conducted three session for each treatment with, on average, four groups per session. Subjects were restricted to participation in one session only and 96 unique subjects were recruited from the student population of Royal Holloway, University of London, during December 2018 and January 2019. The sessions were held at the University’s economics laboratory ExpReSS Lab using z-Tree software for economic experiments (Fishbacher, 2007). Earnings averaged around £10 per subject for a 1-hour session.
2.3 Research hypotheses

In this section we outline our research hypotheses. Our hypotheses are based on predictions from the proposed MCE, which are summarised in Table 2.2, and on behavioural considerations. As mentioned, we label the less likely state 1, which is preferred by biased agents, ‘red’ and we label more likely state 0 ‘blue’.

Our general research question is whether subjects conform to the MCE, or whether they coordinate on a less informative or uninformative equilibrium. A growing experimental literature has documented preferences for truth-telling over payoff-maximisation and suggests a behavioural tendency to select the most informative equilibria (see Abeler, Nosenzo and Raymond (2019) for a comprehensive survey). Furthermore, in both the broadcast game and the network game there is an efficient outcome where the biased agent sends truthful messages and votes for outcome red. But this strategy cannot be an equilibrium because the biased agent deviates to always creating a red message. To see this, first consider the broadcast game. Note that in the MCE a red message from the biased agent induces posterior belief $\rho_i < \frac{1}{2}$ in unbiased agents because she cannot hide her identity. Therefore, each unbiased agent votes for outcome blue and the biased agent votes for outcome red, so the red group decision is selected with probability 0.25. But the red state occurs with probability $\pi = 0.30$. Thus, we note that a strategy where the biased agent creates truthful
messages and unbiased agents vote according the received messages is preferred by agents of both types. Now consider the network game and the MCE. All agents create a red message after receiving a red signal. When either U1 or B2 creates a red message, U1 and B2 vote red and U3 and U4 vote blue. So, outcome red is selected with probability 0.5. When either U3 or U4 creates a red message all agents vote accordingly, and outcome red is selected with probability 1. All unbiased agents create a blue message after receiving a blue signal. The biased agent blocks a blue message in either direction but, in any case, all unbiased agents vote for outcome blue and the biased agent votes for outcome red. So, outcome red is selected with probability 0.25. The biased agent creates a red message after receiving a blue signal. As before, the red outcome is selected with probability 0.5. Therefore, the expected value for the biased agent is $0.3 \cdot 0.75 + 0.7 \cdot 0.3125 = 0.444$. When, alternatively, the biased agent creates truthful messages, unbiased agents vote according to the received messages. The biased agent always votes for outcome red, so the expected value for the biased agent is $0.3 \cdot 1 + 0.7 \cdot 0.25 = 0.475$. Thus, the biased agent has a higher expected value for creating truthful messages. Equally, unbiased agents prefer this outcome because U1 votes red only when it is the true state. Thus, in both games there exists an outcome that is not an equilibrium and that Pareto dominates the MCE. We will refer to this outcome as the efficient state. Several papers in the experimental literature show that agents can coordinate on efficient non-equilibrium
outcomes (e.g. Crawford, Costa-Gomez and Iriberri (2013) includes a survey).

Finally, considering the predictions in the MCE and behavioural preferences for truth-telling and efficiency, our general prediction is summarised as follows:

Subject’s behaviour conforms to the MCE, but (i) biased agents create more truthful messages than predicted, especially in the broadcast treatment, and (ii) biased agents in the network treatment block fewer messages than predicted.
Table 2.2: Equilibrium predictions

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Signal</th>
<th>Recipient</th>
<th>Creation Message</th>
<th>Transmission B2 U3</th>
<th>Voting U1 B2 U3 U4</th>
<th>Outcome probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network</td>
<td>Blue</td>
<td>U1</td>
<td>Blue Block</td>
<td>Blue Red Blue</td>
<td>0.75 0.25</td>
<td>0.69 0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B2</td>
<td>Red Block</td>
<td>Red Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>U3</td>
<td>Blue Block</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>U4</td>
<td>Blue Block Pass</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>U1</td>
<td>Red</td>
<td>Pass Block</td>
<td>Red Red Blue Blue</td>
<td>0.50 0.50</td>
<td>0.25 0.75</td>
</tr>
<tr>
<td></td>
<td>B2</td>
<td>Red</td>
<td>Pass Block</td>
<td>Red Red Blue Blue</td>
<td>0.50 0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U3</td>
<td>Red</td>
<td>Pass</td>
<td>Red Red Red Red</td>
<td>0.00 1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U4</td>
<td>Red</td>
<td>Pass Pass</td>
<td>Red Red Red Red</td>
<td>0.00 1.00</td>
<td></td>
</tr>
<tr>
<td>Broadcast</td>
<td>Blue</td>
<td>Unbiased</td>
<td>Blue</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biased</td>
<td>Red</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td>Red</td>
<td>Unbiased</td>
<td>Red</td>
<td>- -</td>
<td>Red Red Red Red</td>
<td>0.00 1.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biased</td>
<td>Red</td>
<td>- -</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall</td>
<td>0.19 0.81</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>-</td>
<td>-</td>
<td>- -</td>
<td>Blue Red Blue Blue</td>
<td>0.75 0.25</td>
<td></td>
</tr>
</tbody>
</table>
In what follows we formulate four specific hypotheses, one regarding each of the three stages of the game and one regarding the outcomes.

**Message creation.** The biased agent benefits only when a red group decision is implemented and, therefore, has an incentive to create a red message irrespective of the signal. However, the experimental literature on lie aversion (Abeler et al., 2019) suggests possible deviations from this prediction and subsequently more informative behaviour of unbiased agents in either treatment. The higher expected payoffs in the efficient state also provide an incentive for deviation. Note that in the broadcast game deviations from the prediction are costless because a red message is not credible to unbiased agents. Therefore, it could be beneficial to create truthful messages and attempt to coordinate on the efficient state. Unbiased agents prefer the group decision to match the true state of the world and so have an incentive to create messages that match the signal. But note that when unbiased agents receive a blue signal, they are indifferent between creating a truthful message or no message because the other agents vote in the same way in either event. Preferences for truth-telling, however, suggest that this deviation will be sparse. We condense these considerations in our first hypothesis:

**H1: Message creation.** *Biased agents create red messages irrespective of the signal and unbiased agents create truthful messages. Behaviourally we expect that biased agents create more*
truthful messages than predicted, and more so in the broadcast treatment relative to the network treatment.

**Message transmission.** The message transmission stage applies only to biased agent B2 and unbiased agent U3 in the network game. First consider B2. We predict that the biased agent blocks blue messages and transmits red ones. Note that there is no strict incentive for B2 to block a blue message because B2 induces $\rho_1 < \frac{1}{2}$ whether she blocks or transmits a blue message, and unbiased agents vote for blue in either event. Now consider U3. We predict that U3 transmits blue messages and transmits red messages from U4 but blocks red messages from B2. U3 need not doubt the veracity of a red message received from U4, because U4 creates truthful messages. A red message from biased agent B2, however, could be either true or false and induces a posterior belief $\rho_{U3} < \frac{1}{2}$. Thus, U3 does not believe that the red state is the most probable state and she blocks transmission. We predict that U3 transmits blue messages but note that, again, there is no strict incentive to do so. U3 induces $\rho_1 < \frac{1}{2}$ whether she blocks or transmits a blue message and other agents vote the same in either event. Preferences for truth-telling, however, suggest that this deviation will be limited. Our second hypothesis is:

**H2: Message transmission (Network).** *Biased agent B2 transmits red messages and blocks blue messages.* *Unbiased agent*
U3 transmits blue messages, transmits red messages received from U4 and blocks red messages received from B2. Behaviourally we expect that B2 blocks blue messages less often than predicted.

**Voting.** We predict that unbiased agents vote for outcome blue when they receive neither the signal nor a message. When they receive the signal, they vote according to the content. Unbiased agents in the broadcast game vote according to any message received from another unbiased agent, and vote blue after any message from the biased agent. In the network game, unbiased agents U1 and U4 vote according to any message received. U3 votes according to the message received from U4 and votes blue after any message from B2. Finally, biased agent B2 always votes red in either treatment. We expect that in either game unbiased agents deviate from the predictions if the biased agent deviates from the predicted message creation strategy. Unbiased agents vote according to the message received from B2 if she uses a truthful message creation strategy, and vote for outcome blue if she uses any other alternative message creation strategy. Our third hypothesis captures this:

**H3: Voting.** Unbiased agents vote blue in the absence of a message and vote according to the signal when it is received. Unbiased agents in either treatment vote according to any message received from another unbiased agent. In the broadcast treatment, unbiased agents vote blue after any message received from the biased
agent. In the network treatment, $U1$ votes according to any message received from the biased agent and $U3$ votes blue after any message received from the biased agent. Biased agents always vote red. Behaviourally we expect that unbiased agents vote according to messages received from the biased agent more often if the biased agent uses a more truthful message creation strategy.

**Outcomes and Welfare.** Welfare is maximised when the sum of individual payoffs is maximised. This is the case when unbiased agents vote for the true state of the world and the group decision $x$ matches the true state $\theta$. Therefore, we refer to the event $x = \theta$ as the ‘correct group decision’.

We predict that the correct group decision is selected with higher probability in the broadcast treatment relative to the network treatment. In the MCE the correct red (blue) group decision is selected with aggregate probability $0.81$ ($0.75$) in the broadcast treatment and with aggregate probability $0.75$ ($0.69$) in the network treatment. Note also that the correct red group decision is more likely to be selected than the correct blue group decision in either treatment. The red state occurs less often but it benefits from the biased agent always voting red. On the other hand, unbiased agents are more likely to vote for the correct group decision in the blue state. Note that the ordering of the predicted probabilities does not change if the biased agent uses a more truthful message creation strategy. Our fourth and final hypothesis is:
**H4: Outcomes and welfare.** The correct group decision is selected more often in the broadcast treatment relative to the network treatment, and the correct red group decision is selected more often than the correct blue group decision in either treatment.

### 2.4 Experimental results

Our hypotheses consider predictions of aggregate level behaviours within the network treatment and within the broadcast treatment, as well as between the network treatment and the broadcast treatment. The subject pools for the two treatments showed no significant differences in social preferences, lie acceptability and risk attitudes (details are included in Appendix B.7). We test our hypotheses using primarily non-parametric tests where an independent observation is the average decision of a subject located in a fixed position within a four-player network over 25 periods of the game. Most of our sample sizes are small due to the many possible actions in the game and a limited budget. For paired data, we dismiss the Wilcoxon signed-rank test because differences between samples are not symmetrically distributed, and we resort to the lower power sign test to analyse the equality of medians in two-tailed tests. For unpaired data we use the Wilcoxon rank-sum test. We regard results at the 1% significance level and at the 5% significance level as deviating from predictions.
We present four results that correspond to our four hypotheses in section 2.4.1 and, where aggregate behaviour departs from predictions, we analyse individual behaviour in section 2.4.2.

2.4.1 Aggregate behaviour

We follow the outline of the previous section and address each hypothesis in turn.

**Message creation.** We first consider the biased agents. Figure 2.2 displays by treatment the message creation decisions of biased subjects after each signal. B2 subjects created a red message after a red signal 85 percent of the time in the network treatment and 82 percent of the time in the broadcast treatment. The results are more divergent after a blue signal. B2 subjects created a red (blue) message 64 (27) percent of the time in the network treatment and 49 (31) percent of the time in the broadcast treatment. Indeed, biased subjects in both games created more truthful messages than predicted in the MCE, but the difference between the treatments appears small. Biased subjects in the broadcast treatment created empty messages after either a red or a blue signal, but more so after the latter.
We first test for a treatment effect on message creation choices using a two-tailed t-test and the results are displayed in Table 2.3. The test reveals a positive and significant network treatment effect (5% level) for the creation of a red message after a blue signal. We also find a positive and significant broadcast treatment effect for the creation of an empty message after a red signal (1% level) and after a blue signal (5% level). Recall that in the MCE for the broadcast game, the biased agent has no strict incentive to create a message. Note that we do not observe the expected positive broadcast treatment effect for the creation of truthful blue messages.

Table 2.3: B2 message creation - two-tailed t-test

<table>
<thead>
<tr>
<th>State</th>
<th>B2 created</th>
<th>Network prob</th>
<th>Network s.e.</th>
<th>Broadcast prob</th>
<th>Broadcast s.e.</th>
<th>Network vs Broadcast t-test</th>
<th>*p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Red</td>
<td>0.846</td>
<td>0.030</td>
<td>0.824</td>
<td>0.038</td>
<td>0.642</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>0.154</td>
<td>0.083</td>
<td>0.059</td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>0.000</td>
<td>0.000</td>
<td>0.118</td>
<td>0.042</td>
<td>0.005</td>
<td>***</td>
</tr>
<tr>
<td>Blue</td>
<td>Red</td>
<td>0.636</td>
<td>0.040</td>
<td>0.490</td>
<td>0.054</td>
<td>0.029</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Blue</td>
<td>0.273</td>
<td>0.053</td>
<td>0.306</td>
<td>0.054</td>
<td>0.857</td>
<td></td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>0.091</td>
<td>0.037</td>
<td>0.204</td>
<td>0.037</td>
<td>0.018</td>
<td>**</td>
</tr>
</tbody>
</table>

***, **, * denote significance at, respectively, the 1%, 5% and 10% levels.
We now test the hypothesis that biased agents create truthful messages at the same rate irrespective of the signal. For the network treatment we reject the null-hypothesis ($p = 0.0215$) and conclude that truthful message creation depends on the signal, while in the broadcast treatment evidence against the null-hypothesis is only weak ($p = 0.0703$). That is, biased agents in the network treatment conform to the message creation strategy in the MCE but biased agents in the broadcast treatment less so. We elaborate on the last point in section 2.4.2 and now direct our attention to the unbiased agents. Figure 2.3 displays shows for each treatment the message creation decisions after each signal. Unbiased agents in the broadcast treatment are identical and their data is therefore pooled, but unbiased agents in the network treatment need to be considered separately.
Adherence to predictions was excellent in the broadcast treatment, where subjects created a red message after a red signal 90 percent of the time, and a blue message after a blue signal 86 percent of the time. The sign test provides no evidence that truthful message creation depends on the signal ($p = 0.7266$). In the network treatment adherence is more variable, ranging from 72 percent to 88 percent, but again we find no evidence that truthful message creation depends on the signal for U1 ($p = 0.6250$), for U3 ($p = 1.000$) or for U4 ($p = 0.2500$). We summarise our first result as follows:

**Result 1: Message creation.** *Biased agents in the network treatment create red messages irrespective of the signal, but biased agents in the broadcast treatment take account of the signal. Unbiased agents create truthful messages. Additionally,*
biased agents create more truthful messages than predicted in either treatment, albeit at a similar rate.

