Connectors and Influencers

Online Appendix

Frédéric Moisan[1] Syngjoo Choi[2] Sanjeev Goyal[3]

[1]emlyon business school, CNRS, Université Lumière Lyon 2, Université Jean Monnet Saint-Etienne, GATE, 69007 Lyon, France

[2]Department of Economics, Seoul National University, Seoul 08826, Republic of Korea

[3]Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, UK

A Theory

There are no general equilibrium characterization results available for this model. The analysis of Galeotti and Goyal (2010) focuses on polar cases in which $a_1 = 1$ and $a_l = 0$, for all $l \geq 2$ and the case where $a_l = 1$, for all l. Our formulation allows for indirect flow of benefits with decay; this appears to be a natural case.

Proof of Proposition 1: The first step is to observe that in equilibrium, every individual must access at least $\hat{y}$. This is true because if someone is accessing less than $\hat{y}$, then due to the concavity of the $f(.)$ function, she can simply increase her utility by raising effort so that the total access equals $\hat{y}$.

The second step is to show that players will form one link or zero link, for sufficiently large linking costs. Observe that an isolated individual will choose $\hat{y}$. So it follows that in a network with connections, no one will ever choose more than $\hat{y}$. Note that if link costs are close to $c\hat{y}$ then it is not profitable to form links with two individuals who each chooses $\hat{y}$. So the only situation in which an individual, A, may choose two or more links arises if an individual accesses significantly more than $\hat{y}$ through each link. Consider a link between A and B. Iterating on optimal effort, it is true that if B chooses $\hat{y}$ then every neighbor of B must choose 0. So A accesses more than $\hat{y}$ only if B chooses strictly less than $\hat{y}$. If a neighbor of B chooses a positive effort, then it must be the case that this person must meet the first order condition on optimal efforts: her total efforts invested and accessed must equal $\hat{y}$. As this person is a neighbor of B, it follows that A cannot access more than $\hat{y}$ via the link with B. So, A will form at most one link in equilibrium.

The third step considers effort configurations. Take the situation in which some individual (say) A chooses $\hat{y}$. It is optimal for everyone else to choose effort 0 and form a link with this person. And it is clearly optimal for A to choose $\hat{y}$ when faced with zero efforts by everyone else.

To conclude the proof, we need to show that the pure connector outcome is the only possible equilibrium in a situation where no player chooses $\hat{y}$. Observe first that the pure connector outcome is an equilibrium so long as $k < c\hat{y}(n-2)a_2/(1+(n-2)a_2)$ or $n \geq 2 + \frac{k}{(c\hat{y}-k)a_2}$. Observe that $c\hat{y}(n-2)a_2/(1+(n-2)a_2)$ converges to $c\hat{y}$, as n gets large.

The next step is to rule out any other possible equilibrium. The key observation here is that any equilibrium network must have diameter less than or equal to 2. Suppose the diameter of a component is 3 or more. We know from step 2 that the component must be acyclic. So consider two furthest apart leaf nodes. A variant of the 'switching' argument, developed in Bala and Goyal (2000), shows that one of the two leaf players has a strict incentive to deviate. So every component must have diameter 2. Given that the network is acyclic, this implies it must be a star. It is now possible to apply standard agglomeration arguments to deduce that multiple components cannot be sustained in equilibrium.

Finally, the hub player must choose zero. Suppose not. By hypothesis the hub chooses less than $\hat{y}$. Given that $a_1 = 1$ and $a_2 < 1$, both the hub and the spokes cannot be accessing exactly $\hat{y}$. A contradiction that implies that the hub must choose zero effort.

■

We next show that, under discrete values of personal effort, a sufficiently high cost of linking implies a pure influencer equilibrium (for any group size n) and a pure connector ϵ equilibrium (for a sufficiently large group size; e.g., with $\epsilon < 2$ for group size 50).

Proposition 2. *Suppose payoffs are given by (1), $a_1 = 1$, $a_2 \in (0,1)$, $\hat{y} \in X = \{0,1,2,\ldots,\overline{x}\}$, and $c(\hat{y}+a_2-1) < k < c\hat{y}$. Every Nash equilibrium $s^* = (x^*, g^*)$ is such that g^* is a star.*

(a) For $n \geq 3$, there exists a pure influencer outcome: the hub chooses $\hat{y}$ and every spoke chooses 0.

(b) If $n \geq 2 + \frac{\hat{y}-1}{a_2}$, there also exists a pure connector ϵ-equilibrium, in which the hub chooses 0, m spokes invest 1 and the remaining other spokes invest 0.

Proof of Proposition 2: It follows from Proposition 1 that the pure influencer equilibrium always hold, regardless of n. Moreover, the pure connector equilibrium holds only if $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)}$, in which case it requires every spoke to personally invest $\frac{\hat{y}}{1+(n-2)a_2}$. Since $c(\hat{y} + a_2 - 1) < k$ implies $\hat{y} < \frac{k}{c} + 1$, we have that $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)} > 2 + \frac{\hat{y}-1}{a_2}$, and consequently $\frac{\hat{y}}{1+(n-2)a_2} < 1$ for any $n \geq 2 + \frac{\hat{y}-1}{a_2}$. However, since the lowest positive effort that can be made in the game is 1, there is a limited number of $m < n$ players who can benefit from making such minimum positive effort. In this case, each of those positive investors accesses $(m-1)a_2$ from forming a link and therefore earns $f(1+(m-1)a_2) - c - k$. They would earn $f(\hat{y}) - c\hat{y}$ should they form no link and invest $\hat{y}$. They would also earn $f((m-1)a_2) - k$, should they maintain their link and invest 0. The pure connector outcome is an ϵ-equilibrium whenever $\epsilon > f(\hat{y}) - f(1 + (m-1)a_2) + c(\hat{y} - 1) + k$ and $\epsilon > f((m-1)a_2) - f(1+(m-1)a_2) + k$, where m is the number of investing spokes. Given the definition of f from (1), those conditions are satisfied for some values of $\epsilon \geq 0$ and $m \leq n$.

■

The combination of high linking costs and integer action levels leads to some complications with regard to the pure connector outcome that we now spell out. As noted above, a pure connector equilibrium does not exist for $n = 4$ or $n = 8$. In the treatments with 50 and 100 subjects there exists an ϵ-equilibrium in which 18 peripheral individuals choose 1 and the rest of the subjects choose 0. The equilibrium is 'approximate' because the periphery player who chooses effort 1 and forms a link with the hub earns 79.25 whereas she could earn 81 by deleting the link and instead choosing effort level 9. This asymmetry in behavior among the peripheral players makes the coordination problem more challenging than in the pure influencer outcome where it is solely about determining who is to be the hub.

B Simulations: Distance to the star network

In order to better interpret the distance between the networks observed in the experiment and the star network, we compute the same distance from random networks that we simulate. More specifically, given a group size n, we generate networks with m links (with m varying from 0 to $3n$) that are added uniformly at random. Figure A1 presents the resulting distance to star in relation to the different group sizes ($n \in \{4, 8, 50, 100\}$) and the average number of links per individuals (i.e., the total number of links m divided by the group size n). For any group size n and any number of links m, 100 random networks are generated.

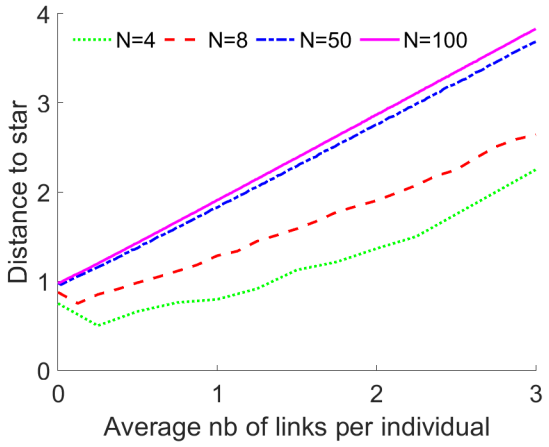


Figure A1: Average distance from random networks to the star (for different group sizes and numbers of links)

Figure A1 reveals interesting results: the distance to star increases almost linearly with the number of links across all group sizes. Moreover, larger random networks are further away from the star than the small random networks. In particular, given $n-1$ links in the network (as in the star network), the average distance to star for random networks is 0.76 for $n = 4$, 1.18 for $n = 8$, 1.81 for $n = 50$, and 1.90 for $n = 100$.

C Experimental platform

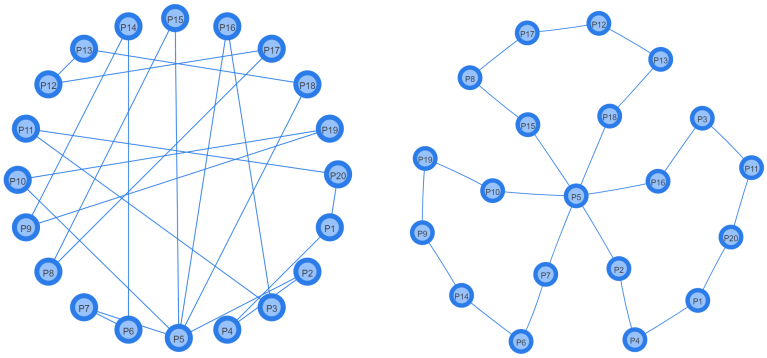
C.1 Network visualization

Existing studies of network formation in economics have considered small group sizes such as 4 or 8 people and visualized evolving networks with fixed positions of nodes (e.g., Goyal et al. (2017); van Leeuwen et al. (2019)). When the group size increases, such a representation of networks with fixed positions of nodes makes it very difficult for subjects to perceive network features. For subjects to learn their optimal choices, they must have a good idea of the evolving networks. An appropriate tool for visualizing networks is thus critical in running the experiment in continuous time. This leads us to develop an experimental software including an interactive network visualization tool that allows the network to automatically reshape itself in response to decisions made by subjects. We use force-directed algorithms to visualize networks in real time (see, e.g., Eades (1984), Fruchterman and Reingold (1991), Hu (2005), Bostock et al. (2011), and Jacomy et al. (2014)).[21]

Clearly, different network environments will offer differ levels of transparency and information on network architecture. Our strategy is to start with a visualization approach that is efficient and that allows us to systematically explore the effects of different variables – such as scale and variations in information on networks. Of course, the experimental platform is flexible enough to incorporate other ways of representing networks and can be used to explore the effects of network visualization itself on human behavior and network formation. Thus, the experiment reported in this paper could be interpreted as benchmark findings with efficient network visualization.