**Message transmission.** We first consider biased agent B2 in the network treatment. B2 subjects transmitted red (blue) messages 88 (51) percent of the time. Indeed, biased agents mostly transmit red messages and block far fewer blue messages than predicted. A sign test that compares the median average blocking rates for red messages and blue messages narrowly fails to provide evidence for a significant difference ($p = 0.0654$). We now focus on unbiased agent U3 in the network treatment. U3 subjects transmitted red (blue) messages 79 (48) percent of the time when received from B2 and 87 (77) percent of the time when received from U4. U3 indeed mostly transmits red messages received from U4 but, contrary to prediction, she also transmits red messages from B2 at nearly the same rate. The sign test provides no evidence for a difference in the median transmission rates of red messages received from B2 and from U4. We address this point separately in section 2.4.2. Our second result is:

**Result 2: Message transmission (network).** *Biased agent B2 transmits red messages and blue message at a similar rate. Unbiased agent U3 transmits blue messages at a lower rate than predicted and transmits red messages irrespective of the sender.*

**Voting.** We first review the voting choices of unbiased agents in the broadcast treatment. Subjects voted for outcome blue 67 percent of the
time when no message was received. This is lower than predicted in the MCE, but close to the prior for the blue state (0.70). Subjects voted according to the signal 95 percent of the time and according to a blue (red) message from another unbiased agent 95 (88) percent of the time. After a message from the biased agents, subjects voted for outcome blue only 45 percent of the time. More precisely, they voted according to a red (blue) message from the biased agent 66 (71) percent of the time. Thus, contrary to predictions, unbiased agents voted according to a message from the biased agent most of the time. We now turn to unbiased agents in the network treatment. U1, U3 and U4 voted for outcome blue after no message was received 81, 88 and 72 percent of the time respectively. Subjects voted according to the signal 91 percent of the time overall, with U1 at 97 percent, U3 at 91 percent and U4 at 86 percent. U3 voted according to a blue (red) message from U4 97 (80) percent of the time, and U4 voted according to a blue (red) message from U3 86 (93) percent of the time. After a red (blue) message was received from the biased agent, U1 voted accordingly 76 (86) percent of the time and U3 voted accordingly 79 (76) percent of the time. Thus, unbiased agents mostly voted in line with predictions except for U3, who voted according to a red message from the biased agent most of the time. Finally, biased subjects voted for outcome red 91 percent of the time in the broadcast treatment and 89 percent of the time in the network treatment. So, biased agents chiefly voted in line with predictions. We arrive at our third result:
Result 3: Voting. Unbiased agents vote blue in the absence of a message, albeit less often than predicted in the broadcast treatment, and vote according to the signal when it is received. Unbiased agents in either treatment vote according to any message received from another unbiased agent. Contrary to predictions, unbiased agents in the broadcast treatment regularly vote according to the message from the biased agent. In the network treatment, U1 votes in line with predictions, but U3 votes according to messages from the biased agent. Biased agents vote for outcome red.

Outcomes and Welfare. Table 2.4 displays the equilibrium predictions for selection of the correct group decision $x = \theta$ alongside observed relative frequencies for each treatment. We test for the equality of means between predicted and observed frequencies using a one-sample t-test. Furthermore, we use the Wilcoxon rank-sum test to compare the distributions of observed rates between the network treatment and the broadcast treatment.

We start with the prediction that the correct group decision is selected more often in the broadcast treatment than in the network treatment. We first review the statistics for the blue state. Table 2.4 shows that in the broadcast treatment, the correct blue group decision was selected with an overall relative frequency of 0.67 versus a predicted probability of 0.75. The
lower observed rate is not significant in a two-tailed t-test ($p = 0.0513$). But note that when the biased agent receives a blue signal, the correct group decision is selected with relative frequency 0.50. This is lower than the predicted probability of 0.75 and the difference is significant at the 5% level ($p = 0.0402$). In the network treatment the blue group decision was selected with an average relative frequency of 0.65 compared with a predicted probability of 0.69. Again, the observed rate is lower, but the difference is not significant ($p = 0.4111$). The correct blue group decision is indeed selected more often in the broadcast treatment relative to the network treatment, but the difference is small and not significant in a Wilcoxon rank-sum test ($p = 0.7910$).
Table 2.4: Predicted probabilities and observed average relative frequencies for the correct group decision

<table>
<thead>
<tr>
<th>State</th>
<th>Signal Recipient</th>
<th>Network</th>
<th>Broadcast</th>
<th>Network vs Broadcast</th>
<th>Wilcoxon ranksum p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Observed</td>
<td>Predicted</td>
<td>Observed</td>
<td>Predicted</td>
</tr>
<tr>
<td>Blue</td>
<td>B2</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50**</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>U1</td>
<td>0.70</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U3</td>
<td>0.74</td>
<td>0.75</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>U4</td>
<td>0.70</td>
<td>0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>0.65</td>
<td>0.69</td>
<td>0.67</td>
<td>0.75</td>
</tr>
<tr>
<td>Red</td>
<td>B2</td>
<td>0.81***</td>
<td>0.50</td>
<td>0.74***</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>U1</td>
<td>0.78**</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>U3</td>
<td>0.97</td>
<td>1.00</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>U4</td>
<td>0.68**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>0.80</td>
<td>0.75</td>
<td>0.87</td>
<td>0.81</td>
</tr>
<tr>
<td>Blue</td>
<td>None</td>
<td>0.72</td>
<td>0.75</td>
<td>0.58</td>
<td>0.75</td>
</tr>
<tr>
<td>Red</td>
<td>None</td>
<td>0.30</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*** and ** denote significance at the 1% and 5% levels, respectively, in a two-tailed one-sample t-test
*** and ** denote significance at the 1% and 5% levels, respectively, in a two-tailed Wilcoxon ranksum test
We now focus our attention on the red state. The correct red group decision in the broadcast treatment was selected with an overall average relative frequency of 0.87 versus a predicted probability of 0.81. The higher aggregate observed rate level is not significantly different from the predicted probability \( (p = 0.2358) \). Note, however, that when the biased agent received a red signal, the correct group decision was selected with relative frequency 0.74. This is significantly higher (1% level) than the predicted probability of 0.25 \( (p = 0.0095) \). In the network treatment the correct red group decision was selected with an overall average relative frequency of 0.80 compared with a predicted probability of 0.75. The difference at the aggregate level is small and not significant \( (p = 0.2923) \).

But again, when the biased agent received a red signal, the correct group decision was selected far more often than predicted with an observed frequency of 0.81 versus a predicted probability of 0.25. The difference is significant at the 1% level \( (p = 0.0076) \). We observe a similar difference when U1 receives a red signal. The correct group decision is selected at an observed rate of 0.78 compared with an 0.50 probability. The observed rates when U1 and B2 receive the signal are higher than predicted because U3 deviates from the predicted strategy and transmits red messages from the biased agent. We conclude that, overall, the correct red group decision is indeed selected more often in the broadcast treatment relative to the network treatment. But the difference between treatments is not significant in a Wilcoxon rank-sum test \( (p = 0.1470) \).
We note that in both treatments the correct blue group decision is selected less often than predicted, and the correct red group decision is selected more often than predicted. The predicted difference is 6 percentage points in either treatment, but we observe a difference of 20 percentage points in the broadcast treatment and of 15 percentage points in the network treatment. Indeed, the correct red group decision is selected more often than the correct blue group decision in either treatment, but the gap is wider than predicted in both treatments.

**Result 4:** Outcomes and welfare. The correct group decision is selected more often in the broadcast treatment relative to the network treatment. The correct red group decision is selected more often than the correct blue group decision in either treatment, but the gap is wider than predicted.

### 2.4.2 Individual behaviour

Aggregate level behaviour deviates from the strategies in the MCE in two crucial cases: (i) biased agent B2 in the broadcast treatment deviates from the message creation strategy by creating truthful messages and (ii) unbiased agent U3 in the network treatment deviates from the message transmission strategy by passing on red messages from the biased agent. We analyse individual behaviour for these cases, starting with biased agents in the broadcast treatment.
To understand individual strategies, we analyse for each group the message creation choices of the biased agent and the subsequent effects on the voting choices of unbiased agents. We measure the relationship between signals and created messages with the pairwise Pearson correlation coefficient $r$. The results of the calculation are displayed in the first column of Table 2.5. A biased agent who conforms to the strategy in the MCE creates a red message regardless of the signal and therefore has correlation coefficient $r = 0$. A biased agent who always creates truthful messages in either state has correlation coefficient $r = 1$, whilst a biased agent who always creates false messages in either state has correlation coefficient $r = -1$. The second and third columns of Table 2.5 show the rates at which unbiased agents voted red and blue after receiving, respectively, a red message and a blue message from the biased agent.

A correlation coefficient in the range $-0.5$ to $0.5$ suggests that the biased agent created messages that were uninformative, and so we expect that unbiased agents vote for outcome blue after receiving either a red message or a blue message. We observe this strategy in groups 2 to 6. Unbiased agents voted according to a red message at an average rate of 0.57 across the five groups, with the lowest rate at 0.17 and the highest rate at 0.83. A correlation coefficient greater than 0.5 indicates that the biased agent created messages that were more truthful and thus informative. This value suggests an attempt to coordinate on the efficient state.
Table 2.5: Pearson correlation coefficient between signal and message creation and the rates of red votes and blue votes from recipient unbiased agents. By agent type in the broadcast treatment.

<table>
<thead>
<tr>
<th>Group</th>
<th>Biased agents create message</th>
<th>Unbiased agents create message</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Signal·Msg</td>
<td>(2) $U_i$ vote red after red msg</td>
</tr>
<tr>
<td></td>
<td>$r$</td>
<td>rate</td>
</tr>
<tr>
<td>1</td>
<td>-0.5774</td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>0.0000</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.0000</td>
<td>0.17</td>
</tr>
<tr>
<td>4</td>
<td>0.2000</td>
<td>0.67</td>
</tr>
<tr>
<td>5</td>
<td>0.2582</td>
<td>0.67</td>
</tr>
<tr>
<td>6</td>
<td>0.3162</td>
<td>0.83</td>
</tr>
<tr>
<td>7</td>
<td>0.6325</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>1.0000***</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>1.0000***</td>
<td>0.78</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***, **, * denote significance at, respectively, the 1%, 5% and 10% levels
Three groups (7, 8 and 9) enjoyed this strategy and unbiased agents voted according to a red message at an average rate of 0.93, with perfect compliance in two groups. The biased agent in group 1 has a correlation coefficient smaller than -0.5 which indicates inverted signalling. Perhaps the biased agent intended to fool the recipients, but unbiased agents still voted according to a red message at a rate of 0.78. After receiving a blue message, unbiased agents voted blue at an overall average rate of 0.69 with little variation between groups. We note that blue messages were informative regardless of the message creation strategy of the biased agent. Finally, column 4 in the panel on the right of Table 2.5 shows that unbiased agents created informative messages with average correlation coefficient of 0.8375. Recipient unbiased agents voted according to a red (blue) message at an average rate of 0.87 (0.94). Generally unbiased agents created truthful messages and other unbiased agents voted accordingly. The data align to our earlier analysis.

Lastly, we consider the effects of individual beliefs and preferences on the message creation choices of agents by type. We use data from the lie acceptability test, the social preferences questionnaire and the beliefs questionnaire as explanatory variables in a multinomial logistic regression that estimates their effects on the probability that the predicted message is created. The social preferences explanatory variables were dropped in the final regression model because they were not significant and prevented convergence. We subsequently calculate the marginal effects of each explanatory variable and the results are displayed in Table 2.6. The
A biased agent who believes that unbiased agents find her blue message credible is 100% more likely to create a red message than a biased agent who does not hold the same belief and the effect is significant at the 1% level. We do not find evidence for significant lie acceptability effects, but a biased agent is 42% more likely to create a red message after a red signal than after a blue signal. The effect is also significant at the 1% level. Unbiased agents who believe that other unbiased agents find their red message credible are 20% more likely to create truthful message, which is significant at the 5% level.

Table 2.6: Marginal effect of beliefs about other agents, the signal and lie acceptability on the creation of the theoretically predicted message. By agent type in the broadcast treatment.

<table>
<thead>
<tr>
<th></th>
<th>Biased</th>
<th>Unbiased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{b} )</td>
<td>( \hat{b} )</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>Unbiased believes Blue</td>
<td>1.000***</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.396)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>Unbiased believes Red</td>
<td>-0.110</td>
<td>0.200**</td>
</tr>
<tr>
<td></td>
<td>(0.345)</td>
<td>(0.079)</td>
</tr>
<tr>
<td>Biased believes Blue</td>
<td></td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.110)</td>
</tr>
<tr>
<td>Biased believes Red</td>
<td></td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.115)</td>
</tr>
<tr>
<td>Signal Red</td>
<td>0.421***</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.151)</td>
<td>-</td>
</tr>
<tr>
<td>Lie Acceptability</td>
<td>-0.023</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>-</td>
</tr>
</tbody>
</table>

***, **, * denote significance at the 1%, 5% and 10% levels respectively
We summarise as follows. At the individual level we find evidence for both behaviour conforming to the MCE and behaviour conforming to the efficient state. Subjects in groups 2 and 3 largely adhere to the predicted strategies and the outcome is close to the MCE. Biased subjects in groups 7, 8 and 9 created more truthful messages than predicted and the result is coordination on the efficient state. We do not find evidence for a lie acceptability effects.

We now analyse for each group in the network treatment the message transmission choices of unbiased agent U3 after receiving a message from the biased agent. We measure the relationship between the messages that U3 receives from B2 and the true state of world with the pairwise Pearson correlation coefficient $r$ again. The results of calculations are displayed in the third column of Table 2.7, while the fourth and the fifth columns show the rates at which U3, respectively, transmits and votes according to a red message. The interpretation of the coefficients is as before. A correlation coefficient in the range of $-0.5$ to $0.5$ suggests that U3 received uninformative messages, and therefore we expect that she blocks transmission to U4. We observe this in the first five groups. The average transmission rate is 0.56 and actions range from never transmitting to always transmitting the message. Transmission choices are consistent with voting choices in all groups but one (3). A correlation coefficient greater than $0.5$ indicates that U3 received more informative message and we expect that U3 transmits the messages. We observe more informative
messages in groups 6 to 12 and U3 indeed transmitted all red message from B2 with one exception (group 10). U3 voted according to a red message at an average rate of 0.89, but we observe some heterogeneity among groups.

Table 2.7: Pearson correlation between signal and message creation for U1 and for B2. Pearson correlation between messages received by U3 and the state. Transmission rates of red message by U3 and red vote rates of U3 conditional on receiving a red message.

<table>
<thead>
<tr>
<th>Group</th>
<th>U1 created $r$ Signal $\cdot$ Msg</th>
<th>B2 created $r$ Signal $\cdot$ Msg</th>
<th>U3 received $r$ Msg $\cdot$ State</th>
<th>Pass red rate U3</th>
<th>Vote red rate U3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-0.6124</td>
<td>-0.2500</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>0.2500</td>
<td>0.1890</td>
<td>0.0259</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>3</td>
<td>1.0000***</td>
<td>-0.4082</td>
<td>0.2182</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>0.1667</td>
<td>0.2182</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>5</td>
<td>0.0000</td>
<td>-</td>
<td>0.3333</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>6</td>
<td>0.7453**</td>
<td>0.0000</td>
<td>0.5270*</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>1.0000***</td>
<td>0.0000</td>
<td>0.6325**</td>
<td>1.00</td>
<td>0.90</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>0.5000</td>
<td>0.6547*</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>1.0000***</td>
<td>0.3780</td>
<td>0.6614***</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>10</td>
<td>1.0000***</td>
<td>0.4472</td>
<td>0.6708**</td>
<td>0.75</td>
<td>1.00</td>
</tr>
<tr>
<td>11</td>
<td>1.0000***</td>
<td>0.5000</td>
<td>0.7500*</td>
<td>1.00</td>
<td>0.50</td>
</tr>
<tr>
<td>12</td>
<td>1.0000***</td>
<td>1.0000***</td>
<td>1.0000***</td>
<td>1.00</td>
<td>0.80</td>
</tr>
</tbody>
</table>

***, **, * denote significance at, respectively, the 1%, 5% and 10% levels.

The first and second columns of Table 2.7 show the correlation coefficients between signals and created messages for U1 and B2 respectively. U1 always created truthful messages except in groups 2 and 5. Biased subjects B2 created informative messages only in groups 1 and 12, with signalling inverted in the former. Further analysis reveals that
B2 subjects always transmitted red messages from U1 and regularly transmitted blue messages from U1. We did not observe a relationship between this last point and the behaviour of U3 subjects.

In summary, U3 received informative message in seven of 12 groups and in these groups U3 justifiably transmitted red messages from the biased agent and voted accordingly. Compliance to a red message is lower in the five groups where U3 received uninformative messages. Thus, we find some evidence to suggest that U3 can discriminate between a more truthful and a less truthful biased agent. We do not have enough data to identify a relationship between the transmission of blue messages by B2 and the behaviour of U3.

2.5 Conclusion

We have designed and reported on an experiment that examines the effects of rumours in an environment with biased agents and unbiased agents. Our experiments implement a general model proposed by BDK and we are the first to experimentally explore the role of network structure on the efficacy of rumours. Firstly, we find that biased agents in the broadcast treatment deviate from the predicted message creation strategy by creating more truthful messages. As a result, we observe coordination on the efficient state in some groups, but we cannot identify whether the biased agent is motivated by a preference for truth-telling or by a taste for efficiency. This question may be answered by comparing our results to an
additional broadcast treatment where the identity of sender, and thus the creator, of the message is unknown. Secondly, unbiased agent U3 in the network treatment transmits and votes according to a red message irrespective of the sender. At an individual level we find justification for these choices because in most groups U3 receives informative messages. We also find some suggestive evidence that U3 can discriminate between more and less truthful biased agents. Thirdly, the correct group decision in either state is selected more often in the broadcast treatment relative to the network treatment, but the difference is not significant. Either environment leads to the same set of outcomes, but the behaviours that underpin the outcomes differ. The main factor in the broadcast treatment is the message creation strategy of the biased agent, and in the network treatment it is the transmission strategy of U3. Question marks remain around this last point. Biased agents make a costless deviation from the MCE by regularly transmitting blue messages. The choice could be motivated by fairness considerations or could be a strategic one to establish a reputation of credibility. Either way, the action influences the perception that U3 has of the biased agent. The question whether the biased agent deviates intentionally and, if so, driven by which motivation remains open.

We have shed barely a speck of light on the role of network structure on the efficacy of rumours, and our research has raised more questions than it has answered. We hope that our results inspire others to further explore the interaction between network structure and rumour propagation.
CHAPTER 3:

Network Intervention: Temporary Incentives and Persistence

3.1 Introduction

This paper reports the design and result of a partial pilot study. Our research contributes to a growing literature that considers a central planner who designs optimal network intervention strategies (e.g. Valente, 2012; Demange, 2017; Galeotti, Golub and Goyal, 2019). The central planner aims to induce behavioural change using incentives that consider the network structure and an individual’s position within the network. We also consider the nature of game (strategic complements or strategic substitutes), and the statistic that captures neighbour’s actions which determine payoffs (e.g. the mean, the minimum or the maximum). Furthermore, the planner wishes to limit intervention and use a temporary incentive only to shift the system to an efficient outcome (e.g. Brandts and Cooper, 2006; Brandts and Cooper, 2006b).

Due to time and budget constraints, we present results only for the case of strategic complements where the relevant statistic is the minimum. We illustrate this context with an example inspired by Hirshleifer (1983). Imagine a perfectly circular island where each citizen owns a wedge-shaped slice of land that runs from the centre to the sea. The island needs dikes if it is to be prevented from occasional flooding. Each citizen decides for themselves whether to build a dike or not. If everyone builds a dike, the
island never floods. But if one citizen does not build a dike, the island floods from time to time and the efforts of those who built dikes is wasted. We conduct our pilot study in this environment. Our goal is to explore the effects of a temporary financial incentive to shift the system from an inefficient to an efficient equilibrium, and to understand the conditions under which the effect persists after the incentive is removed.

We conduct an experiment where subjects engage in an abstract decision-making task in an online game. The game is a repeated game of strategic complements with incomplete information. Our network game follows the design introduced by Galeotti, Goyal, Jackson, Vega Redondo and Yariv (2010) in terms of information environment and uses an extension by Feri and Pin (2019) for the aggregation of externalities.

In each of 30 rounds, subjects are in a new directed network structure with the same 10 players. The network has five positions that have one connection and five positions that have three connections. Knowing only their number of connections, players choose to be active at a cost (e.g. build a dike) or to be inactive (e.g. not build a dike). Payoffs are determined by the minimum of neighbour’s actions, so the choice to be active is riskier for players with three connections. We then study the effect of an incentive. The incentive is no cost for the choice to be active for players with degree 3 during the second set of 10 rounds. The incentive expires in the third set of 10 rounds when, once again, all players pay the cost for the choice to be active. We have two treatments that differ in the probability distribution of the network formation process for degree 3. We then focus on three
questions: (1) Is a financial incentive targeted at players with degree 3 enough to move the system to the efficient outcome? (2) If the shift occurs, does the effect persist after the incentive is removed? And, (3) does the network formation process affect outcomes.

The data reveal some interesting initial insights. Firstly, the targeted incentive successfully moved the system to the efficient equilibrium in both treatments. Secondly, the shift was not persistent in either treatment - at least at the aggregate level. Thirdly, we observe heterogeneity in outcomes between group in only one treatment.

The rest of this paper is organised as follows. We describe the game, the equilibrium predictions the implementation in section 3.2. We present statistics and econometric results in section 3.4 and we briefly conclude in section 3.5.

3.2 Experimental design

We define the game in section 3.2.1 and we outline our predictions in section 3.2.2. We describe the experiment implementation in section 3.2.3.

3.2.1 The network game

In the experiment in this paper we focus on a game of strategic complements played between 10 players arranged in a directed random network $g$. Each player is represented by a node. A link from player $i$ to player $j$, denoted by $ij \in g$, indicates that the action of player $j$ affects the
payoff of player \( i \) (and not vice versa). Let \( N_i \) be the set of players (neighbours) that receive a link from player \( i \), i.e. \( N_i \in \{\forall j : ij \in g \} \). Lastly, \( d_i \) denotes the degree of player \( i \), i.e. \( d_i = |N_i| \in K \equiv \{1, 3\} \), i.e. players have either degree 1 or 3.

Each player \( i \) chooses an action \( a_i \in \{0, 1\} \) where 0 denotes the choice to be \textit{inactive} and 1 denotes the choice to be \textit{active}, \( a_i = 1 \) has a cost of \( c \) and \( a_i = 0 \) is costless. The payoff that player \( i \) receives depends on her own action and the actions chosen by her neighbours. Let \( e_i \) be the endowment of player \( i \), then her payoff is:

\[
\pi_i = a_i(b \cdot s_i - c) + e_i
\]

where \( b \) is a bonus and \( s_i \) is the minimum action played by \( i \)'s neighbours, i.e. \( s_i = \min\{a_j\}_{j \in N_i} \). We assume that \( b > c \) and \( e_i \geq c \).

When choosing an action, players do not know the network structure. They know only their degree in the realised network and the probability distribution over their neighbour’s degrees. Thus, players play a game of incomplete information. A strategy for player \( i \) is a mapping \( \sigma_i : K \rightarrow \Delta\{0, 1\} \) where \( \Delta\{0, 1\} \) is the set of probability distributions on \( \{0, 1\} \). That is, \( \sigma_i = \{\sigma_{ik}\}_{k \in \{1, 3\}} \) where \( \sigma_{ik} \) is the mixed strategy played by player \( i \) of degree \( k \) and denotes the probability to play action 1.

In what follows we consider symmetric Bayesian Nash equilibria (BNE), where every agent who has the same information and faces the same ex ante conditions (i.e. each agent \( i \) with the same degree \( k \)) chooses the same strategy, i.e. \( \sigma_{ik} = \sigma_{jk} \forall ij \) and \( \forall k \). Furthermore, in an \textit{all equal} strategy profile all players take the same action, i.e. \( \sigma_{ik} = \sigma_{jk}, \forall ij \) and \( \forall k, k' \).
Our network design allows us to apply the analysis and results of Feri and Pin (2019). For any strategy profile, the function $s_i$ is weakly degree-decreasing. That is, the expected value of $s_i$ is not increasing with the degree of player $i$ and weakly first order stochastic dominance (FOSD) degree-decreasing. Degree-decreasing is defined as follows. Let $E_k(s_i|\sigma)$ be the expected value of $s$ for node $i$ of degree $k$ when the strategy is $\sigma$. Given our network formation process, a statistic $s$ is degree-decreasing if for every $\sigma$ we have that $E_3(s_i|\sigma) < E_1(s_i|\sigma)$. A statistic $s$ is weakly degree-decreasing when the conditions are satisfied only for all strategy profiles that are not all equal. Now let $\Phi_k(s_i|\sigma)$ be the cumulative probability distribution on $s$. Given our network formation process, a statistic $s$ is FOSD degree-decreasing if for every $\sigma$, and $x \in \mathbb{R}$, we have that $\Phi_3(x|\sigma) \geq \Phi_1(x|\sigma)$ with strict inequality for some $x$. A statistic $s$ is weakly degree-decreasing when the strict inequalities hold only for every strategy profile that is not all equal $\sigma$.

A straight application of the results in Proposition 2 of Feri and Pin (2019) shows that the symmetric equilibria in this game are either all equal or FOSD decreasing, i.e. $\sigma_1$ FOSD $\sigma_3$. It follows that, in the context of pure strategies, FOSD decreasing means that $a_1 = 1$ and $a_3 = 0$.

It is easy to proof that there exists an equilibrium where all players choose action 0, an equilibrium where all players choose action 1, and at least one equilibrium where players with degree 1 are active and those with degree 3 are inactive.
In the experiment we implement a network with 10 subjects where the degree of each subject is either 1 or 3. The network is reconfigured every round and in any realised network five subjects have degree 1 and five subjects have degree 3. We design two treatments that differ in the connection probabilities for degree 3 as shown in Table 3.1. The connection probabilities are common knowledge and in either treatment $b = 10$, $c = 3.5$ and $e_i = 3.5 \forall i$.

<table>
<thead>
<tr>
<th>Player degree</th>
<th>Number of neighbours with degree = 1</th>
<th>Number of neighbours with degree = 3</th>
<th>Probabilities Treatment 1</th>
<th>Probabilities Treatment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>degree = 1</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>degree = 3</td>
<td>3</td>
<td>0</td>
<td>0.125</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0.375</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>0.375</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>3</td>
<td>0.125</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### 3.2.2 Predictions

We focus on symmetric Bayesian Nash equilibria (BNE), where all players of the same degree choose the same actions. Therefore, all players of the same degree have the same strategy. Each treatment has the same set of equilibria, which includes three pure strategy equilibria and two mixed strategy equilibria. The equilibria are outlined in Proposition 1 and summarised in Table 3.2.
Proposition 1: Both treatments have the following five symmetric BNEs: (i) all players choose action 0 (ii) players with degree 1 choose action 1 and players with degree 3 choose action 0, (iii) all players choose action 1, (iv) players with degree 1 play a mixed strategy and players with degree 3 choose action 0, and (v) players with degree 1 choose action 1 and players with degree 3 play a mixed strategy.

The proof below is based on the following considerations. We test the equilibrium conditions for the five strategy profiles described in Proposition 1. In Claim 1 we prove that the expected value of action 1 for a player with degree 1 is strictly higher than the expected value of action 1 for a player with degree 3, for all strategy profiles except the ones which are all equal. By this result we can prove that all other strategy profiles are not BNE.