The network structure in Figure A2a can be represented in a transparent manner in Figure A2b with the network visualization tool we use. In our large-scale experiment, this visualization tool improves graphical clarity of evolving networks and helps subjects distin-

[21] Such algorithms are common, and have been previously used in Gallo and Yan (2015).



(a) Fixed visualization (b) Adaptive visualization

Figure A2: Examples of network visualization

guish between those who are more connected and those who are less connected. Note that this feature does not aim at reproducing any cognitive representation of networks (i.e., how people mentally visualize networks), but instead attempts to facilitate access to information about the network structure, which people may use to make decisions. It is important to emphasize that this tool allows interaction between the subject and the network: while the nodes are subject to the above attraction and repulsion forces, they can also be freely manipulated by the participant through the usual drag-select functionality. The creation and removal of links is also interactive through double-clicking on corresponding nodes. This network visualization tool is built on the open source Javascript library *vis.js*.

The force-directed algorithms of the network visualization tool use attraction and repulsion forces between nodes in the network and gravity force toward the center of the screen, in order to readjust their positions in two-dimensional space and improve the overall visibility on the subjects' screen.

Any two nodes o and o' in the network repulse each other with a *repulsion force* $F_r(o, o')$ in order to avoid overlaps and allow a sparse visualization of the network. It is modelled as a decreasing function of the Euclidean distance between two nodes $dist(o, o')$, implying

that close nodes repulse more than distant nodes. Two connected nodes o and o' in the network apply an *attractive force* $F_a(o, o')$ towards each other to allow for visual proximity. A classical approach of modelling attraction force is a linear and positive relation with the distance, implying that close nodes attract less than distant nodes. Finally, every node o applies a *gravity force* $F_g(o)$ to the center of the spatialisation space O to pull the entire network towards the center of the screen. In particular, such a force allows disconnected components to be within reasonable distance from each other, and therefore more easily visualized on the screen.

The net force vector applied to any node o resulting from the above three forces is then given by the following form of weighted sum (where F_x and F_y represent corresponding force vectors applied to the x and y axes of the Euclidean space respectively):

$$F_x(o) = \frac{x_O - x_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{x_{o'} - x_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{x_{o''} - x_o}{dist(o, o'')} F_r(o, o'') \quad (3)$$

$$F_y(o) = \frac{y_O - y_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{y_{o'} - y_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{y_{o''} - y_o}{dist(o, o'')} F_r(o, o'') \quad (4)$$

Note that the computation of the repulsion force for every node can be a complex task, especially in the context of large networks. In order to address this issue, the experimental software approximates this computation using the well-known algorithm introduced by Barnes and Hut (1986). More concretely, it finds groupings of nodes in proximity and determines a repulsion force $F_r(o, c)$ between node o and the group of nodes with a center of mass c, in replacement of the brute force method of computing repulsion forces between all pairs of nodes. More details of this approximation algorithm are provided at the following website: `http://networks.econ.cam.ac.uk/net_formation/connectors_influencers.html`.

We turn back to Figure A2 to derive some intuition of how the net force equations

aggregate forces for every node and the network is visualized in the two-dimensional space. The adaptive visualization in Figure A2b is obtained by using the force-directed algorithm. The network has a petal-like structure with three independent sub-components connected through a common player, P5. The visualization algorithm makes P5 to be located at the center of the screen because the neighbors of P5 repluse each other and surround P5, while each pair of P5's neighbors belonging to the same sub-component are in close proximity and positioned side by side. The three forces then operate to make the rest of players located to draw non-overlapping petal-like structures.

Dynamic adjustment. The above equations (3) and (4) describe the net forces that are applied for the visualization of the network, given the positions of all nodes and the links between nodes. When the network changes, the algorithm updates dynamically the network visualization by computing the corresponding velocity of nodes on both coordinate axes.

In order to get a sense of how the network visualization is updated, we turn again to the example of network visualization in Figure A2 and show how the algorithm makes the transition from the fixed visualization in Figure A2a to the adaptive visualization in Figure A2b. Six (slow-motion) snap shots of the transition are presented at the following website: `http://networks.econ.cam.ac.uk/net_formation/connectors_influencers.html`. They show how the hub player, P5, moves from the bottom of the fixed circle to the center of the screen, and the petal-like structures emerge. This dynamic adjustment occurs rapidly to arrive at Figure A2b.

In our large-scale experiment, this visualization tool improves graphical clarity of evolving networks and helps subjects distinguish between those who are more connected and those who are less connected. It is wothwhile to note that this tool allows interaction between the subject and the network: while the nodes are subject to the above attrac-

tion and repulsion forces, they can also be freely manipulated by the participant through the usual drag-select functionality. The creation and removal of links is also interactive through double-clicking on corresponding nodes. This network visualization tool is built on the open source Javascript library *vis.js*.

Model parameter setting used in the experiment: $K_g = -2000$, $K_s = 0.04$, $K_{cg} = 0.3$, $L = 95$, $D = 0.09$, $T = 0.5$, $V_{min} = 0.3$, $V_{max} = 10$.

Continuous time with asynchronous choices. We elaborate on some of the technical aspects of the platform in relation to these points. Running the continuous time experiments in large groups poses a number of technical challenges. First, every action made by a subject on her computer must be updated instantly on the computer screens of all other participants through the server computer. Network visualization must also be correspondingly updated in real time. As the group size increases, the information flows across the computer network increases dramatically. This can cause communication congestion and lagged responses. Another challenge with a large scale experiment is that it is constrained by the limited capacity of existing laboratories. Large groups that cannot fit into a single lab therefore require remote interactions between subjects in different geographical locations (that is, across different labs). In order to handle both of these technical challenges, we use a Websocket protocol with enhanced two-way communication between the server and subjects' computers. It fits naturally into the environment of asychronous choices in real time and the updates are made only when necessary. Our Websocket technology relies on the Javascript run-time environment *Node.js*. Since it only requires an internet connection and is compatible with most existing web browsers (e.g., Google Chrome, Mozilla Firefox, Internet Explorer), this technology makes no specific restriction on the physical location of every participant. **Information on Payoffs.** We turn finally to information on payoffs: clearly subjects need to be able to see their own payoffs in order to learn the prof-

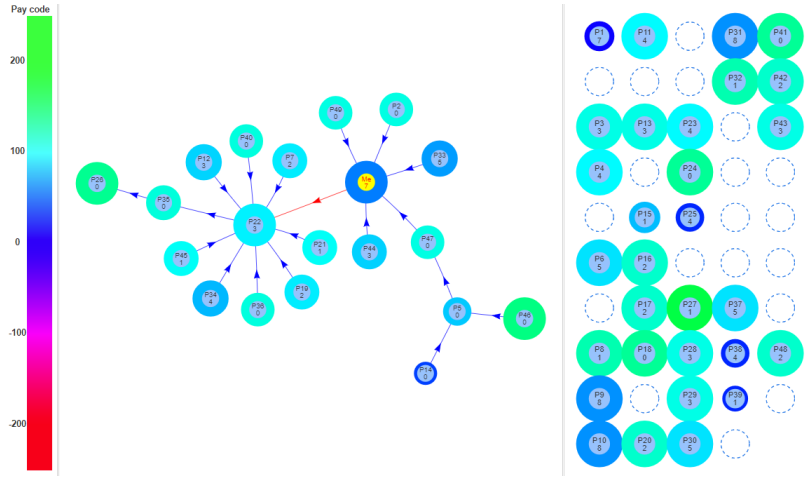


Figure A3: Payoff Information

itability of different linking and effort combinations. In the baseline treatments, subjects are shown their own payoffs but *not* others' payoffs. A subject is also shown the efforts and public good access for all other subjects. To facilitate learning, we add information about every player's payoff through a set of color codes as illustrated by Figure A3. Specifically, the border of every node is coloured: the colour varies from green (high positive payoff) to red (high negative payoff). The scale of the colour code is presented at all times on the left hand side, as in Figure A3.

D Experimental instructions

[In the following instructions, N is to be replaced with any value from $\{3, 7, 49, 99\}$ *depending on the treatment]*

Please read the following instructions carefully. **These instructions are the same for all the participants**. The instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. One of

the experimenters will answer your question.

You can earn money by earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you have earned will be exchanged at the following exchange rate:

100 points = 1 Euro

The money you earn will be paid out in cash at the end of the experiment. The other participants will not see how much you earned.

Details of the experiment

The experiment consists of 6 (six) independent rounds of the same form. The first round is for practice and does not count for your payment. The next 5 rounds will be counted for your payment.

At the beginning of each round, you will be grouped with N other participants. This group will remain fixed throughout the 6 rounds. Each of the participants will be randomly assigned an identification number of the form "Px" where x is a number between 1 and N. Those numbers will be randomly changed across every round of the experiment. The actual identity of the participants will not be revealed to you during or after the experiment. The participants will always be represented as blue circles on the decision screen. You are always represented as a yellow circle identified as "ME".