Proof: We denote with subscript $k$ all the quantities that are common for all agents with degree $k$. Then a symmetric strategy profile is given by the vector $\sigma = (\sigma_1, \sigma_3)$, where $\sigma_1$ and $\sigma_3$ are the mixed strategy profiles played by, respectively, players of degree 1 and degree 3. Let $E_k(\pi|\sigma_i, \sigma_{-i})$ be the expected payoff for a player of degree $k$ from playing strategy $\sigma_i$ when the other players play strategy profile $\sigma_{-i}$. Then the expected payoffs for the pure strategies are:
\[ E_1(\pi|1, \sigma_{-i}) = 10(0.5\sigma_1 + 0.5\sigma_3) = 5(\sigma_1 + \sigma_3) \]

\[ E_3(\pi|1, \sigma_{-i}) = 10(0.125\sigma_1^3 + 0.375\sigma_1^2\sigma_3 + 0.375\sigma_1\sigma_3^2 + 0.125\sigma_3^3) = \\
1.25(\sigma_1^3 + 3\sigma_1^2\sigma_3 + 3\sigma_1\sigma_3^2 + \sigma_3^3) \text{ in treatment 1} \]

\[ E_3(\pi|1, \sigma_{-i}) = 10(0.05\sigma_1^3 + 0.45\sigma_1^2\sigma_3 + 0.45\sigma_1\sigma_3^2 + 0.05\sigma_3^3) = 0.5(\sigma_1^3 + \\
9\sigma_1^2\sigma_3 + 9\sigma_1\sigma_3^2 + \sigma_3^3) \text{ in treatment 2} \]

\[ E_1(\pi|0, \sigma_{-i}) = E_3(\pi|0, \sigma_{-i}) = 3.5 \]

Claim 1: \( E_1(\pi|1, \sigma_{-i}) > E_3(\pi|1, \sigma_{-i}) \) \( \forall \sigma_{-i} = \{\sigma_1, \sigma_3\} \) such that \( \sigma_1 + \sigma_3 \in (0,2) \). Consider treatment 1. The claim requires \( 5(\sigma_1 + \sigma_3) > 1.25(\sigma_1^3 + \\
3\sigma_1^2\sigma_3 + 3\sigma_1\sigma_3^2 + \sigma_3^3) \). This inequality can be written as \( 4(\sigma_1 + \sigma_3) > \\
(\sigma_1 + \sigma_3)^3 \) that is satisfied when \( \sigma_1 + \sigma_3 \in (0,2) \). Consider treatment 2. The claim requires \( 5(\sigma_1 + \sigma_3) > 0.05(\sigma_1^3 + 9\sigma_1^2\sigma_3 + 9\sigma_1\sigma_3^2 + \sigma_3^3) \). Multiplying and dividing the right hand side by 3 we get \( 0.15 \left( \frac{\sigma_1^3}{3} + 3\sigma_1^2\sigma_3 + 3\sigma_1\sigma_3^2 + \\
\frac{\sigma_3^3}{3} \right) \). Note that this is smaller than the righthand side of the inequality for treatment 1. Finally, it is directly verifiable that inequality is satisfied when \( \sigma_1 + \sigma_3 \in (0,2) \). QED

Now we prove that \( \sigma^* = (0,0) \) is a BNE. Note that under this strategy profile \( s_i = 0 \) for all players. The expected equilibrium payoffs are \( E_1(\pi|0, \sigma_i^*) = E_3(\pi|0, \sigma_i^*) = 3.5 \). For a player either of degree 1 or 3 to deviate to a strategy \( \sigma_i > 0 \), i.e. a strategy where action 1 is played with strictly positive probability is not profitable because the expected payoff from action 1 is 0, i.e. \( E_1(\pi|1, \sigma_i^*) = E_3(\pi|1, \sigma_i^*) = 0 \). Then the expected payoffs from \( \sigma_i \) are \( E_1(\pi|\sigma_i, \sigma_i^*) = E_3(\pi|\sigma_i, \sigma_i^*) = 3.5(1 - \sigma_i) < 3.5, \forall \sigma_i > 0 \).
Now we prove that $\sigma^* = (1,1)$ is a BNE. Note that under this strategy profile $s_i = 1$ for all players. The expected equilibrium payoffs are $E_1(\pi|1,\sigma^*_i) = E_3(\pi|1,\sigma^*_i) = 10$. For a player either of degree 1 or 3 to deviate to a strategy $\sigma_i < 1$, i.e. a strategy where action 0 is played with strictly positive probability is not profitable because the expected payoff from action 0 is 3.5, i.e. $E_1(\pi|0,\sigma^*_i) = E_3(\pi|0,\sigma^*_i) = 3.5$. Then the expected payoffs from $\sigma_i$ are $E_1(\pi|\sigma_i,\sigma^*_i) = E_3(\pi|\sigma_i,\sigma^*_i) = 10\sigma_i + 3.5(1 - \sigma_i) < 10, \forall \sigma_i < 0$.

Now we prove that $\sigma^* = (1,0)$ is a BNE. Note that under this strategy profile

$$s_i = \begin{cases} 1 \text{ with probability } 0.5 & \text{ if } d_i = 1 \text{ in both treatments} \\ 0 \text{ with probability } 0.5 & \end{cases}$$

$$s_i = \begin{cases} 1 \text{ with probability } 0.125 & \text{ if } d_i = 3 \text{ and treatment 1} \\ 0 \text{ with probability } 0.875 & \end{cases}$$

$$s_i = \begin{cases} 1 \text{ with probability } 0.05 & \text{ if } d_i = 3 \text{ and treatment 2.} \\ 0 \text{ with probability } 0.95 & \end{cases}$$

The expected equilibrium payoffs are $E_1(\pi|1,\sigma^*_i) = 5$ and $E_3(\pi|0,\sigma^*_i) = 3.5$ in both treatments. For a player of degree 1 to deviate to a strategy $\sigma_i < 1$, i.e. a strategy where action 0 is played with strictly positive probability is not profitable because the expected payoff from action 0 is 3.5, i.e. $E_1(\pi|0,\sigma^*_i) = 3.5$. Indeed the expected payoffs from $\sigma_i$ are $E_1(\pi|\sigma_i,\sigma^*_i) = 5\sigma_i + 3.5(1 - \sigma_i) < 5, \forall \sigma_i < 1$. For a player of degree 3 to deviate to a strategy $\sigma_i > 0$, i.e. a strategy where action 1 is played with strictly positive probability is not profitable because the expected payoff from action 1 is $E_3(\pi|1,\sigma^*_i) = 1.25$ in treatment 1 and $E_3(\pi|1,\sigma^*_i) = 0.5$ in treatment 2. Then the expected payoffs from $\sigma_i$ are $E_3(\pi|\sigma_i,\sigma^*_i) = 1.25\sigma_i + 0.5$.
\[3.5(1 - \sigma_i) < 3.5, \forall \sigma_i > 0 \text{ in treatment 1 and } E_3(\pi|\sigma_i, \sigma^*_i) = 0.5\sigma_i + 3.5(1 - \sigma_i) < 3.5, \forall \sigma_i > 0 \text{ in treatment 2.}\]

Now we prove that \(\sigma^* = (0,1)\) is not a BNE. Equilibrium conditions require \(E_1(\pi|1, \sigma^*_i) \leq 3.5\) and \(E_3(\pi|1, \sigma^*_i) \geq 3.5\), which is a contradiction with Claim 1.

Now we prove that \(\sigma^* = (\sigma_1, 1)\) is not a BNE \(\forall \sigma_1 \in (0,1)\). The equilibrium conditions are \(E_1(\pi|1, \sigma^*_1) = E_1(\pi|0, \sigma^*_1) = 3.5\) and \(E_3(\pi|1, \sigma^*_3) \geq 3.5\), which is a contradiction with Claim 1.

Now we prove that in the class of strategy profiles \(\sigma = (\sigma_1, 0)\) where \(\sigma_1 \in (0,1)\) only \(\sigma^* = (0.35,0)\) is a BNE in both treatments. The equilibrium condition on player with degree 1 is that \(E_1(\pi|1, \sigma^*_i) = E_1(\pi|0, \sigma^*_i)\), i.e.

\[10\sigma_1 = 3.5.\]

This is satisfied only for \(\sigma_1 = 0.35\). For player with degree 3 the equilibrium payoff is 3.5. For player \(i\) of degree 3 the deviation to a strategy \(\sigma_i > 0\) is not profitable because it produces an expected payoff of

\[E_3(\pi|\sigma_i, \sigma^*_i) = 0.125 \cdot 10 \cdot 0.35^3 < 3.5 \text{ in treatment 1 and } E_3(\pi|\sigma_i, \sigma^*_i) = 0.05 \cdot 10 \cdot 0.35^3 < 3.5 \text{ in treatment 2.}\]

Now we prove that \(\sigma^* = (0, \sigma_3)\) is not a BNE \(\forall \sigma_3 \in (0,1)\). The equilibrium conditions are \(E_3(\pi|1, \sigma^*_i) = E_3(\pi|0, \sigma^*_i) = 3.5\) and \(E_1(\pi|1, \sigma^*_3) \leq 3.5\), which is a contradiction with Claim 1.

Now we prove that in each treatment there exist a \(\overline{\sigma}_3 \in (0,1)\) such that \(\sigma^* = (1, \overline{\sigma}_3)\) is a BNE. The equilibrium condition on player with degree 3 is that \(E_3(\pi|1, \sigma^*_i) = E_3(\pi|0, \sigma^*_i)\), i.e. in treatment 1 we have \(10 \cdot (0.125 \overline{\sigma}_3^3 + 0.375\overline{\sigma}_3^2 + 0.375\overline{\sigma}_3 + 0.125) = 3.5\) that is satisfied only for \(\overline{\sigma}_3 = 0.41\).

Similarly, we find that for treatment 2 the analogous condition is satisfied.
only for $\bar{\sigma}_3 = 0.45$. We note that the expected equilibrium payoff for a player with degree 1 is $E_1(\pi|1, \sigma^*_i) > 5$. By the same arguments as above a deviation to a strategy $\sigma_i < 1$ is never profitable.

Now we prove that $\sigma^* = (\sigma_1, \sigma_3)$ is not a BNE in both treatments. The equilibrium conditions are $E_3(\pi|1, \sigma^*_i) = E_3(\pi|0, \sigma^*_i) = 3.5$ and $E_1(\pi|1, \sigma^*_i) = E_1(\pi|0, \sigma^*_i) = 3.5$, which is a contradiction with Claim 1. QED

### Table 3.2: Equilibrium predictions

<table>
<thead>
<tr>
<th>Equilibrium</th>
<th>Probability to be active</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment 1</td>
</tr>
<tr>
<td></td>
<td>degree = 1</td>
</tr>
<tr>
<td>Secure</td>
<td>0</td>
</tr>
<tr>
<td>Intermediate 1</td>
<td>0.35</td>
</tr>
<tr>
<td>Intermediate 2</td>
<td>1</td>
</tr>
<tr>
<td>Intermediate 3</td>
<td>1</td>
</tr>
<tr>
<td>Payoff dominant</td>
<td>1</td>
</tr>
</tbody>
</table>

The equilibria can be Pareto ranked. The equilibrium payoffs are simply the probabilities to be active given in Table 3.2 multiplied by 10. Hence, $(0,0)$ is the secure and inefficient equilibrium and $(1,1)$ is the payoff dominant equilibrium. The remaining ones are intermediate equilibria and are ranked according to their expected payoffs.

The game is played for 30 rounds with the same group of 10 participants. In rounds 1-10 and rounds 21-30 the game is the one described above. In rounds 11-20 we introduced the following variant: each
player with degree 3 receives a discount if she chooses action 1. She does not pay the cost $c$. Thus, if she chooses action 1, she earns a payoff of 13.5 if all her neighbours also choose action 1, and she earns 3.5 if one or more of her neighbours choose action 0.

**Proposition 2:** Both treatments have the following two symmetric BNEs in the game with the incentive: (i) all players choose action 0, and (ii) all players choose action 1.

**Proof:**

The expected payoffs for players with degree 1 remain unchanged since the incentive does not apply. For players with degree 3 the expected payoffs are:

- $E_3(\pi|1, \sigma_{-i}) = 3.5 + 10(0.125\sigma_1^3 + 0.375\sigma_1^2\sigma_3 + 0.375\sigma_1\sigma_3^2 + 0.125\sigma_3^3) = 3.5 + 1.25(\sigma_1^3 + 3\sigma_1^2\sigma_3 + 3\sigma_1\sigma_3^2 + \sigma_3^3)$ in treatment 1
- $E_3(\pi|1, \sigma_{-i}) = 3.5 + 10(0.05\sigma_1^3 + 0.45\sigma_1^2\sigma_3 + 0.45\sigma_1\sigma_3^2 + 0.05\sigma_3^3) = 3.5 + 0.5(\sigma_1^3 + 9\sigma_1^2\sigma_3 + 9\sigma_1\sigma_3^2 + \sigma_3^3)$ in treatment 2

As before, $\sigma^* = (0, 0)$ is a BNE and $\sigma^* = (0, 1)$ is not a BNE. The proofs are identical to the ones set out for Proposition 1.

Now we prove that $\sigma^* = (1, 1)$ is a BNE. Note that under this strategy profile $s_i = 1$ for all players. The expected equilibrium payoffs are $E_1(\pi|1, \sigma^*_i) = 10$ and $E_3(\pi|1, \sigma^*_i) = 13.5$. For a player either of degree 1 or 3 to deviate to a strategy $\sigma_i < 1$, i.e. a strategy where action 0 is played with strictly positive probability is not profitable because the expected
payoff from action 0 is 3.5, i.e. $E_1(\pi|0, \sigma_{i*}) = E_3(\pi|0, \sigma_{i*}) = 3.5$. Then the expected payoffs from $\sigma_i$ are $E_1(\pi|\sigma_i, \sigma_{-i}) = 10\sigma_i + 3.5(1 - \sigma_i) < 10, \forall \sigma_i < 0$ and $E_3(\pi|\sigma_i, \sigma_{-i}) = 13.5\sigma_i + 3.5(1 - \sigma_i) < 13.5, \forall \sigma_i < 0$.

Now we prove that $\sigma^* = (1,0)$ is not a BNE. Note that under this strategy profile still

$$s_i = \begin{cases} 1 & \text{with probability } 0.5 \\ 0 & \text{with probability } 0.5 \end{cases}$$ if $d_i = 1$ in both treatments

$$s_i = \begin{cases} 1 & \text{with probability } 0.125 \\ 0 & \text{with probability } 0.875 \end{cases}$$ if $d_i = 3$ and treatment 1

$$s_i = \begin{cases} 1 & \text{with probability } 0.05 \\ 0 & \text{with probability } 0.95 \end{cases}$$ if $d_i = 3$ and treatment 2.

The expected equilibrium payoffs are $E_1(\pi|1, \sigma_{i*}) = 5$ and $E_3(\pi|0, \sigma_{i*}) = 3.5$ in both treatments. For a player of degree 3 to deviate to a strategy $\sigma_i > 0$, i.e. a strategy where action 1 is played with strictly positive probability is profitable because the expected payoff from action 1 is $E_3(\pi|1, \sigma_{i*}) = 3.5 + 0.125 \cdot 10 = 4.75$ in treatment 1 and $E_3(\pi|1, \sigma_{i*}) = 3.5 + 0.05 \cdot 10 = 4$ in treatment 2. Then the expected payoff from $\sigma_i$ is $E_3(\pi|\sigma_i, \sigma_{-i}) = 4.75\sigma_i + 3.5(1 - \sigma_i) > 3.5, \forall \sigma_i > 0$ in treatment 1 and $E_3(\pi|\sigma_i, \sigma_{-i}) = 3.5\sigma_i + 3.5(1 - \sigma_i) > 3.5, \forall \sigma_i > 0$ in treatment 2. By the same arguments as above there are no mixed strategy BNE for either degree 1 or degree 3.

Now we prove that strategy profiles $\sigma^* = (\sigma_1, \sigma_3)$ where $\sigma_3 \in (0,1)$ are not equilibria $\forall \sigma_1$. The equilibrium condition for a player with degree 3 is that $E_3(\pi|1, \sigma_{i*}) = E_3(\pi|0, \sigma_{i*})$. We note that the right-hand side is equal to 3.5 while the left-hand side (under incentives) is strictly greater than 3.5 for all strategy profiles different from (0,0). Then the condition cannot be satisfied for all strategy profiles where $\sigma_3 \in (0,1)$.
Now we prove that strategy profiles $\sigma^* = (\sigma_1, 0)$ are not equilibria $\forall \sigma_1$. A player with degree 3 has incentive to deviate to action 1 because $E_3(\pi|0, \sigma_{-i}^*) = 3.5$ and $E_3(\pi|1, \sigma_{-i}^*) > 3.5$ for all strategy profiles different from $(0, 0)$.

Now we prove that strategy profiles $\sigma^* = (\sigma_1, 1)$ are not equilibria $\forall \sigma_1$. The equilibrium conditions on a player with degree 1 are $E_1(\pi|0, \sigma_{-i}^*) = E_1(\pi|1, \sigma_{-i}^*)$. The left-hand side is equal to 3.5 while for the right-hand side is straightforward to show that is strictly larger than 5.