Each round will last **6 (six) mins: the first minute will be a trial period, only the latter 5 minutes will be relevant for the earnings**. Your earnings in a given round will be based on everyone's choice **at a randomly selected moment in the last 5 mins of the round**. In other words, any decision made before or after that randomly chosen moment will not be used to determine your points. This precise moment will be

announced to everyone only at the end of the round, along with the corresponding behavior and earnings.

At the beginning of the experiment, you are given an initial balance of 500 points. Your final earnings at the end of the experiment will consist of the sum of points you earn across the 5 last rounds plus this initial capital (the first round will be used to familiarize yourself with the game and will have no influence on your earnings). Note that if your final earnings (i.e., the sum of your earnings across the 5 last rounds plus the initial endowment) go below 0, your final earnings will be simply treated as 0.

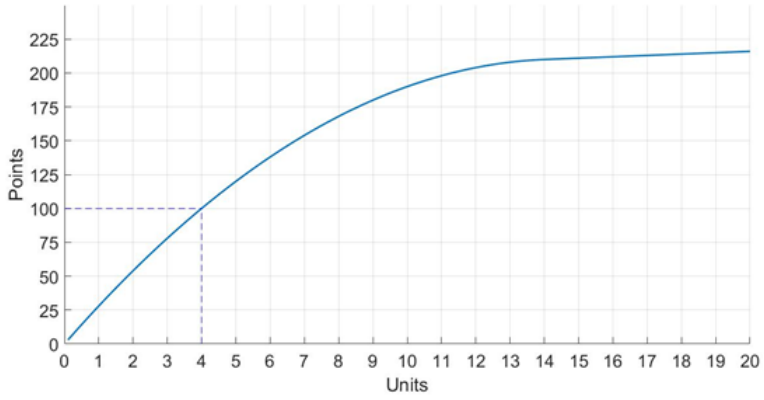
In each round, every participant will have choose two types of actions:

- **How many any units to buy/invest**: You may buy at most 20 units. Each unit costs you **11 points**.
- **Add/delete links with other participants**: You are linked with another person if you form a link with that person or that person forms a link with you (or both). You do not pay any fee for links formed by others. The people that you are linked with (regardless of whether you or they form the links) are called your neighbors. You automatically have access to **all units bought by your neighbors** as well as **half of the units bought by your neighbors' neighbors** (see below for an example). Each link you form costs you **95 points**.

You may revise your choices at any moment before the round ends. During a round, you will also be informed about every other participant's most recent decision (units bought and formed links), which will be updated every 5 seconds or whenever you change your own choice.

At any moment, the total number of units you have access to (i.e., units you bought + units bought by your neighbors + units bought by your neighbors' neighbors) generates

points for you according to the following figure (for example, accessing 4 units generates 100 points, as shown by the dotted lines):



Moreover, having access to 20+m units generates 216+m points.

The computer screen will be split into two parts:

- **The middle side of the screen presents you and your local neighborhood**. More precisely, you will see your neighbors, the neighbors of your neighbors, and the neighbors of neighbors' neighbors. In other words, you will see the participants that are up to 3 links away from you.

- **The right side of the screen presents participants outside of your local neighborhood**.

- **The left side of the screen presents the code for the players' net earnings in the network**. *[Payoff information treatment only]* The inner circle of each node from the middle or right part side of the screen is characterized by some color, which varies from **green** (high positive net payoff) to **red** (high negative net payoff) depending on the player's corresponding net earnings.

Each node is described by their identification number "Px" and the number of units

that they buy. Identification numbers "Px" are randomly assigned in every round. Therefore, every player is likely to have a different ID in different rounds. In the initial state of the network, nobody buys any unit and no link is formed.

Tutorial

Please follow this simple tutorial simulating a simple virtual scenario on the computer screen. In this tutorial you are interacting with 9 other players. In the initial state, you are not linked with anyone and you do not buy any units: you start at 0 points.

1. The slider allows you to choose how many units you wish to buy yourself. For example, buying 4 units costs you 44 points (= 4 units $\times$ 11 points, in red on the screen) and generates 100 points (according to the figure from the previous page, in green on the screen).

2. Initially, the nodes on the right side of the screen represent all other players (in this simulation, those players are not real people). The size of node reflects the total number of units bought by that node and the units accessed via the network. For example, P1-P4 are the largest nodes because these players have access to the most units.

3. You may choose to form a link with any player by simply double clicking on the corresponding node. For example, forming a link with P4 reveals that P1, P2, and P3 each form a link with P4, and P9 forms a link with P1. Forming a link with P4 costs you 95 points (in red on the screen), but it also gives you access to 8.5 units (7 from P4 + 0.5 $\times$ 1 from P1 + 0.5 $\times$ 1 from P2 + 0.5 $\times$ 1 from P3), which generates 174 points (according to the above figure, describing the benefit function in green on

the screen). If you do not buy any additional unit yourself, your resulting net payoff is **79 points (= 174 points − 1 link × 95 points)**.

4. After forming a link with P4, you observe that some nodes remain unobserved (P5, P6, P7, and P8 on the right side). However, forming an additional link with P9 (by double clicking on the corresponding node) reveals that those nodes all form a link with P9. You were not allowed to observe them before because they were 4 nodes away from you (for example, P5 were connected to you via P4, P1, and P9). You can now observe them because they are only 2 nodes away from you (for example, P5 is connected to you via P9 only). Remember that you can only see players that are at most 3 nodes away. Assuming you still do not buy any unit yourself, your resulting net payoff is **16 points (= 206 points from accessing 12.5 units − 2 links × 95 points)**.

5. Alternatively, you may choose to remove a link that you previously formed by double clicking on the corresponding node. For example, after forming links with P4 and P9, removing the link with P4 leads to players P2 and P3 becoming unobserved again, as they are now more than 3 nodes away from you.

6. Note that varying the amount of units you buy directly affects the sizes of the nodes you are linked with as well as their neighbors. Indeed, the amount of units they each have access to includes the units you buy (the larger this amount, the larger the node).

7. You may also shape the visual structure of the network by dragging nodes as it pleases you.

Summary

Here is a brief description of information available on the decision screen:

1. The timer indicates elapsed time since the beginning of the round. Any round lasts **6 mins**. A moment will be randomly selected **in the last 5 mins** to determine everyone's payoff. The time displayed will turn red when entering this interval.

2. **Only decisions made at the randomly selected moment in the round** matter to directly determine the earnings. The payoff may be negative at the end of a round. However, starting from a balance of 500 pts, any negative total of points at the end of the 5 rounds will be equivalent to 0 point.

3. The amount of units you have access is equal to the sum of **(1)** the units bought by you, **(2)** the units bought by your neighbors, and **(3)** half of the units bought by your neighbors' neighbors.

4. You are represented as the yellow node, and your ID is "ME".

5. Every other node's ID is represented as "Px" (inside the node) where x is a number. Every node has a unique ID, which is randomly reassigned in every round.

6. The size of each node determines **how many units that node has access to** (units bought personally plus units accessed from others, directly and indirectly).

7. The amount of units **bought personally by** a player is mentioned inside the corresponding node.

8. *[Payoff information treatment only]* The color of each node determines **that node's net earnings** according to the code depicted on the left side of the screen.

E Questionnaires

At the end of the experiment, subjects answered a set of surveys aiming at measuring various types of individual differences. More precisely, incentivized measures of comprehension in network game, social preferences, and risk preferences were used. Finally non incentivized personality measures were used before which subjects filled up a debriefing questionnaire that includes demographics information.

E.1 Comprehension check

In order to assess the subjects' comprehension of the network game played during the experiment, we provided 5 questions, each of which with a unique correct answer. Each correct answer was rewarded with 0.1 euro for the subject.

The following first 2 questions were used across all treatments (correct answers are "11 pts" to question 1, and "95 pts" to question 2).

Question 1: In the previous game, how many points did investing one unit cost you?

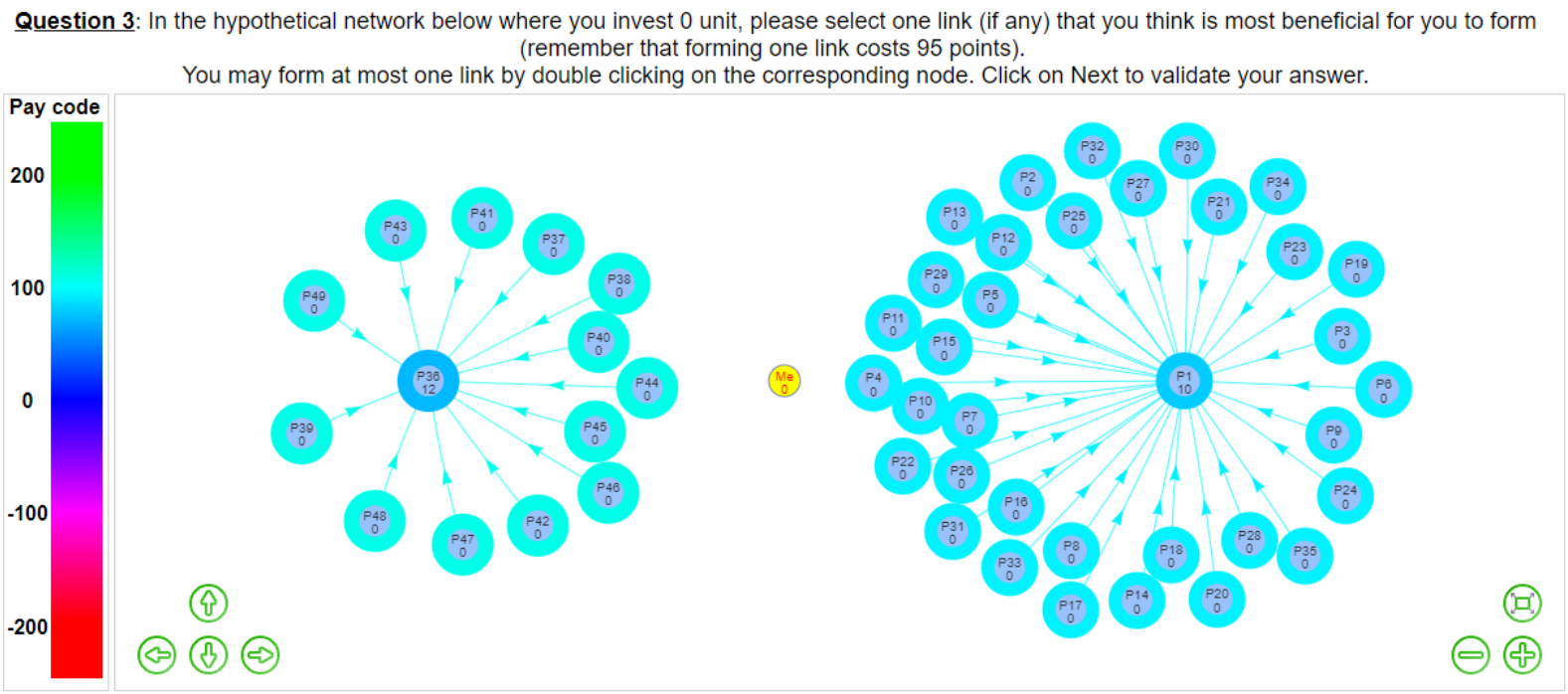
- 1 pts
- 11 pts
- 21 pts
- 31 pts
- 41 pts
- 51 pts
- more than 51 pts

Question 2: In the previous game, how many points did forming a link cost you?

- 0 pt
- 25 pts
- 45 pts
- 65 pts
- 95 pts
- 115 pts
- more than 115 pts

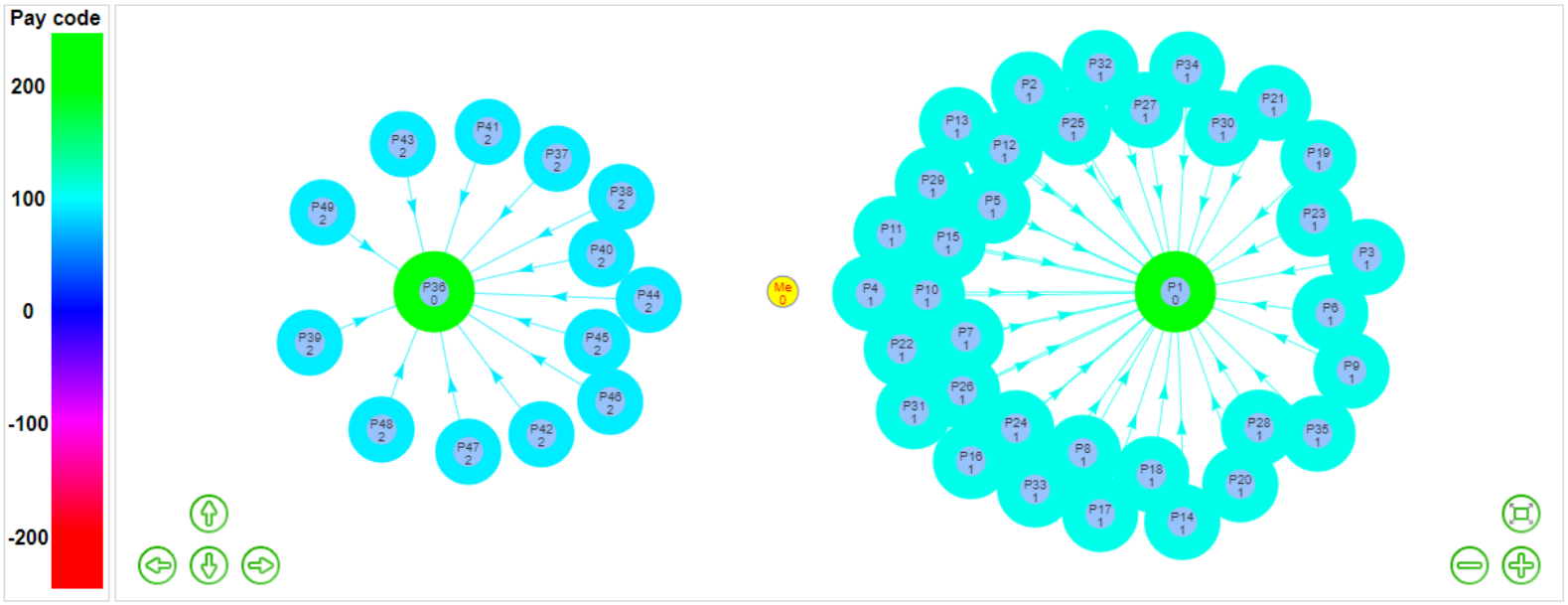
The third question depicted below was used in the payoff information treatment with

$n = 50$ (the correct answer is "P36"). This question was adapted in all other treatments by matching the number of nodes to the group size in the experiment, and by removing the colors in the baseline treatments.

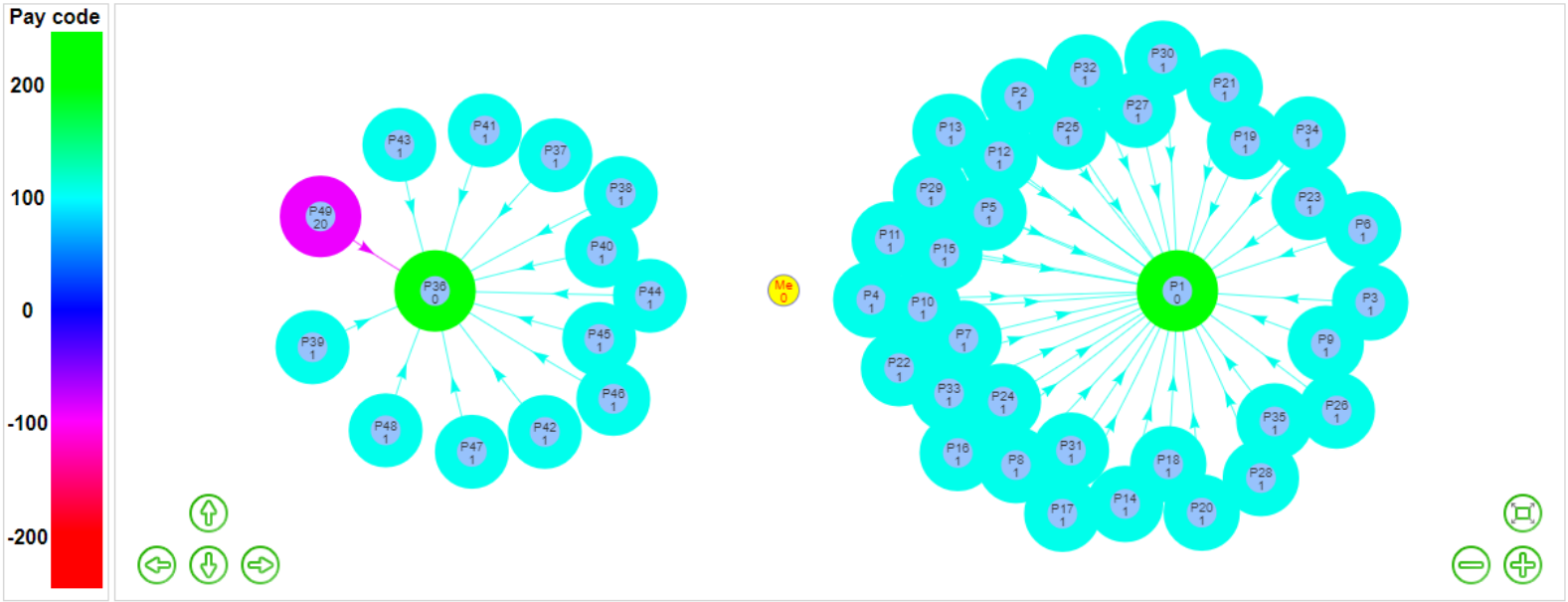


The following questions 4 and 5 below were also used in the payoff information treatment with $n = 50$ (correct answers are "P1" for both questions 4 and 5). These questions were however adapted only in other treatments where $n > 4$ by again matching the number of nodes to the group size in the experiment. The reason for filtering the small group treatments (with $n = 4$) is that the limited number of nodes did not allow representing the corresponding scenarios. As before, these questions were also adapted to the baseline treatments by simply removing the colors to match the design of the actual game that subjects played.

Question 4: In the hypothetical network below where you invest 0 unit, please select one link (if any) that you think is most beneficial for you to form (remember that forming one link costs 95 points).
You may form at most one link by double clicking on the corresponding node. Click on Next to validate your answer.