QED.

The incentive rules out equilibria where players with degree 1 choose action 1 with some positive probability and players with degree 3 choose action 0. Indeed, the incentive provides players with degree 3 a strictly higher expected payoff when there is a strictly positive probability to have three neighbours who choose action 1. Therefore, action 0 is weakly dominated for players with degree 3.

### 3.2.3 Implementation

Recall that action 0 is labelled ‘inactive’ and action 1 is labelled ‘active’. We will follow this convention in the remainder of the paper. The experiment was conducted as follows. Participants downloaded App Lab, an online economics laboratory environment, to their mobile device. They registered as a user and subscribed to the game using a code. At the
scheduled time, the online game opened. First, players entered their PayPal email address to receive payment (see Appendix C.1). Next, players read the instructions and answered a set of questions to check their understanding of the game (included in Appendix C.2) before the start of the first round.

During the game payoffs were in tokens. At the start of each round, participant received an endowment of 3.5 tokens and were privately informed that they had one connection or that they had three connections. They faced a choice to be active at a cost 3.5 tokens or to be inactive at no cost. If they chose to be active, they earned 10 tokens if all their neighbours were also active and they earned 0 tokens if one or more of their neighbours were inactive. If they chose to be inactive, they kept their endowment of 3.5 tokens and were unaffected by the actions of their connections. Each subject received private information about events at the end of each round. They were reminded whether they had one connection or three connections, and they were informed how many of their connections chose to be active and inactive. Finally, their choice and the resulting payoff were displayed.

Before the start of round 11, all subjects were informed that a discount now applied and players with three connections did not pay the cost for the active choice. Before the start of round 21, all subjects were informed that the discount had expired, and once again everyone paid the cost for the active choice.
After 30 rounds of play the game finished and three rounds were randomly selected for payment: one round from rounds 1 to 10, one round from rounds 11 to 20, and one round from rounds 21 to 30. Before showing participants their final payoff, we tested risk attitudes. Subjects were offered the option to invest up to half of their participation fee in a risky project. The risk test followed the design by Charness and Gneezy (2010) and the implemented version is included in Appendix C.3. The final payoffs consisted of the points earned in the three randomly selected rounds, the (remainder of) the participation fee and the return on the risky project, if applicable. Finally, tokens were converted to cash at a rate of 1 token = GBP 0.40. Average payments were around £12 for a 30-minute session, plus the time taken to install the app and to register as a user.

We conducted a total of 10 sessions, five sessions of each treatment. A total of 100 participants were recruited from the experimental subject pool of the ExpReSS Lab at Royal Holloway, University of London.

3.3 Results

In this section we analyse our experimental data. We focus on answering the three questions that we can address within this pilot. Firstly, can the discount targeted at players with degree 3 move coordination to the efficient outcome? Secondly, if so, does the effect persist after the discount expires? And thirdly, does the probability distribution of the network formation process affect outcomes?
We review descriptive results in section 3.3.1 and we perform an econometric analysis in section 3.3.2.

### 3.3.1 Descriptive results

Table 3.3 displays for each treatment the frequency of active choices in each set of 10 rounds. We first note that a complete set of observations yields 250 decisions for each degree in each set of 10 rounds, but we collected between 232 and 248 decisions. During two sessions for Treatment 1 and during one session for Treatment 2, one participant left before the start of the first round. Furthermore, the game was played online, and participants could physically be in any location that allows an internet connection. Occasionally a participant missed a single round, either due to an issue with their connection or for some other reason. In these cases, the participant re-joined the game in the current round when they re-appeared.

#### Table 3.3: Activity by treatment and degree

| Rounds | Treatment 1 | | Treatment 2 | | |
|--------|-------------|-----------------|-------------|
|        | 1-10 | 11-20 | 21-30 | 1-10 | 11-20 | 21-30 |
| Active | # total | # total | # total | # total | # total | # total |
|        | % | % | % | % | % | % |
| degree = 1 | 193 | 236 | 211 | 244 | 185 | 236 |
|          | 82% | 86% | 78% | 78% | 81% | 74% |
| degree = 3 | 149 | 232 | 214 | 236 | 160 | 244 |
|          | 64% | 91% | 66% | 51% | 85% | 39% |
In both treatments, activity rates of players with degree 1 started around 80 percent in the first 10 rounds, then increased by 3 to 4 percent when the incentive took effect, and in the last 10 rounds fell to 4 percent below the initial activity rate. The activity rates for players with degree 3 followed the same trend, but differences were much larger both within and between treatments. The incentive increased the activity rate by 27 percentage points from 64 to 91 percent in Treatment 1, and by 34 percentage points from 51 to 85 percent in Treatment 2. After the incentive expired, the activity rate returned to its pre-incentive level in Treatment 1, but it dropped to 39 percent in Treatment 2.

We take a closer look at the evolution of play in Figure 3.1, which shows the evolution of aggregate activity rates by degree for each treatment. In the first 10 rounds of Treatment 1, we observe a steady activity rate for players with degree 1 and a declining activity rate for players with degree 3. Activity rates in the last 10 rounds are a little lower but stable for players of either degree. The picture is different for players with degree 3 in Treatment 2. The decline in activity rates that started in the first 10 rounds continued in the last 10 rounds. The discount for players with degree 3 improved coordination while it was available but after it expired, activity rates stabilised in Treatment 1 and continued to decline in Treatment 2.
3.3.2 Econometrics

We estimate the probability of being active using a logistic function where the dependent variable is the action (0 for inactive, 1 for active), and the independent variables include session, round, participant, risk aversion and the interaction between the set of rounds (rounds 1-10 for the baseline, round 11-20 for the incentive, round 21-30 for post-incentive) and degree. Errors are clustered at the session level. Table 3.4 shows the results of four regression models. We find that model_1 is the most parsimonious and provides the best fit, but we note that risk aversion is weakly significant in Treatment 1.
### Table 3.4: Logit regressions for the choice to be active or inactive

<table>
<thead>
<tr>
<th>Variable</th>
<th>model_1</th>
<th>model_2</th>
<th>model_3</th>
<th>model_4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session</td>
<td></td>
<td></td>
<td></td>
<td>-0.089</td>
</tr>
<tr>
<td>Round</td>
<td></td>
<td></td>
<td></td>
<td>0.013</td>
</tr>
<tr>
<td>Participant</td>
<td></td>
<td></td>
<td></td>
<td>0.037</td>
</tr>
<tr>
<td>Risk</td>
<td>-0.149 *</td>
<td>-0.145 *</td>
<td>-0.144 *</td>
<td></td>
</tr>
<tr>
<td>Set # degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 1-10 # degree 3</td>
<td>-0.916 ***</td>
<td>-0.915 ***</td>
<td>-0.906 ***</td>
<td>-0.907 ***</td>
</tr>
<tr>
<td>Round 11-20 # degree 1</td>
<td>0.354</td>
<td>0.368 *</td>
<td>0.376 *</td>
<td>0.248</td>
</tr>
<tr>
<td>Round 11-20 # degree 3</td>
<td>0.773 ***</td>
<td>0.787 ***</td>
<td>0.791 ***</td>
<td>0.665 **</td>
</tr>
<tr>
<td>Round 21-30 # degree 1</td>
<td>-0.213</td>
<td>-0.216</td>
<td>-0.213</td>
<td>-0.474 **</td>
</tr>
<tr>
<td>Round 21-30 # degree 3</td>
<td>-0.857</td>
<td>-0.854</td>
<td>-0.848</td>
<td>-1.113 ***</td>
</tr>
<tr>
<td>Constant</td>
<td>1.501 ***</td>
<td>1.156 ***</td>
<td>0.967 ***</td>
<td>1.158</td>
</tr>
<tr>
<td>N</td>
<td>1,428</td>
<td>1,428</td>
<td>1,428</td>
<td>1,428</td>
</tr>
<tr>
<td>$r^2$</td>
<td>0.0548</td>
<td>0.0676</td>
<td>0.0692</td>
<td>0.0719</td>
</tr>
</tbody>
</table>

| **Treatment 2**           |         |         |         |         |
| Session                   |         |         |         | 0.123   |
| Round                     |         |         |         | 0.032   |
| Participant               |         |         |         | -0.099 *|
| Risk                      | -0.095  | -0.083  | -0.088  |         |
| Set # degree              |         |         |         |         |
| Round 1-10 # degree 3     | 1.225 ***| 1.234 ***| 1.265 ***| 1.270 ***|
| Round 11-20 # degree 1    | 0.217   | 0.220   | 0.232   | 0.552 **|
| Round 11-20 # degree 3    | 0.454   | 0.453   | 0.441   | 0.771 * |
| Round 21-30 # degree 1    | -0.235  | -0.239  | -0.260  | 0.385   |
| Round 21-30 # degree 3    | -1.701 ***| -1.711 ***| -1.726 ***| -1.101 |
| Constant                  | 1.258 ***| 1.038 ***| 1.629 ***| 0.801   |
| N                         | 1,463   | 1,463   | 1,463   | 1,463   |
| $r^2$                     | 0.1036  | 0.1080  | 0.1195  | 0.1251  |

***, **, * denote significance at 1%, 5% and 10% levels, respectively.
We first explore the effect of having degree 3. We use model_1 and calculate the marginal effect of having degree 3 as the change in the probability to be active compared to the base level of having degree 1 in the first 10 rounds. Table 3.5 displays the results. A player with degree 3 in Treatment 1 (Treatment 2) is 17.6 (27) percent less likely to be active in the first 10 rounds (Baseline) than a player with degree 1 in the first 10 rounds. The effect is significant at the 1% level for both treatments. When the discount takes effect in the second 10 rounds (Incentive), players with degree 3 are as likely to be active as players with degree 1 in the first 10 rounds in either treatment. After the discount expires in the last 10 rounds (Post-incentive), players with degree 3 are 12.8 (34.4) percent less likely to be active than players with degree 1 in the first 10 rounds. Note that in Treatment 1, the difference with degree 1 is smaller Post-incentive than in the Baseline. On the contrary, in Treatment 2 the probability to be active is significantly lower at the 1% level Post-incentive, and more so than in the Baseline.

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1</th>
<th>Treatment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>-0.176***</td>
<td>-0.270***</td>
</tr>
<tr>
<td><strong>Incentive</strong></td>
<td>0.042</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>Post-incentive</strong></td>
<td>-0.128</td>
<td>-0.344***</td>
</tr>
</tbody>
</table>

***, **, * denote significance at 1%, 5% and 10% levels, respectively
In our final analysis we explore the marginal effects of the incentive and of persistence by comparing degree 1 to degree 1 and degree 3 to degree 3 between sets of rounds. We calculate the marginal effect of the incentive as the change in the probability to be active in the second 10 rounds, when the discount applies, compared to the first 10 periods. The marginal effect of persistence is the change in the probability to be active in the last 10 rounds, after the discount expires, compared to the second 10 rounds that have the discount. Furthermore, we calculate marginal effects for model_1 without the risk aversion variable and for model_2 with the risk aversion variable. Within model_2, we consider players who did not invest any of their participation fee as high risk averse and players who invested the maximum as low risk averse. Table 3.6 displays the results. N/A refers to the model without risk aversion.

Table 3.6: Marginal effects of incentive by treatment, degree and risk aversion on the probability to be active

<table>
<thead>
<tr>
<th></th>
<th>Treatment 1</th>
<th></th>
<th>Treatment 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk aversion</td>
<td></td>
<td>Risk aversion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N/A</td>
<td>High</td>
<td>Low</td>
<td>N/A</td>
</tr>
<tr>
<td>degree = 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive</td>
<td>0.265 ***</td>
<td>0.315 ***</td>
<td>0.392 ***</td>
<td>0.339 ***</td>
</tr>
<tr>
<td>Persistence</td>
<td>-0.251</td>
<td>-0.300 *</td>
<td>-0.377 **</td>
<td>-0.456 ***</td>
</tr>
<tr>
<td>degree = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incentive</td>
<td>0.047 *</td>
<td>0.061 *</td>
<td>0.084</td>
<td>0.035</td>
</tr>
<tr>
<td>Persistence</td>
<td>-0.081</td>
<td>-0.102</td>
<td>-0.137</td>
<td>-0.078 *</td>
</tr>
</tbody>
</table>

***, **, * denote significance at 1%, 5% and 10% levels, respectively
We first consider degree 3. A player with degree 3 in Treatment 1 (Treatment 2) is 26.5 (33.9) percent more likely to be active than a player with degree 3 in the first 10 rounds when we do not consider risk aversion. The effect of the discount is significant at the 1% level. The result is the same when we consider risk aversion, but the increase in the probability to be active is larger for those with low risk aversion than for those with high risk aversion. The marginal effect of persistence is negative in either treatment, irrespective of risk attitude. In Treatment 1, the fall in the probability to be active is larger and more significant (5% level) for those with low risk aversion. But this is less surprising when we consider the larger positive effect of the discount. In Treatment 2 we find no suggestion of persistence as the fall in the probability to be active is statistically significant at the 1% level for any risk attitude.

We now consider degree 1. We observe weakly significant (10% level) positive spill-over effects of the incentive for the high risk averse in Treatment 1. In Treatment 2, the lower probability to be active for players with degree 3 negatively affects the probability to be active for players with degree 1 and the effect is significant at the 5% level regardless of risk attitude.
3.4 Conclusion

We design and report on a partial pilot study that explores the effect of a temporary financial incentive aimed at a subgroup of agents in a network. We focus on an environment where agents have either one connection or three connections and where a temporary financial incentive is available only to those with three connections. Actions have complements and payoffs are determined by the minimum of neighbour’s actions. We have two treatments that differ only in the probability distribution over neighbour’s degrees for agents with three connections. In our first treatment we find that the incentive has a significant positive effect and players coordinate on the efficient equilibrium. But activity rates started from a high baseline and we could not clearly identify whether the effect was persistent. In our second treatment we discourage activity from players with degree 3 in the baseline by decreasing the probability of three connections who each have only one connection. We again observe that the incentive has a significant positive effect, but we conclude that the effect was not persistent. Given the subtlety of the adjustment between Treatment 1 and Treatment 2, however, the difference between outcomes is impressive.