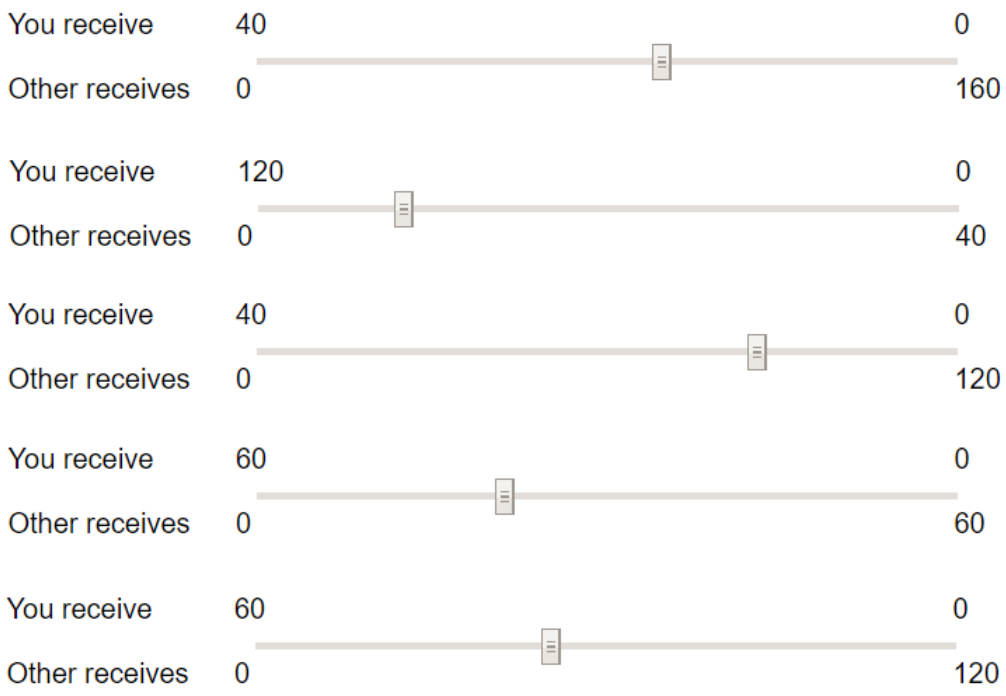


Question 5: In the hypothetical network below where you invest 0 unit, please select one link (if any) that you think is most beneficial for you to form (remember that forming one link costs 95 points).
You may form at most one link by double clicking on the corresponding node. Click on Next to validate your answer.



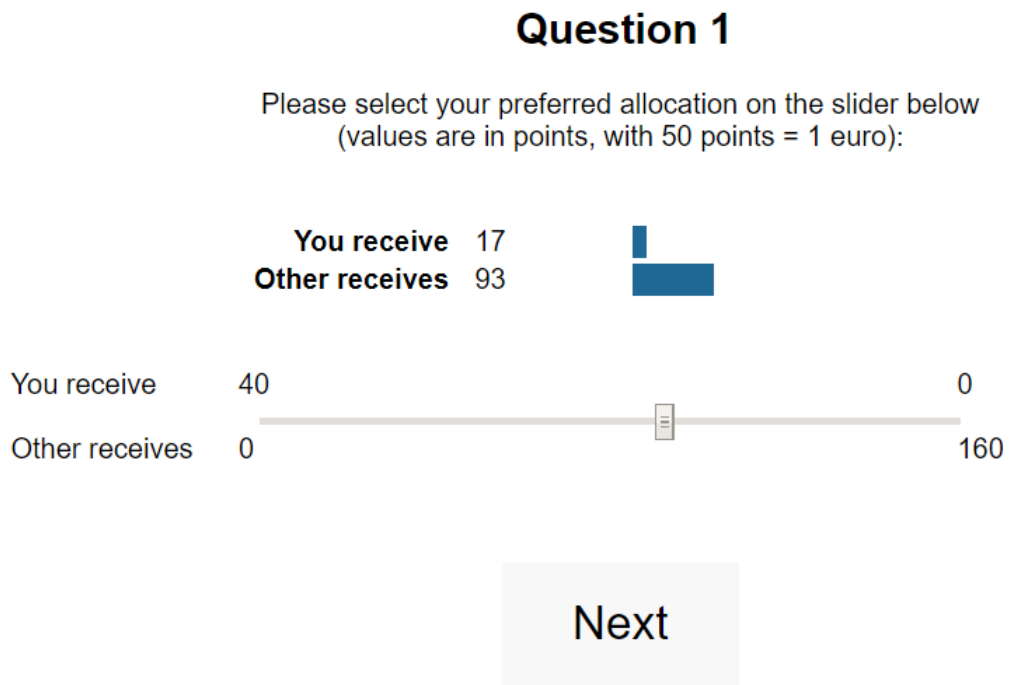
E.2 Social preferences

The social preferences measure was adapted from Andreoni and Miller (2002) and involved a series of five money allocation tasks between the decision maker and some anonymous external participants of another experiment at the LINEEX lab (corresponding payments were therefore made to these external passive participants). The five tasks used in our experiment were represented through sliders as shown in the following figure:



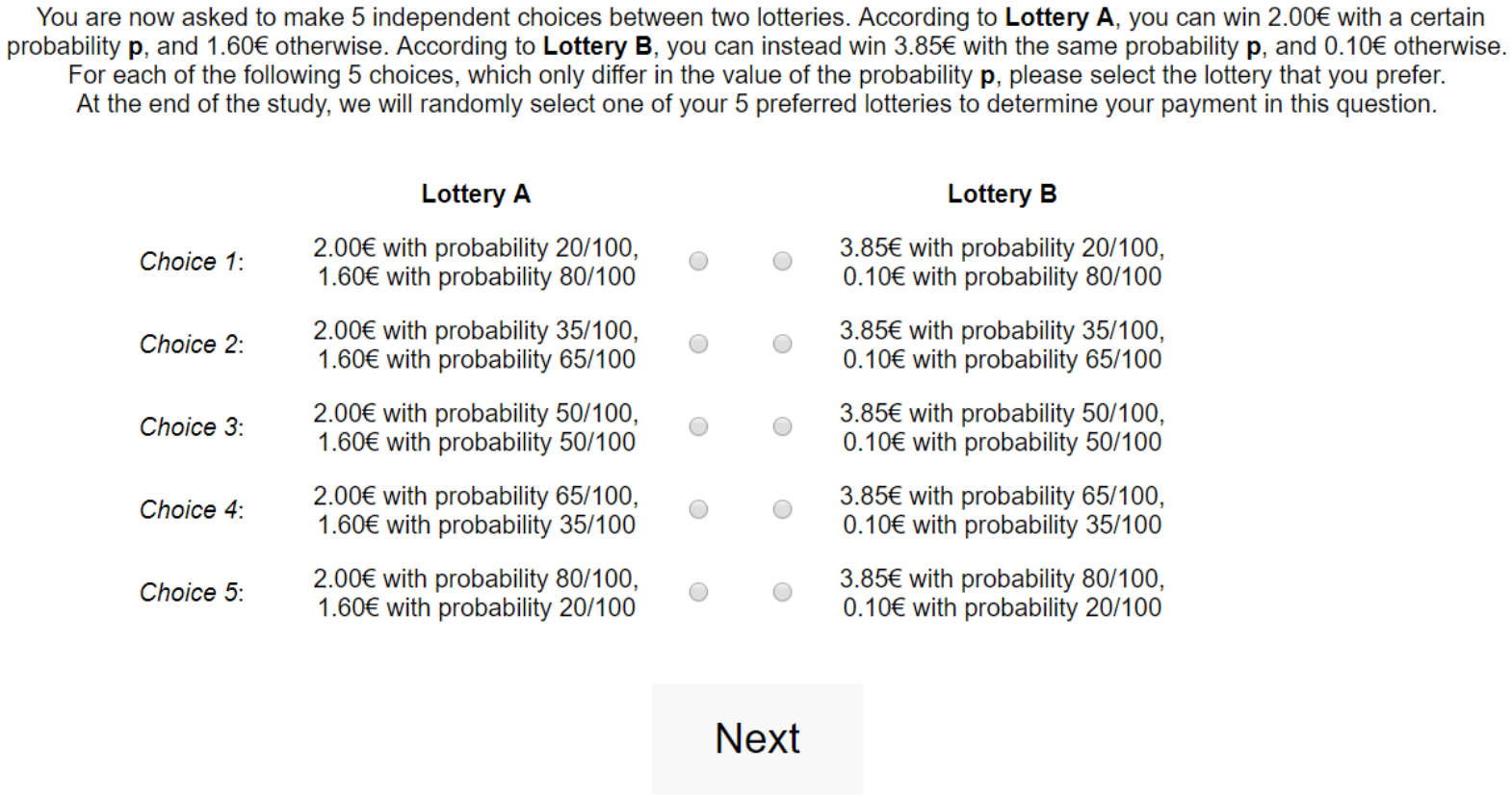
Note however that each question was presented in a different screen, and the order of presentation was randomized for every subject. Furthermore, 50 points were worth 1 euro both the subject, and the other anonymous external participant. Detailed instructions provided to the subjects, as well as a screenshot highlighting one of the above five questions are described below.

Instructions: You are asked to answer a series of 5 questions, each of which consists of selecting an allocation of points that you most prefer between yourself and an anonymous randomly selected person who is participating to a different experiment in this lab. At the end of the study, we will randomly select your allocation for 1 of the 5 questions to determine the payments for both you and the other person in this part. Your decisions will remain unknown to the other persons you are matched with.



E.3 Risk preferences

The risk preference measure was adapted from Holt and Laury (2002) and consisted of a series of five binary choices between lotteries, presented as in the figure below.



E.4 Personality test

Non incentivized measures were used through a simplified version of the Big Five personality inventory test adapted from Rammstedt and John (2007), as shown below.

How well do the following statements describe your personality?

I see myself as someone who...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1. ... is reserved	○	○	○	○	○
2. ... is generally trusting	○	○	○	○	○
3. ... tends to be lazy	○	○	○	○	○
4. ... is relaxed, handles stress well	○	○	○	○	○
5. ... has few artistic interests	○	○	○	○	○
6. ... is outgoing, sociable	○	○	○	○	○
7. ... tends to find fault with others	○	○	○	○	○
8. ... does a thorough job	○	○	○	○	○
9. ... gets nervous easily	○	○	○	○	○
10. ... has an active imagination	○	○	○	○	○

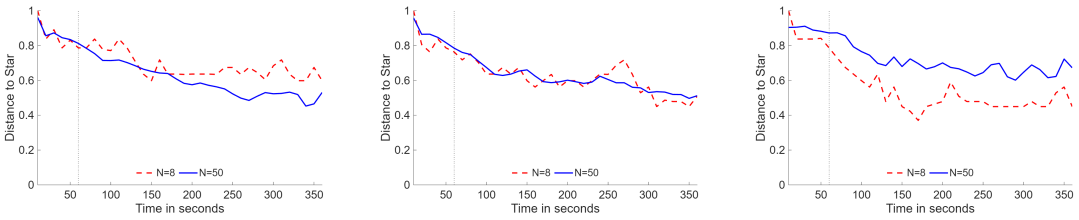
Next

F Additional Results

F.1 Main Treatments

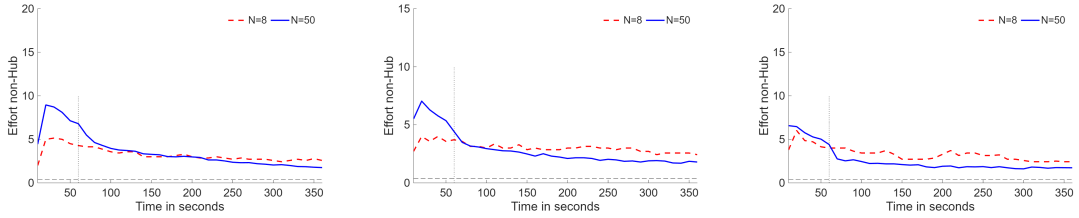
F.1.1 Dynamics of network statistics, efforts, and payoffs

We present time series of the main statistics used in the paper, over the 6 minutes of the game for each treatment (for any second, the median value is used across rounds and groups). More precisely, Figure A4 shows that the distance to the star network is constantly decreasing for all treatments. Figure A5 shows that the efforts of non-hubs, in



(a) Baseline (b) PayInfo (c) Counterfactual

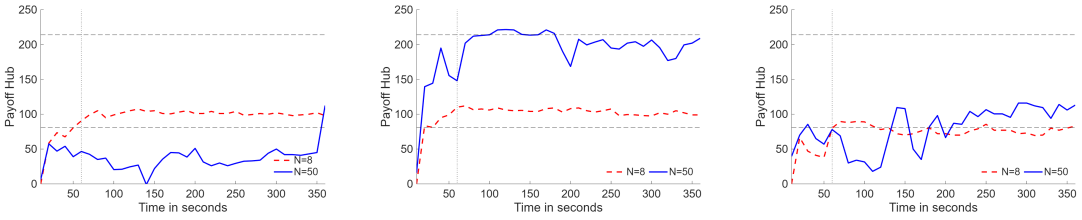
Figure A4: Dynamics of median distance to the star network



(a) Baseline (b) PayInfo (c) Counterfactual

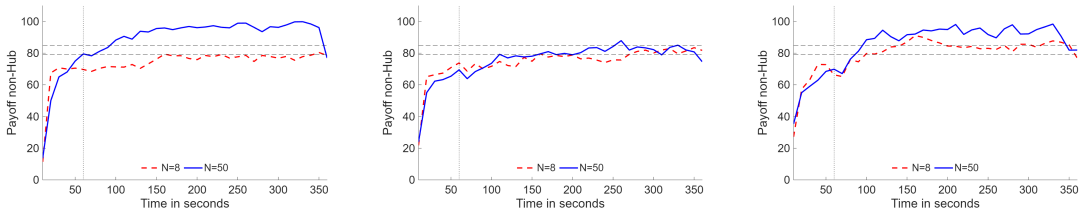
Figure A5: Dynamics of median effort for the non-hubs

aggregate, are mostly stable (only slightly decreasing). Figure A6 shows that payoff for the hub is constantly negative in Baseline (as a result of large effort levels), constantly positive in PayInfo (as a result of low effort level), and constantly growing in Counterfactual, fluctuating around the pure influencer equilibrium prediction. Finally, Figure A7 shows that payoffs of non-hubs are above theoretical expectations in Baseline and Counterfactual, and consistent with theoretical predictions in PayInfo.



(a) Baseline (b) PayInfo (c) Counterfactual

Figure A6: Dynamics of median payoff for the most connected subject (hub)

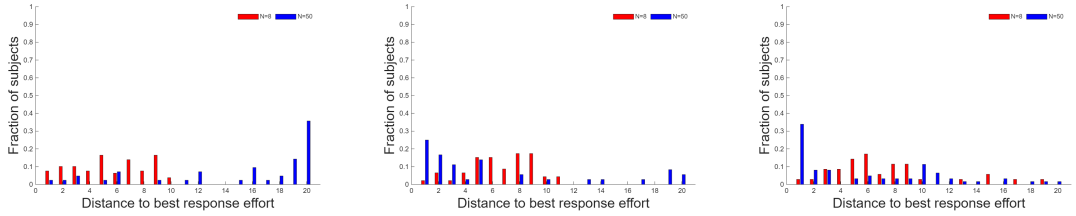


(a) Baseline (b) PayInfo (c) Counterfactual

Figure A7: Dynamics of median payoff for the non-hubs

F.1.2 Distance to best response

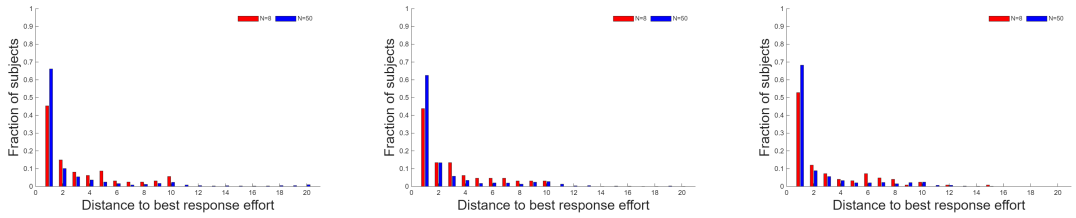
Let us consider hubs as any subjects who have become the most connected individual at any moment of any round during the experiment. Using this classification, we first observe that hubs represent a significant fraction of the population in small groups $N = 8$ (33% in Baseline, 19% in PayInfo, 23% in Counterfactual), but these numbers become extremely small in large groups $N = 50$ (2.8% in Baseline, 2.4% in PayInfo, 6.2% in Counterfactual). Figure A9 shows that, while there is heterogeneity in the errors made by subjects, a large fraction of non-hubs choose an effort consistent with the myopic best response value (within 1 unit): in Baseline, 45% for $N = 8$, and 66% for $N = 50$; in PayInfo, 44% for $N = 8$, and 63% for $N = 50$; in Counterfactual, 53% for $N = 8$, and 68% for $N = 50$.



(a) Baseline (b) PayInfo (c) Counterfactual

Figure A8: Average distribution of distance to best response effort for hubs

Notes: In these figures, each observation corresponds to a distribution computed at a moment between 40 seconds and 10 seconds before the end of a given round for a given group.



(a) Baseline (b) PayInfo (c) Counterfactual

Figure A9: Average distribution of distance to best response effort for non-hubs

Notes: In these figures, each observation corresponds to a distribution computed at a moment between 40 seconds and 10 seconds before the end of a given round for a given group.

F.1.3 Regression analyses

We here report regression analyses about the most connected individual (hub) in each treatment. Table A1 compares payoff for the hub with that of others (non-hubs). We find that in small groups, the hub earns significantly more in Baseline and PayInfo. However, no effect is found in Counterfactual. In the large groups, the hub earns significantly less in Baseline than non-hubs (more than 60 pts difference, as a result of choosing large effort levels by the hub) whereas they earn significantly more in PayInfo (more than 106 pts difference, as a result of low effort level by the hub). In the Counterfactual treatment, we find no significant difference in payoff between hub and non-hubs, which is in line with the theoretical predictions.

Table A4 compares the treatment effects on the time spent by subjects as the most connected individual, and the number of hubs in any given round. The following observations can be made: the hub is significantly more stable in the large group and similarly so for both Baseline and PayInfo treatments. However, there is significantly more turnover for the hub in the large group of Counterfactual than in the other treatments.

Table A1: Median regression analysis of payoff difference between hubs and non-hubs in all treatments (group level data between 40 seconds and 10 seconds before the end)

	Baseline		PayInfo		Counterfactual	
	N=8	N=50	N=8	N=50	N=8	N=50
Hub	22.22***	-61.89***	20.11***	106.24**	-11.15	24.35
	(5.60)	(10.73)	(6.63)	(48.29)	(16.26)	(15.23)
Non-hubs	78.53	96.01	79.28	81.23	87.11	83.03
# obs.	60	60	60	60	40	40

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Dependent variables are computed as mean values between 40 seconds and 10 seconds before the end of a round. There are 2 observations per round: the mean payoff of the hub, and the mean payoff among all others. All regressions include a constant, and dummy variables for the hub payoff (boolean) and rounds. Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds.

Table A2: Efforts for the most connected subject and others

Treatment	Type	N	mean	25th percentile	50th percentile	75th percentile
Baseline	Hub	8	8.08	6.67	8.68	9.87
		50	16.18	13.7	18.2	20
	Non-hub	8	2.73	2.03	2.63	3.6
		50	1.88	1.58	1.85	2.11
PayInfo	Hub	8	8.78	7	9	10
		50	5.5	0	2	7
	Non-hub	8	2.65	1.71	2.57	3.07
		50	1.77	1.41	1.73	2.19
Counterfactual	Hub	8	11.07	7.97	9.72	14.33
		50	7.49	5.25	8.5	10.22
	Non-hub	8	2.37	1.73	2.41	3.56
		50	1.62	1.01	1.8	2.15

Notes: A data point corresponds to the mean of all values between between 40 seconds and 10 seconds before the end of a given round (30 data points per treatment except for Counterfactual with 20 data points).