We plan to expand the research to include further interactions between the nature of the game and the nature of the statistic that determines the payoffs.
APPENDIX

APPENDIX A: INTERVENTION – CONDITIONAL INCENTIVES

A.1 Logistic regression with random effects

Definitions

\[ s_1 = 1 \text{ if treatment is Baseline, 0 otherwise} \]
\[ s_2 = 1 \text{ if treatment is TH14, 0 otherwise} \]
\[ s_3 = 1 \text{ if treatment is TH20, 0 otherwise} \]
\[ s_4 = 1 \text{ if treatment is TH14_D1, 0 otherwise} \]
\[ d_1 = 1 \text{ if player's degree=1, 0 otherwise} \]
\[ d_2 = 1 \text{ if player's degree=2, 0 otherwise} \]
\[ d_3 = 1 \text{ if player's degree=3, 0 otherwise} \]
\[ d_4 = 1 \text{ if player's degree=4, 0 otherwise} \]
\[ d_{2s_2} = \text{interaction variable between degree 2 and treatment TH14} \]
\[ d_{2s_3} = \text{interaction variable between degree 2 and treatment TH20} \]
\[ d_{2s_4} = \text{interaction variable between degree 2 and treatment TH14_D1} \]
\[ d_{3s_2} = \text{interaction variable between degree 3 and treatment TH14} \]
\[ d_{3s_3} = \text{interaction variable between degree 3 and treatment TH20} \]
\[ d_{3s_4} = \text{interaction variable between degree 3 and treatment TH14_D1} \]
\[ d_{4s_2} = \text{interaction variable between degree 4 and treatment TH14} \]
\[ d_{4s_3} = \text{interaction variable between degree 4 and treatment TH20} \]
\[ d_{4s_4} = \text{interaction variable between degree 4 and treatment TH14_D1} \]
\[ s_{2p} = \text{interaction variable between treatment TH14 and period} \]
\[ s_{3p} = \text{interaction variable between treatment TH20 and period} \]
\[ s_{4p} = \text{interaction variable between treatment TH14_D1 and period} \]
\[ d_{2p} = \text{interaction variable between degree 2 and period} \]
\[ d_{3p} = \text{interaction variable between degree 3 and period} \]
d4p = interaction variable between degree 4 and period

d2s2p = interaction variable between degree 2, treatment TH14 and period

d2s3p = interaction variable between degree 2, treatment TH20 and period

d2s4p = interaction variable between degree 2, treatment TH14_D1 and period

d3s2p = interaction variable between degree 3, treatment TH14 and period

d3s3p = interaction variable between degree 3, treatment TH20 and period

d3s4p = interaction variable between degree 3, treatment TH14_D1 and period

d4s2p = interaction variable between degree 4, treatment TH14 and period

d4s3p = interaction variable between degree 4, treatment TH20 and period

d4s4p = interaction variable between degree 4, treatment TH14_D1 and period
Random-effects logistic regression

| choice | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-------|--------|-----------|-------|-----|------------------------|
| period | -.0552117 | .0129211 | -4.27 | 0.000 | -.0805367 | -.0298867 |
| s2     | 5.155412 | .5045481 | 10.22 | 0.000 | 4.166516 | 6.144308 |
| s3     | 1.204348 | .4771722 | 2.52  | 0.012 | .2691075 | 2.135958 |
| s4     | 2.49967 | .4680518 | 5.34  | 0.000 | 1.582305 | 3.417034 |
| d2     | 2.822659 | .3016963 | 9.36  | 0.000 | 2.231345 | 3.413973 |
| d3     | 3.416411 | .3422982 | 12.60 | 0.000 | 3.643719 | 4.985503 |
| d4     | 8.662458 | .8112826 | 10.68 | 0.000 | 7.072374 | 10.25254 |
| d2s2   | -.0807463 | .4821607 | -0.17 | 0.867 | -1.025764 | .8642731 |
| d2s3   | -.3186328 | .3931919 | -0.81 | 0.418 | -1.089275 | .4520091 |
| d2s4   | -.621769 | .3898536 | -1.59 | 0.111 | -.3855686 | .14233 |
| d3s2   | -.0357116 | .6189444 | -0.06 | 0.954 | -1.24882 | 1.177397 |
| d3s3   | -.2763421 | .4625493 | -0.60 | 0.550 | -1.182922 | .6302379 |
| d3s4   | -.4875382 | .4751892 | -1.03 | 0.305 | -.418892 | .4438154 |
| d4s2   | -.290862 | 1.749438 | -1.31 | 0.190 | -5.719698 | 1.137974 |
| d4s3   | -.9348778 | 1.205789 | -0.78 | 0.438 | -3.29818 | 1.426425 |
| d4s4   | -3.115681 | .9921802 | -3.14 | 0.002 | -5.060318 | -1.171043 |
| s2p    | .0603676 | .0160926 | 3.75  | 0.000 | .0288266 | .0919085 |
| s3p    | -.0070769 | .0163951 | -0.43 | 0.666 | -.0392106 | .0205569 |
| s4p    | .0150906 | .0149058 | 1.01  | 0.311 | -.0141243 | .0443055 |
| d2p    | .0043435 | .0149591 | 0.29  | 0.772 | -.0249757 | .0336628 |
| d3p    | .0138418 | .0158131 | 0.88  | 0.381 | -.0171512 | .0468348 |
| d4p    | -.0355539 | .0284227 | -1.23 | 0.219 | -.0922806 | .0211729 |
| d2s2p  | -.0648539 | .0213877 | -3.03 | 0.002 | -.106773 | -.0223948 |
| d2s3p  | .0029072 | .0193513 | 0.15  | 0.881 | -.0350207 | .0408352 |
| d2s4p  | -.0382656 | .0184464 | -2.07 | 0.038 | -.07441 | -.0021031 |
| d3s2p  | -.0688064 | .0253311 | -2.72 | 0.007 | -.1184544 | -.0191584 |
| d3s3p  | .0035079 | .0209414 | 0.17  | 0.867 | -.0375365 | .0445523 |
| d3s4p  | -.047844 | .0205719 | -2.33 | 0.020 | -.0861642 | -.0075237 |
| d4s2p  | .0234794 | .0708386 | 0.33  | 0.740 | -.1153622 | .1623201 |
| d4s3p  | .0022231 | .042171 | 0.05  | 0.958 | -.0804305 | .0884768 |
| d4s4p  | -.0235623 | .0355405 | -0.66 | 0.507 | -.0932205 | .0406959 |
| _cons  | -.2894054 | .3569421 | -8.11 | 0.000 | -3.593640 | -2.194461 |

/lnsig2u  1.616716 .1025931 1.415637 1.817795

sigma_u  2.24422 .1151207 2.029559 2.481585
rho      .6048867 .0245196 .555962 .651797

LR test of rho=0: chibar2(01) = 3377.34 Prob >= chibar2 = 0.000
A.2 Marginal effects estimation

Definitions:

\( s_{2\_d1} = \text{TH14 vs Baseline, degree 1} \)
\( s_{2\_d2} = \text{TH14 vs Baseline, degree 2} \)
\( s_{2\_d3} = \text{TH14 vs Baseline, degree 3} \)
\( s_{2\_d4} = \text{TH14 vs Baseline, degree 4} \)
\( s_{3\_d1} = \text{TH20 vs Baseline, degree 1} \)
\( s_{3\_d2} = \text{TH20 vs Baseline, degree 2} \)
\( s_{3\_d3} = \text{TH20 vs Baseline, degree 3} \)
\( s_{3\_d4} = \text{TH20 vs Baseline, degree 4} \)
\( s_{4\_d1} = \text{TH14\_D1 vs Baseline, degree 1} \)
\( s_{4\_d2} = \text{TH14\_D1 vs Baseline, degree 2} \)
\( s_{4\_d3} = \text{TH14\_D1 vs Baseline, degree 3} \)
\( s_{4\_d4} = \text{TH14\_D1 vs Baseline, degree 4} \)
\( s_{23\_d1} = \text{TH20 vs TH14, degree 1} \)
\( s_{23\_d2} = \text{TH20 vs TH14, degree 2} \)
\( s_{23\_d3} = \text{TH20 vs TH14, degree 3} \)
\( s_{23\_d4} = \text{TH20 vs TH14, degree 4} \)
\( s_{43\_d1} = \text{TH20 vs TH14\_D1, degree 1} \)
\( s_{43\_d2} = \text{TH20 vs TH14\_D1, degree 2} \)
\( s_{43\_d3} = \text{TH20 vs TH14\_D1, degree 3} \)
\( s_{43\_d4} = \text{TH20 vs TH14\_D1, degree 4} \)
\( s_{42\_d1} = \text{TH14 vs TH14\_D1, degree 1} \)
\( s_{42\_d2} = \text{TH14 vs TH14\_D1, degree 2} \)
\( s_{42\_d3} = \text{TH14 vs TH14\_D1, degree 3} \)
\( s_{42\_d4} = \text{TH14 vs TH14\_D1, degree 4} \)
| choice  | Coef.     | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------|-----------|-----------|------|------|----------------------|
| s2_d1   | 0.8960614 | 0.0239391 | 37.43| 0.000| 0.8491416             |
|         |           |           |      |      | 0.9429812             |
| s2_d2   | 0.7282332 | 0.051058  | 14.26| 0.000| 0.6281614             |
|         |           |           |      |      | 0.8283049             |
| s2_d3   | 0.3520153 | 0.0628993 | 5.60 | 0.000| 0.228735              |
|         |           |           |      |      | 0.4752955             |
| s2_d4   | 0.0186292 | 0.007701  | 2.42 | 0.016| 0.0035356             |
|         |           |           |      |      | 0.0337228             |
| s3_d1   | 0.0324109 | 0.0144848 | 2.24 | 0.025| 0.0040211             |
|         |           |           |      |      | 0.0608006             |
| s3_d2   | 0.1770245 | 0.082047  | 2.16 | 0.031| 0.0162153             |
|         |           |           |      |      | 0.3378336             |
| s3_d3   | 0.1658758 | 0.0757941 | 2.19 | 0.029| 0.017322              |
|         |           |           |      |      | 0.3144296             |
| s3_d4   | 0.002935  | 0.0103873 | 0.28 | 0.778| -0.0174237            |
|         |           |           |      |      | 0.0232936             |
| s4_d1   | 0.2140273 | 0.0480427 | 4.45 | 0.000| 0.1198654             |
|         |           |           |      |      | 0.3081893             |
| s4_d2   | 0.3289068 | 0.0827962 | 3.97 | 0.000| 0.1666132             |
|         |           |           |      |      | 0.4911684             |
| s4_d3   | 0.2313815 | 0.0701109 | 3.30 | 0.001| 0.0939666             |
|         |           |           |      |      | 0.3687964             |
| s4_d4   | -0.0215678| 0.015445  | -1.40| 0.163| -0.0518395             |
|         |           |           |      |      | 0.0087039             |
| s23_d1  | -0.8636505| 0.0269208 | -32.08| 0.000| -0.9164143            |
|         |           |           |      |      | -0.8108868            |
| s23_d2  | -0.5512087| 0.0648788 | -8.50| 0.000| -0.6783687            |
|         |           |           |      |      | -0.4240487            |
| s23_d3  | -0.1861395| 0.0423245 | -4.40| 0.000| -0.269094             |
|         |           |           |      |      | -0.103185             |
| s23_d4  | -0.0156942| 0.0099895 | -2.54| 0.012| -0.0293933            |
|         |           |           |      |      | -0.001951             |
| s43_d1  | -0.181165 | 0.0496416 | -3.66| 0.000| -0.2788339            |
|         |           |           |      |      | -0.0843991            |
| s43_d2  | -0.1518663| 0.0919491 | -1.65| 0.099| -0.3320832            |
|         |           |           |      |      | 0.0283506             |
| s43_d3  | -0.0655057| 0.0524069 | -1.25| 0.211| -0.1682213            |
|         |           |           |      |      | 0.0372099             |
| s43_d4  | 0.0245027 | 0.0151346 | 1.62 | 0.105| -0.0051605            |
|         |           |           |      |      | 0.0541659             |
| s42_d1  | 0.6820341 | 0.0530894 | 12.85| 0.000| 0.5779803             |
|         |           |           |      |      | 0.7860879             |
| s42_d2  | 0.3993424 | 0.0657084 | 6.08 | 0.000| 0.2705562             |
|         |           |           |      |      | 0.5281285             |
| s42_d3  | 0.1206338 | 0.0309846 | 3.89 | 0.000| 0.0599051             |
|         |           |           |      |      | 0.1813624             |
| s42_d4  | 0.040197  | 0.0134375 | 2.99 | 0.003| 0.0183599             |
|         |           |           |      |      | 0.066534              |
A.3: Evolution of play by treatment and by session

BASELINE GAME

ACTIVITY RATE BY GROUP AND DEGREE

TREATMENT: BASELINE

degree = 1  
degree = 2  
degree = 3  
degree = 4

Period
UNIVERSAL PREMIUM GAME WITH $X = 20$

ACTIVITY RATE BY GROUP AND DEGREE

TREATMENT: TH20

degree = 1  degree = 2  degree = 3  degree = 4

Period

124
UNIVERSAL PREMIUM GAME WITH $X = 14$

ACTIVITY RATE BY GROUP AND DEGREE

TREATMENT: TH14

degree = 1

degree = 2

degree = 3

degree = 4

Period
TARGETED PREMIUM GAME

ACTIVITY RATE BY GROUP AND DEGREE

TREATMENT: TH14_D1

degree = 1
degree = 2
degree = 3
degree = 4

Period
A.4: Game instructions

Only instructions for TH14 are included. The corresponding instructions for TH14_D1, TH20 and Baseline are analogues.

INSTRUCTIONS

The aim of this experiment is to study how individuals make decisions in certain contexts. The instructions are simple. If you follow them carefully you will earn a non-negligible amount of money in cash (GBP) at the end of the experiment. During the experiment, your earnings will be accounted in ECU (Experimental Currency Units). Individual earnings will remain private, as nobody will know the other participants’ earnings. Any communication among you is strictly forbidden and will result in an immediate exclusion from the experiment.

1. The experiment consists of 40 periods, and there are 20 participants, including yourself. The participants will remain the same throughout the experiment. At each period, you and each of the remaining nineteen participants will be assigned one position of the following NETWORK. The positions in the network are numbered from 1 to 20.
2. In the network, a link is represented by a line (connection) between two positions. For example, position 16 has four links: it is linked to positions 6, 7, 15 and 19 (but it is not linked to the remaining positions).

Note that there are four classes of positions in the network, identified by different colours.

- There are eight yellow positions: Those positions with one link (1, 2, 3, 6, 9, 10, 12 and 14).
- There are six green positions: Those positions with two links (7, 8, 13, 15, 17 and 18).
- There are four blue positions: Those positions with three links (5, 11, 19 and 20).
- There are two red positions: Those positions with four links (4 and 16).

3. At each period, you (and the other participants) are randomly assigned by the computer to a position from 1 to 20 in the network, all of them being equally likely. The assignment process is random: At each period, you are equally likely to be located in each of the 20 positions of the network.

At each period, you will only be informed of the colour of your position, that is, you will know how many links your assigned position has: 1 link (yellow), 2 links (green), 3 links (blue) or 4 links (red). However, you will not be informed of which is your exact position.
For example, if at a particular period you are informed that your position has 3 links (blue), then you know that you can be in position 5, 11, 19 or 20, and that you can be in any of them with the same probability. Note that, in such a case, you also know that you cannot be in yellow, green or red positions.

Your earnings for the period are affected by your decisions and the decisions of the other participants, as specified below.