Table A3: Payoffs for the most connected subject and others

Treatment	Type	N	mean	25th percentile	50th percentile	75th percentile
Baseline	Hub	8	91.66	88.2	99.48	107.07
		50	2.31	11.97	47.57	71.27
	Non-hub	8	77.48	63.68	78.53	89.11
		50	95.72	85.86	96.01	110.57
PayInfo	Hub	8	98.35	81	99	113
		50	39.67	40	200.75	219.5
	Non-hub	8	81.08	74.57	80.68	89.21
		50	82.42	65.62	85.57	96.04
Counterfactual	Hub	8	75.02	50.32	77.33	100.53
		50	-356.06	50.75	105.6	140.03
	Non-hub	8	86.66	72.82	87.11	95.82
		50	90.38	69.45	83.03	101.09

Notes: A data point corresponds to the mean of all values between between 40 seconds and 10 seconds before the end of a given round (30 data points per treatment except for Counterfactual with 20 data points). In the Counterfactual treatment, the presence of outliers justifies the negative mean while all quartiles are positive.

Table A4: Regression analysis of the hub dynamics

	Max time as hub	Mean time as hub	Nb of hubs
	OLS	OLS	OLS
N=50	72.83**	66.27***	-2.43***
	(30.73)	(22.28)	(0.73)
PayInfo	25.17	9.42	-0.57
	(35.50)	(15.53)	(0.81)
Counterfactual	42.12	18.13	-1.13
	(38.44)	(17.28)	(1.10)
PayInfo × ($N = 50$)	-13.33	-3.37	0.17
	(41.59)	(33.29)	(1.05)
Counterfactual × ($N = 50$)	-95.28**	-81.74***	5.98**
	(44.77)	(27.79)	(2.45)
$N = 8$	123.33	28.37	6.03
# obs.	160	160	160

Notes: The reference group size is $N = 8$ baseline. OLS regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

F.1.4 Non-parametric tests

In this section, we report non parametric tests to support the regression analyses presented separately. Those tests are performed at the group level, which means they consider 6 observations per group size in both Baseline and PayInfo treatments, and 4 observations per group size in the Counterfactual treatment. Table A5 shows that in the Baseline treatment, there is a significant group size effect on the effort and payoff for the hub, and on the non-hubs' efforts. In the PayInfo treatment, there is only a significant group size effect on the hub's effort. Regarding the Counterfactual treatment, we find no group size effect, which is consistent with observing influencer outcomes in both $N = 8$ and $N = 50$.

Table A5: Mann-Whitney U tests of group size effects on different metrics for each treatment (group level data between 40 seconds and 10 seconds before the end)

Treatment	Metric	# obs.	U statistic	z statistic	p-value
Baseline	Distance to star	12	13.000	-0.801	0.4233
	Effort hub	12	0.000	2.903	0.0037
	Effort non-hub	12	2.000	-2.562	0.0104
	Payoff hub	12	0.000	-2.882	0.0039
	Payoff non-hub	12	6.000	1.922	0.0547
PayInfo	Distance to star	12	13.000	0.801	0.4233
	Effort hub	12	1.000	-2.722	0.0065
	Effort non-hub	12	6.000	-1.922	0.0547
	Payoff hub	12	12.000	0.961	0.3367
	Payoff non-hub	12	17.000	0.160	0.8728
Counterfactual	Distance to star	8	5.000	0.866	0.3865
	Effort hub	8	4.000	-1.155	0.2482
	Effort non-hub	8	7.000	-0.289	0.7728
	Payoff hub	8	4.000	1.155	0.2482
	Payoff non-hub	8	8.000	0.000	1.0000

Notes: each data point summarizes one group: within the group, we averaged the outcome for each round over the interval 40–10 s before round end, then took the median across rounds.

Tables A6 and A7 compare information treatments for any group size. More specifically,

Table A6 shows there is no effect of showing others' payoff information for the small groups, but it significantly affects the hub's effort in the large group, consistently with our main results. Similarly, Table A7 reveals the same patter regarding the effect of showing counterfactual payoff information in small and large groups.

Table A6: Mann-Whitney U tests for treatment effects (Baseline Vs. PayInfo) on all metrics (group level data between 40 seconds and 10 seconds before the end)

N	metrics	# obs.	U statistic	z statistic	p-value
8	Distance to star	12	12.000	-0.961	0.3367
	Effort hub	12	15.500	-0.401	0.6884
	Effort non-hub	12	14.000	-0.641	0.5218
	Payoff hub	12	16.000	0.320	0.7488
	Payoff non-hub	12	17.000	0.160	0.8728
50	Distance to star	12	18.000	0.000	1.0000
	Effort hub	12	0.000	-2.903	0.0037
	Effort non-hub	12	12.000	-0.961	0.3367
	Payoff hub	12	12.000	0.961	0.3367
	Payoff non-hub	12	6.000	-1.922	0.0547

Notes: each data point summarizes one group: within the group, we averaged the outcome for each round over the interval 40–10 s before round end, then took the median across rounds.There are 6 observations from each treatment.

Table A7: Mann-Whitney U tests for treatment effects (Baseline Vs. Counterfactual) on all metrics (group level data between 40 seconds and 10 seconds before the end)

N	metrics	# obs.	U statistic	z statistic	p-value
8	Distance to star	10	16.000	-0.853	0.3938
	Effort hub	10	6.000	1.279	0.2008
	Effort non-hub	10	14.000	-0.426	0.6698
	Payoff hub	10	15.000	-1.919	0.0550
	Payoff non-hub	10	6.000	1.279	0.2008
50	Distance to star	10	9.000	0.640	0.5224
	Effort hub	10	12.000	-2.590	0.0096
	Effort non-hub	10	13.000	-0.213	0.8312
	Payoff hub	10	6.000	1.279	0.2008
	Payoff non-hub	10	15.000	-0.640	0.5224

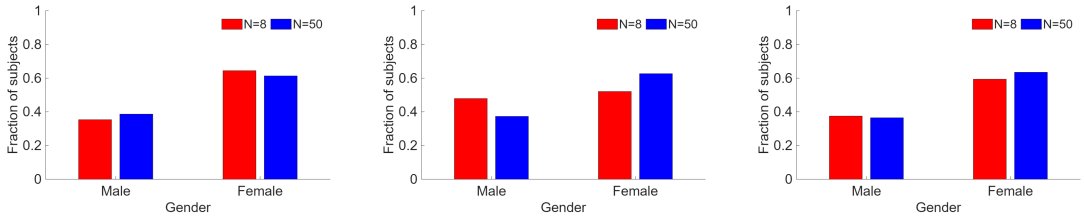
Notes: each data point summarizes one group: within the group, we averaged the outcome for each round over the interval 40–10 s before round end, then took the median across rounds. There are 6 observations from the Baseline treatment and 4 observations from the Counterfactual treatment.

F.1.5 Individual characteristics

F.1.5.1 Descriptive statistics

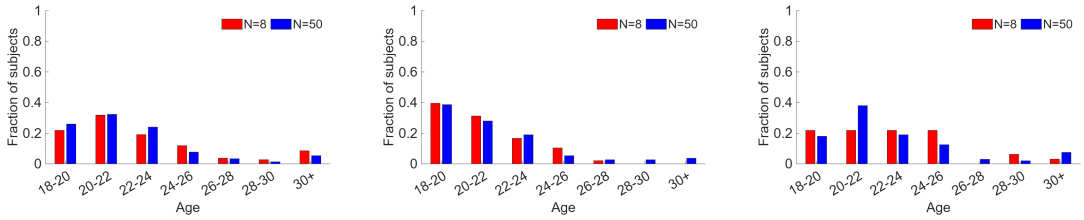
We here report statistics about individual characteristics, as measured through the post game surveys in the experiment (see Section E for more details about the questionnaires). The risk aversion index, whose distribution is presented in Figure A12, corresponds to the fraction of times the subject chose the least risky option, out of the 5 choices. The social preference index, whose distribution is presented in Figure A13, is computed as the average ratio of payoff allocated to other divided by the sum of allocation to self and other across the 5 allocation decision made by the subject. The comprehension index, whose distribution is presented in Figure A14, corresponds to the fraction of accurate answers across the 5 comprehension questions. Overall, the distributions of individual

characteristics are consistent across information treatments.