4.- At each period, knowing the network and your position, you will be asked to make a choice: to be ACTIVE or INACTIVE (the other participants are asked to make the same choice). Your payoff for the period will depend on your choice and on the choices of those participants located in positions linked to yours, as well as overall activity in the network. If you choose to be INACTIVE, your period payoff is 50 ECU. If you choose to be ACTIVE, your period payoff is calculated as follows: First, add 100 ECU per participant linked to you that also chooses to be ACTIVE; then, divide the result by 3. If in any period 14 or more participants in the network, including yourself, are active, 33,33 ECU are added to your period payoff. Hence,
• If you choose to be **ACTIVE** your period payoff can be:

- **166,66 ECU** if 4 participants linked to you choose to be ACTIVE and any **14 or more** participants, including yourself, are ACTIVE: $\frac{100+100+100+100}{3} + 33,33$, or

- **133,33 ECU** if 4 participants linked to you choose to be ACTIVE and **less than 14** participants, including yourself, are ACTIVE: $\frac{100+100+100+100}{3}$, or

- **133,33 ECU** if 3 participants linked to you choose to be ACTIVE and any **14 or more** participants, including yourself, are ACTIVE: $\frac{100+100+100}{3} + 33,33$, or

- **100,00 ECU** if 3 participants linked to you choose to be ACTIVE and **less than 14** participants, including yourself, are ACTIVE: $\frac{100+100+100}{3}$, or

- **100,00 ECU** if 2 participants linked to you choose to be ACTIVE and any **14 or more** participants, including yourself, are ACTIVE: $\frac{100+100}{3} + 33,33$, or

- **66,66 ECU** if 2 participants linked to you choose to be ACTIVE and **less than 14** participants, including yourself, are ACTIVE: $\frac{100+100}{3}$, or

- **66,66 ECU** if 1 participant linked to you chooses to be ACTIVE and any **14 or more** participants, including yourself, are ACTIVE: $\frac{100}{3} + 33,33$, or

- **33,33 ECU** if 1 participant linked to you chooses to be ACTIVE and **less than 14** participants, including yourself, are ACTIVE: $\frac{100}{3}$, or

- **33,33 ECU** if no participant linked to you chooses to be ACTIVE and any **14 or more** participants, including yourself, are ACTIVE $\frac{0}{3} + 33,33$, or

- **0,00 ECU** if no participant linked to you chooses to be ACTIVE and **less than 14** participants, including yourself, are ACTIVE: $\frac{0}{3}$
5. At the end of every period, you will get information about current and past periods. The information consists of:
   - Your position in the network.
   - Your choice (ACTIVE or INACTIVE).
   - The number of participants linked to you that chose to be ACTIVE.
   - The number of participants in the network that chose to be ACTIVE.
   - Your (period) payoff from participants linked to you (component A).
   - Your (period) payoff from overall activity (component B).
   - Your total (period) payoff.

6. Payoffs. At the end of the experiment, you will be paid the earnings that you achieved in 4 periods, that will be randomly selected across the 40 periods of play (all periods selected with the same probability).
   These earnings are transformed to cash at the exchange rate of 40 ECU = 1 GBP. In addition to your earnings in the experiment, you will receive GBP 4.00 for your participation.

7. The experiment starts with 5 trial periods that do not count for earnings. Trial periods are indicated as such in the top left corner of the screen. After the 5 trial periods the experiment starts without a break.
APPENDIX B: NETWORK STRUCTURE AND RUMOURS

B.1: Game instructions

Instructions for the network treatment. The instructions for the broadcast treatment are analogues.

Welcome to the RHUL ExpReSS Lab

Welcome and thank you for participating in today’s experiment.

Please switch your phone to silent and put it in the container provided on your desk now, so we can have your undivided attention.

Please do not touch the computer until you are instructed to do so. When using the computer, please use it as instructed only and do not attempt to browse the internet or to launch any programs that are not part of the experiment.

You will use the computer for the entire experiment. **DO NOT** socialise, talk or communicate with other participants – doing so will result in immediate exclusion.

You can start reading the instructions now. The instructions are in two sections:

1. The Experiment: This section presents a summary of the game
2. Instructions: This section provides detailed instructions and screen shots of the game

If you have a question at any time, please raise your hand.
**The Experiment**

Today’s experiment is an experiment in decision making. You are in a World that is either **Blue** or **Red** - as determined by the computer.

Overall, **7 times out of 10 the World is Blue**, and **3 times out of 10 the World is Red**. You earn money by voting for Blue or Red. Which one depends on your preference and on your beliefs. Most players prefer to vote for Blue when the World is Blue and for Red when the World is Red, but some prefer to vote for Red regardless. Once everyone has voted, **one vote is randomly selected** by the computer and it is **implemented** for all players in your group. The vote that is implemented determines how much money you can earn.

Before you vote, you may (or may not) have some information about whether the World is Blue or Red. **9 times out of 10**, the computer generates a **signal** that truthfully reveals the colour of the World. The signal is sent to **one player** only. A player that receives the signal can create one of **three messages**: a) **Blue**, b) **Red**, or c) **None**. If a message is created, it is sent to the creator’s **neighbours** – who, in turn, can decide to **pass on** (or not) the message to their neighbours. **1 time out of 10**, the computer generates **no signal**.

You play for 25 rounds (periods). You are in a group of **four players** (including yourself) that remains the same for all 25 periods. In the first period, you are assigned to a position in the group and you remain in this position.
**Instructions**

If the computer generates a signal, the player who receives it sees the screen in Figure 1.

![Signal recipient screen](image)

Figure 1: Signal recipient screen

The top of the screen informs you of your position: U1, B2, U3 or U4. In this example, you are in **position U1** and your **neighbour is B2**. The picture in the middle panel shows the same information: a large black arrow points to your position and a smaller arrow shows that any message you create is sent to your neighbour B2.

Note that, in the picture, the four positions are connected by lines. You can send a message only to the positions that you are connected to by a line (your **neighbours**):
• U1 can send to B2
• B2 can send to U1 and U3
• U3 can send to B2 and U4
• U4 can send to U3

You can also receive messages only from your neighbours.

You can communicate only with your neighbours – who are connected to you by lines.

The bottom panel of Figure 1 shows that you received the signal and that the World is Blue. You can now choose to create a message ‘Blue’ or a message ‘Red’ regardless of what the signal says. You can also choose ‘None’ and not create any message. Select the radio button of your choice (1) and then click ‘OK’ (2).

If you do not receive the signal, you may receive a message from one of your neighbours. In the example in Figure 2 below, U1 received the signal and created message ‘Red’. U1’s neighbour B2 receives the message and sees the following screen:
The screen is like Figure 1, only now the bottom panel tells you that ‘You received a message’ from U1, and that the message says: ‘The World is Red’. Click the radio buttons (3): ‘Yes’ to pass on the message to your neighbour U3, or ‘No’ to not pass on the message. Click ‘OK’ to submit your choice. In any period, you receive **one message at most**.

When you receive a message, you can pass it on only to a neighbour that did not send it to you: **for example**, B2 receives a message from U1 and can send it to U3, or B2 receives a message from U3 and can send it to U1.
Once all messages are sent, you **vote** for a colour of the World. Figure 3 below shows the voting screen.

The top left of the voting screen reminds you of your position and the top right shows what happened so far. This example is for position U4, who received the signal and created message ‘*Blue*’.

You vote in the bottom panel by **selecting the button** (4) for one of two **options**: ‘*Blue*’ or ‘*Red*’. Your vote may determine how much you earn.

The dotted circle in Figure 3 highlights that, if U4 votes for ‘*Blue*’, the **potential payoff** is GBP 10.00 if the World turns out to be *Blue* and GBP
0.00 if the World turns out to be Red. Press ‘OK’ to submit your vote. The colour of the World is revealed to you after voting.

**Positions U1, U3 and U4 have the payoffs shown in the picture.** These positions are easily recognised by their round shape and blue/red colour. **The payoffs for position B2,** which is a red square shape, are different and shown in **Figure 4 on the next page.**

![Figure 4: Voting screen (for B2)](image.png)
The payoffs for B2 do not depend on the colour of the World. If B2 votes ‘Blue’, his potential payoff is GBP 0.00 (5). If B2 votes ‘Red’, his potential payoff is GBP 10.00.

Note that there are two types of position identified by a shape and a letter:

- Three blue/red **CIRCLES** denoted by ‘U’: U1, U3 and U4
- One red **SQUARE** denoted by ‘B’: B2

Each player in your group votes, but **only one vote is selected at random and implemented**: that is, the potential earnings of everyone in your group depend on the implemented vote.

**Take the example in Figure 3**: say, vote ‘Blue’ by U4 is selected and implemented. It turns out that the World is **Blue**. Regardless of their own votes, U1 and U3 have a payoff of GBP 10.00 – just like U4. The square position (B2), however, has a payoff of GBP 0.00 whenever vote ‘Blue’ is implemented.

The results of each period are reported to all players in your group. Figure 5 shows and example of the screen you see after voting is completed.
The top part of the screen (6) summarises the events in this period. The middle part (7) tells you how everyone in your group voted, and the bottom part (8) shows you the results.

You play for 25 periods. At the end, **ONE PERIOD is randomly selected for payment.** In addition to your earnings, you also receive a GBP 5.00 show-up fee. Figure 7 on the next page shows the screen you see.
Figure 7: Final results

Please complete the three sets of questions and the mini-game when the screen opens.

Once everyone has finished reading, please check your understanding by answering five questions on the computer. Before the game begins, you will also play three practise periods – the practise periods do not count for your payoffs.

Please remember to raise your hand whenever anything is unclear.
SUMMARY OF THE GAME

1. You are in a group of four players (including yourself)

2. You are assigned to one of the four positions in Figure 1

IN EACH PERIOD:

3. The state of the world is Blue 7/10 times and Red 3/10 times

4. 9/10 times a signal is created that reveals the state of the world to one randomly selected player in your group
   a. This player creates a message, or
   b. This player does not create a message

5. When there is a message, it is passed until
   a. A player decides not to pass it, or
   b. Everyone has received it

6. Each player submits a vote

7. The computer selects one player, whose vote is implemented for everyone in the group

8. The colour of the World is revealed, and payoffs are calculated depending on
B.2: Check Your Understanding Questions

The ‘Check Your Understanding’ questions are answered after reading the instructions and before the three trial periods. Participants answer a question and on the next screen are shown if they were right or wrong, along with an explanation of the correct answer. Participants in Public Broadcast treatment (PB) answer three questions (1, 4 and 5), and participants in the Network treatment (N) answer the same three questions plus two additional questions (2 and 3). Correct answers are bold and underlined.

Participants answer the following questions.

1) Q (N + PB): You were selected to receive the signal and you decide to create a message. Does the message you create have to match the signal?
   A: Yes; **No**.

2) Q (N only): You are in position B2 and receive the signal. If you choose to create a message, who will be the recipient(s)?
   A: U1; U3; **U1 and U3**; Everyone.

3) Q (N only): You are in position U3 and receive a message from U4. If you choose to pass on the message, who will be the recipient(s)?
   A: **B2**: U4; B2 and U4; Everyone.

4) Q (N + PB): Players in your group voted as follows. U1: Blue, B2: Red, U3: Blue, U4: Red. You are in position U4. The world is Blue. The vote by U3 is implemented. What is your potential payoff? (Hint: use Figure)
   A: **GBP 10.00**.
5) Q (N + PB): Players in your group voted as follows. U1: Blue, B2: Red, U3: Blue, U4: Red. You are in position B2. The world is Blue. The vote by U3 is implemented. What is your potential payoff (Hint: use Figure 4).

A: **GBP 0.00**: GBP 10.00.
B.3: Beliefs and expectations

Questions for the network game are shown below for each position. Questions for the public broadcast game are the same except for questions related to transmission, which are omitted.
Questions for position U3

Generally, did you believe message Red received from B2?  
Generally, did you believe message Blue received from B2?  
Generally, did you believe message Red received from U4?  
Generally, did you believe message Blue received from U4?  
Generally, if you sent / passed on message Red, did you expect U4 would believe it?  
Generally, if you sent / passed on message Blue, did you expect U4 would believe it?  
Generally, if you sent / passed on message Red, did you expect B2 would believe it?  
Generally, if you sent / passed on message Red, did you expect B2 would pass it on?  
Generally, if you sent / passed on message Blue, did you expect B2 would believe it?  
Generally, if you sent / passed on message Blue, did you expect B2 would pass it on?

Questions for position U4

Generally, did you believe message Red received from U3?  
Generally, did you believe message Blue received from U3?  
Generally, if you sent message Red, did you expect U3 would believe it?  
Generally, if you sent message Red, did you expect U3 would pass it on?  
Generally, if you sent message Blue, did you expect U3 would believe it?  
Generally, if you sent message Blue, did you expect U3 would pass it on?
B.4: Social preferences questionnaire

Now imagine that you are paired with another subject. You face four pairs of different allocations between you and your partner. Please choose which one you prefer.

1) A: £2 for you and £2 for your partner. B: £2 for you and £1 for your partner.

2) A: £2 for you and £2 for your partner. B: £3 for you and £1 for your partner.

3) A: £2 for you and £2 for your partner. B: £2 for you and £4 for your partner.

4) A: £2 for you and £2 for your partner. B: £3 for you and £5 for your partner.

B.5: Lie aversion questionnaire

Please answer the following questions:

1) Lying is immoral. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

2) It is ok to lie in order to achieve one’s goals. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

3) There is no excuse for lying to someone else. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

4) Honesty is always the best policy. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

5) It is often better to lie than to hurt someone’s feelings. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

6) Lying is just wrong. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

7) Lying is no big deal. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

8) There is nothing wrong with bending the truth now and then. □ strongly agree □ agree □ neutral □ disagree □ strongly disagree

Scoring rules. Items are scored so that higher scores reflect higher levels of lie acceptability: strongly agree (1 point); agree (2 points); neutral (3 points); disagree (4 points); and, strongly disagree (5 points). Questions 2, 5, 7 and 8 are reverse scored.
B.6: Risk test

**Investment Game**

Earlier you earned GBP 0.00 in addition to your show-up fee of GBP 5.00.

In this final game you can invest your GBP 5.00 show-up fee in a risky option.

If your investment is successful, you receive back 2.5 times your investment.
If your investment is unsuccessful, you lose the money you invested.
The chance of success of your investment is 50% (regardless of the amount you invest).

After all investment decisions are made, one player is randomly selected and their decision is implemented.
Only the selected player is awarded the return on investment in the final payment.
The investment decisions of all other players are not implemented and their final payments are unaffected.

How much of your show-up fee would you like to invest in the risky option?

- [ ] GBP 5.00
- [ ] GBP 4.00
- [ ] GBP 3.00
- [ ] GBP 2.00
- [ ] GBP 1.00
- [ ] GBP 0.00

B.7: Network and broadcast treatment subject pool comparison

<table>
<thead>
<tr>
<th>Category</th>
<th>Attitude</th>
<th>Biased agents</th>
<th>Unbiased agents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><em>p</em>-value</td>
<td><em>p</em>-value</td>
</tr>
<tr>
<td>Social Preferences</td>
<td>Pro-sociality</td>
<td>0.6227</td>
<td>0.6453</td>
</tr>
<tr>
<td></td>
<td>Costly Pro-sociality</td>
<td>0.6798</td>
<td>0.2417</td>
</tr>
<tr>
<td></td>
<td>Envy</td>
<td>1.0000</td>
<td>0.7968</td>
</tr>
<tr>
<td></td>
<td>Costly Envy</td>
<td>1.0000</td>
<td>0.8143</td>
</tr>
<tr>
<td>Risk</td>
<td>Risk</td>
<td>0.1553</td>
<td>0.7784</td>
</tr>
<tr>
<td>Lie Acceptability</td>
<td>Average score</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8847</td>
<td>0.4329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5594</td>
<td>0.5981</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1125</td>
<td>0.9617</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.8263</td>
<td>0.3638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5287</td>
<td>0.1997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1736</td>
<td>0.2044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.6751</td>
<td>0.1897</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2350</td>
<td>0.6780</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2927</td>
<td>0.8508</td>
</tr>
</tbody>
</table>
B.8: Network treatment marginal effects estimation

Marginal effect coefficients are interpreted as the percentage point change in the probability of creating the theoretically predicted message when the belief is held versus not held. Marginal effects were estimated following a multinomial logistic regression with errors clustered at the group level. Social preferences were dropped from the regression model because they did not produce any effect (alone or in interactions) and prevented convergence.

<table>
<thead>
<tr>
<th></th>
<th>Position U1</th>
<th>Position B2</th>
<th>Position U3</th>
<th>Position U4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b )</td>
<td>( b )</td>
<td>( b )</td>
<td>( b )</td>
</tr>
<tr>
<td></td>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
<td>(s.e.)</td>
</tr>
<tr>
<td>U1 believes Blue</td>
<td>( -0.001^* )</td>
<td>( 0.000 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.001) )</td>
<td>( (0.000) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U1 believes Red</td>
<td>( 0.138^{***} )</td>
<td>( 0.000 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.046) )</td>
<td>( (0.000) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2 believes Red</td>
<td>( 0.602^{***} )</td>
<td>( 0.000 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.203) )</td>
<td>( (0.000) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2 passes on Red</td>
<td>( 0.002 )</td>
<td>( 0.047^{***} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.003) )</td>
<td>( (0.016) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 believes Blue</td>
<td></td>
<td>( 0.149 )</td>
<td>( 0.706^{***} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( (0.101) )</td>
<td>( (0.228) )</td>
<td></td>
</tr>
<tr>
<td>U3 passes on Blue</td>
<td>( 0.258^{**} )</td>
<td></td>
<td>( 0.007 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.114) )</td>
<td></td>
<td>( (0.063) )</td>
<td></td>
</tr>
<tr>
<td>U3 believes Red</td>
<td>( 0.457^{***} )</td>
<td></td>
<td>( 0.925^{***} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.139) )</td>
<td></td>
<td>( (0.063) )</td>
<td></td>
</tr>
<tr>
<td>U3 passes on Red</td>
<td>( 0.338^{***} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.086) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4 believes Blue</td>
<td></td>
<td>( 0.000 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( (0.000) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U4 believes Red</td>
<td></td>
<td>( 0.000 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( (0.000) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Signal Red</td>
<td>( 0.000 )</td>
<td>( 0.208^{**} )</td>
<td>( 0.000^{**} )</td>
<td>( -0.091 )</td>
</tr>
<tr>
<td></td>
<td>( (0.002) )</td>
<td>( (0.090) )</td>
<td>( (0.000) )</td>
<td>( (0.097) )</td>
</tr>
<tr>
<td>Lie Acceptability</td>
<td>( 0.000 )</td>
<td>( -0.070^{**} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( (0.001) )</td>
<td>( (0.030) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***, **, * denote significance at the 1%, 5% and 10% levels respectively
B.9: Posterior beliefs calculations (Table 2.1)

Unbiased agents form posterior beliefs using Bayes’ rule to evaluate for the event that $\theta = 1$. There are three scenarios: (i) An agent receives message 1, (ii) an agent receives message 0, and (iii) an agent receives no message.

(i) An agent who receives message 1 evaluates:
$$\rho_i(1) = Pr(\theta = 1 | m(j) = 1)$$
$$= \frac{Pr(m(j) = 1 | \theta = 1) Pr(\theta = 1)}{Pr(m(j) = 1 | \theta = 1) Pr(\theta = 1) + Pr(m(j) = 1 | \theta = 0) Pr(\theta = 0)}$$

(ii) An agent who receives message 0 evaluates:
$$\rho_i = Pr(\theta = 1 | m(j) = 0)$$
$$= \frac{Pr(m(j) = 0 | \theta = 1) Pr(\theta = 1)}{Pr(m(j) = 0 | \theta = 1) Pr(\theta = 1) + Pr(m(j) = 0 | \theta = 0) Pr(\theta = 0)}$$

An unbiased agent who receives message 0 from another unbiased agent has posterior belief $\rho_i(0) = \frac{0}{0.7} = 0$. An unbiased agent who receives message 0 from the biased agent has posterior belief $\rho_i(0) = \frac{1}{2^{0.3+2^{0.7}}} = 0.3$.

(iii) An agent who receives no message evaluates:
$$\rho_i(\emptyset) = Pr(\theta = 1 | m(j) = \emptyset)$$
$$= \frac{Pr(m(j) = \emptyset | \theta = 1) Pr(\theta = 1)}{Pr(m(j) = \emptyset | \theta = 1) Pr(\theta = 1) + Pr(m(j) = \emptyset | \theta = 0) Pr(\theta = 0)}$$

We first calculate posterior beliefs $\rho_i(1)$ and $\rho_i(\emptyset)$ for each unbiased agent in the network treatment. Consider $U1$ when receiving message 1. $\rho_{U1}(1) =$
\[
Pr ( \theta = 1 \mid m(B2) = 1 ) = \frac{3\left(\frac{0.9}{4}\right) \cdot 0.3}{3\left(\frac{0.9}{4}\right) \cdot 0.3 + \left(\frac{0.9}{4}\right) \cdot 0.7} = 0.5625.
\]
Consider U1 when receiving no message. The probability of no message when \( \theta = 1 \) is \[
\frac{0.1}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.1290
\]
and the probability of no message when \( \theta = 0 \) is \[
\frac{0.1 + 2 \cdot \left(\frac{0.9}{4}\right)}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.7097.
\]
Then \( \rho_{U1}(\emptyset) = \frac{0.1290 \cdot 0.3}{0.1290 \cdot 0.3 + 0.7097 \cdot 0.7} = 0.0723. \)
Consider agent U3 when receiving message 1. \( \rho_{U3}(1) = Pr ( \theta = 1 \mid m(B2) = 1 ) = \frac{2\left(\frac{0.9}{4}\right) \cdot 0.3}{2\left(\frac{0.9}{4}\right) \cdot 0.3 + \left(\frac{0.9}{4}\right) \cdot 0.7} = 0.4615 \) and \( \rho_{U3}(1) = Pr ( \theta = 1 \mid m(U4) = 1 ) = \frac{\left(\frac{0.9}{4}\right) \cdot 0.3}{\left(\frac{0.9}{4}\right) \cdot 0.3 + 0 \cdot 0.7} = 1. \)
Consider agent U3 when receiving no message. The probability of no message when \( \theta = 1 \) is \[
\frac{0.1}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.1290
\]
and the probability of no message when \( \theta = 0 \) is \[
\frac{0.1 + \left(\frac{0.9}{4}\right)}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.4194.
\]
Then \( \rho_{U3}(\emptyset) = \frac{0.1290 \cdot 0.3}{0.1290 \cdot 0.3 + 0.4194 \cdot 0.7} = 0.1165. \)
Consider agent U4 when receiving message 1. \( \rho_{U3}(1) = Pr ( \theta = 1 \mid m(U3) = 1 ) = \frac{3\left(\frac{0.9}{4}\right) \cdot 0.3}{3\left(\frac{0.9}{4}\right) \cdot 0.3 + \left(\frac{0.9}{4}\right) \cdot 0.7} = 0.5625. \)
Consider agent U4 when receiving no message. The probability of no message when \( \theta = 1 \) is \[
\frac{0.1 + 2 \cdot \left(\frac{0.9}{4}\right)}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.7097
\]
and the probability of no message when \( \theta = 0 \) is \[
\frac{0.1 + \left(\frac{0.9}{4}\right)}{0.1 + 3 \cdot \left(\frac{0.9}{4}\right)} = 0.4194.
\]
Then \( \rho_{U4}(\emptyset) = \frac{0.7097 \cdot 0.3}{0.7097 \cdot 0.3 + 0.4194 \cdot 0.7} = 0.4204. \)

We now calculate posterior beliefs \( \rho_i(1) \) and \( \rho_i(\emptyset) \) for each unbiased agent in the broadcast treatment. Unbiased agents in the broadcast treatment are identical so we calculate posterior beliefs for only one case, U1.
Consider unbiased agent $U_1$ when receiving message 1. $\rho_{U_1}(1) = \Pr(\theta = 1 \mid m(B2) = 1) = \frac{1 \cdot 0.3}{1 \cdot 0.3 + 1 \cdot 0.7} = 0.3$ and $\rho_{U_1}(1) = \Pr(\theta = 1 \mid m(U3) = 1) = \frac{1 \cdot 0.3}{1 \cdot 0.3 + 0 \cdot 0.7} = 1$. Consider unbiased agent $U_1$ when receiving no message. The probability of no message when $\theta = 1$ is 

$$\frac{0.1}{0.1 + 3 \cdot \frac{0.9}{4}} = 0.1290$$

and the probability of no message when $\theta = 0$ is 

$$\frac{0.1}{0.1 + 3 \cdot \frac{0.9}{4}} = 0.1290.$$ 

Then $\rho_{U_1}(\emptyset) = \frac{0.1290 \cdot 0.3}{0.1290 \cdot 0.3 + 0.1290 \cdot 0.7} = 0.3$. 

152
APPENDIX C: INTERVENTION - PERSISTENCE

C.1: App screenshots

App home screen

First page of game

Payment Information
The money you earn during this experiment will be paid into your PayPal account today or tomorrow. If you have any problems with your payment, please follow the instructions in the invitation email.

Please enter your PayPal email address and your name (example: a.participant@hotmail.com Adrian Participant):

Next:
C.2: Game instructions

Welcome!

In this experiment you play a game in which you can earn experimental tokens. At the end of the experiment, the tokens are exchanged for real money at a rate of 1 token for GBP 0.40. In addition to your earnings in the game, you also receive 10 tokens (GBP 4.00) for participating when you complete the game. The experiment is expected to last no more than 60 minutes.

You are in a group of ten players, including yourself, and you are arranged in a network. There are two possibilities: either one other player in your group connects to you, or three other players in your group connect to you. It is equally likely that you have one connection or that you have three connections.

If you have one connection, there are two possibilities:

- **10 times out of 20** your connection has one connection
- **10 times out of 20** your connection has three connections

If you have three connections, there are four possibilities:

- **1 time out of 20**, all of your connections have one connection each
- **9 times out of 20**, two of your connections have one connection each and one of your connections has three connections
- **9 times out of 20**, one of your connections has one connection and two of your connections have three connections each
- **1 time out of 20**, all of your connections have three connections each

You only know that you have one connection or that you have three connections, and the probabilities given above. Below are two examples to illustrate.

**Example 1:** One player connects to you. You have no information about the other players, who are shown for illustration only.

![Example 1](image1)

**Example 2:** Three players connect to you. You have no information about the other players, who are shown for illustration only.

![Example 2](image2)

You receive an endowment of 3.5 tokens. You can use your endowment to play the game. In the game, you decide to be active or inactive. There is a cost for being active of 3.5 tokens and there is no cost for being inactive. Your payoffs depends on your decision and the decision(s) of your connection(s) as follows:

- If you are active and all your connections are active, your payoff is 10 tokens
- If you are active and one or more of your connections are inactive, your payoff is 0 tokens
- If you are inactive, your payoff is 3.5 tokens regardless

The payoffs are displayed in the table below.

<table>
<thead>
<tr>
<th>Your choice</th>
<th>Your connection(s) choice(s)</th>
<th>Your endowment</th>
<th>Your cost</th>
<th>Your earnings</th>
<th>Your payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>All are active</td>
<td>3.5</td>
<td>3.5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Active</td>
<td>One or more are inactive</td>
<td>3.5</td>
<td>3.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inactive</td>
<td></td>
<td>3.5</td>
<td>0</td>
<td>0</td>
<td>3.5</td>
</tr>
</tbody>
</table>

You play the game for 30 rounds. You receive your endowment at the start of each round. The way in which the network is connected and the players that connect to you change in each round.

After 30 rounds, the computer randomly selects three rounds: one from rounds 1–10, one from rounds 11–20 and one from rounds 21–30. Your payoffs in these three rounds are added to your final payment.
Check Your Understanding

1. If I am active and all my connections are active, my payoff is my endowment of _______ plus my earning of _______, minus my cost of _______, which equals _______ tokens.

2. If I am active and one or more of my connections are inactive, my payoff is my endowment of _______ plus my earning of _______, minus my cost of _______, which equals _______ tokens.

3. If I am inactive, my payoff is my endowment of _______ plus my earnings of _______, minus my cost of _______, which equals _______ tokens.

4. It is equally likely that I have one connection or three connections. True or false?

<table>
<thead>
<tr>
<th>Your choice</th>
<th>Your connection(s) choice(s)</th>
<th>Your endowment</th>
<th>Your cost</th>
<th>Your earnings</th>
<th>Your payoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>All are active</td>
<td>3.5</td>
<td>3.5</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Active</td>
<td>One or more are inactive</td>
<td>3.5</td>
<td>3.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inactive</td>
<td></td>
<td>3.5</td>
<td>0</td>
<td>0</td>
<td>3.5</td>
</tr>
</tbody>
</table>

C.3: Risk test

Investment Game

You have finished the game.

You now have the option to invest in a risky project if you would like. You can invest up to half of your participation fee of 10 tokens (GBP 4.00). The other half of you participation fee and any tokens that you do not invest are still paid to you for sure.

The risky project has a 50% chance of success:

- If the project is successful, you will receive 2.5 times the amount you chose to invest.
- If the project is unsuccessful, you will lose the amount you chose to invest.

Note that you can pick any number between 0 and 5 tokens, including 0 or 5.

Once all players have submitted their investment choices, the computer randomly selects one player. If you are the selected player, your investment is implemented and the result counted in your final payment. If you are not the selected player, your investment is not implemented and your final payment is not affected by your investment choice.

Please select your investment:

[Slider]

Your choice: 1.8
C.4: Activity rates by treatment and session

TREATMENT 1:

[Graphs showing activity rates by treatment and session]
TREATMENT 2:
BIBLIOGRAPHY


Available at SSRN: https://ssrn.com/abstract=2995809 or http://dx.doi.org/10.2139/ssrn.2995809


https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm


https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies