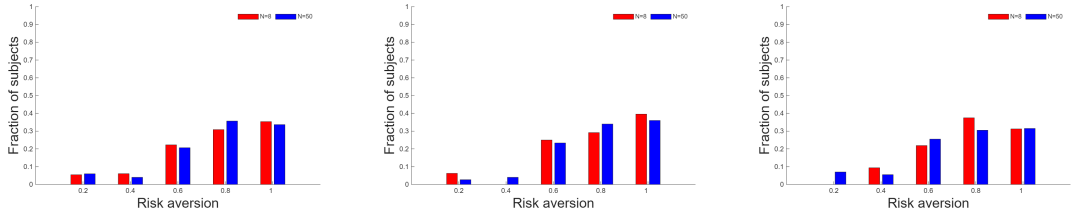
(a) Baseline (b) PayInfo (c) Counterfactual

Figure A10: Gender distribution



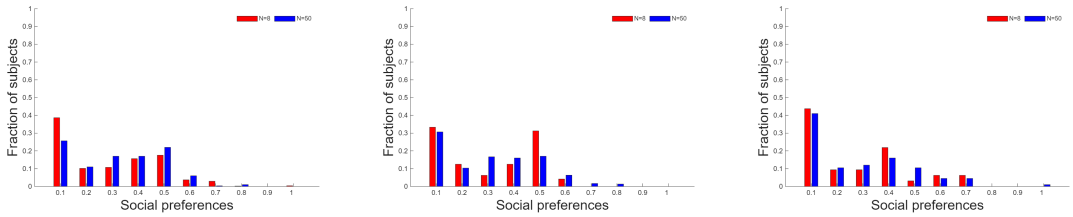
(a) Baseline (b) PayInfo (c) Counterfactual

Figure A11: Age distribution



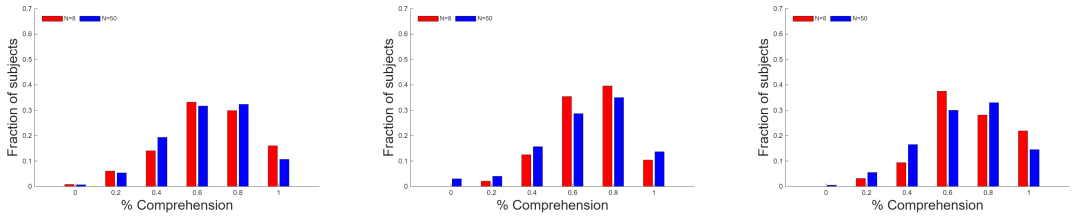
(a) Baseline (b) PayInfo (c) Counterfactual

Figure A12: Risk aversion



(a) Baseline (b) PayInfo (c) Counterfactual

Figure A13: Social preferences



(a) Baseline (b) PayInfo (c) Counterfactual

Figure A14: Comprehension

F.1.5.2 Who emerges as a hub?

In the Baseline large group, hubs faced excessive competition and as a result, made overinvestment and earned low payoff. In contrast, in the PayInfo large group, the hub, once it emerges, became dominant, shaded their efforts sufficiently low, and eventually earned high payoff. A natural question is then who emerges as a hub and whether the likelihood of being a hub is related to particular individual characteristics. Using individual characteristics measured through post game questionnaires (i.e., demographic information, attitude towards risk, social preferences, personality traits, comprehension of the task), we check whether any of these are associated with the frequency of being a hub during the course of the experiment.

We compute the total time spent as a hub across the last five minutes of the 5 rounds for

each individual. Table A8 presents the Tobit regression results of this total time as a hub on individual characteristics. First, in all large group treatments, female subjects are less likely to become a hub. It suggests that female subjects tend to avoid joining competition for the hub, which is in line with the literature of gender differences in competitive environments (Niederle and Vesterlund (2007, 2011)). The same pattern of gender difference is also found in the Baseline small group but not in the PayInfo and Counterfactual small groups. Second, in the Baseline large group treatment, more prosocial individuals tend to spend less time as the hub. This rules out any prosocial motivation to choose extreme effort levels. On the other hand, the social preference measure is not significantly associated with the frequency of being a hub in the other treatments. Similarly, higher comprehension test score lead to longer time spent as the hub only in the large group Counterfactual. Finally, risk aversion is not significantly associated with the frequency of a hub in any treatment.

Table A8: Tobit regressions of the total time as a hub on individual characteristics

	Baseline		Payinfo		Counterfactual	
	$N = 8$	$N = 50$	$N = 8$	$N = 50$	$N = 8$	$N = 50$
Age	-0.17	2.54	-4.98	6.80***	-0.04	2.74
	(1.48)	(2.47)	(4.27)	(1.81)	(5.34)	(1.89)
Female	-67.95***	-87.29***	3.89	-91.32***	-65.18**	-51.92***
	(17.26)	(29.81)	(21.32)	(26.95)	(27.04)	(13.17)
Risk aversion	-2.50	7.35	0.03	-4.46	-33.46**	7.60*
	(7.83)	(11.97)	(10.40)	(10.11)	(13.90)	(4.46)
Social preference	-30.94	-163.35**	51.47	30.56	209.94**	-15.32
	(38.01)	(63.74)	(59.65)	(61.11)	(98.08)	(27.26)
Comprehension	-6.69	-2.98	17.29	17.36	-5.85	11.79**
	(5.85)	(15.58)	(13.94)	(11.48)	(13.62)	(4.97)
# obs.	240	1500	240	1500	160	1000

Clustered standard errors at the individual subject level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Dependent variables are the total seconds of being most connected in last 5 minutes, for any subject in a given round. All regressions include a constant, Big 5 personality traits, education levels, ethinic background, religion, and round and group dummies. Regressions are computed with standard errors clustered at the individual subject level, allowing within-individual correlation across rounds.

F.1.5.3 Non-hubs behavior

Let us consider non-hubs as any subjects who have never become the most connected individual at any moment of any round during the experiment. For each of these subjects, we quantify the errors that they make by measuring how distant their behavior is from myopic best responses. Since our previous results revealed important deviations in effort from equilibrium predictions, we focus on the distance to best response effort. Table A9 presents regression analyses to identify potential effects of the different metrics measured on the distance to best response effort. The results reveal a significant effect of comprehension in large groups for both Baseline and PayInfo treatments: individuals with a better comprehension of the game tend to choose effort levels closer to the myopic best response. This effect is significant at the 5% level. However, comprehension has no effect in the Counterfactual treatment, as expected (the experimental manipulation specifically helped subjects navigate the complexity of the environment). This finding suggests that the complexity of the decision environment in the large group setting can explain why a minority of subjects deviates from best response in Baseline and PayInfo.

Table A9: Tobit regressions of the distance to best response effort for the non-hubs on individual characteristics

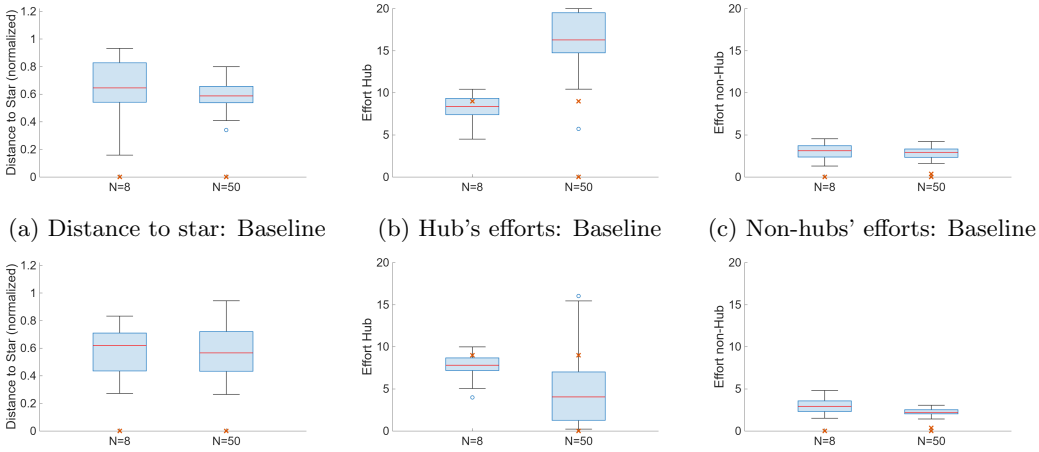
	Baseline		Payinfo		Counterfactual	
	$N = 8$	$N = 50$	$N = 8$	$N = 50$	$N = 8$	$N = 50$
Age	-0.07	0.07	-0.27	0.06	0.41***	-0.02
	(0.08)	(0.07)	(0.17)	(0.07)	(0.15)	(0.04)
Female	-0.23	-0.80	0.43	0.33	0.82	-1.09**
	(0.91)	(0.55)	(0.77)	(0.40)	(0.89)	(0.43)
Risk aversion	-0.27	-0.16	0.17	0.25	0.20	0.08
	(0.47)	(0.19)	(0.29)	(0.16)	(0.36)	(0.18)
Social preference	2.51	1.81	0.86	1.64	1.99	1.11
	(2.07)	(1.43)	(2.06)	(1.09)	(3.58)	(1.08)
Comprehension	-0.03	-0.62**	0.23	-0.55***	0.36	0.02
	(0.34)	(0.30)	(0.35)	(0.18)	(0.54)	(0.19)
# obs.	161	1458	194	1464	120	938

Clustered standard errors at the individual subject level in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: Dependent variables are the average distances to best response effort for any non-hub subject between 40 seconds and 10 seconds before the end of a round. All regressions include a constant, Big 5 personality traits, education levels, ethnic background, religion, and round and group dummies. Regressions are computed with standard errors clustered at the individual subject level, allowing within-individual correlation across rounds.

F.1.6 Robustness check

We replicate all data analyses from the main text while aggregating various statistics across the entire payoff effective period of each round (instead of focusing on values between 40 seconds and 10 seconds before the end of a round). More precisely, we now consider a data point as the mean of all values in the last 300 seconds (5 minutes) of a given round.



(a) Distance to star: Baseline (b) Hub's efforts: Baseline (c) Non-hubs' efforts: Baseline

(d) Distance to star: PayInfo (e) Hub's efforts: PayInfo (f) Non-hubs' efforts: PayInfo

Figure A15: Networks and efforts across treatments

Notes: A data point corresponds to the mean of all values in the last 300 seconds (5 minutes) of a given round (30 data points per treatment). A red cross represents a theoretical benchmark. The red line in the box plot represents a median, the top and bottom edges of the box are the upper and lower quartiles, respectively, outliers (represented by circles) correspond to values more than 1.5 times the interquartile range away from the top or bottom of the box, and the whiskers above and below the box represent the non-outlier maximum and minimum, respectively. Mean normalized distance to star: 0.65 for N=8, 0.59 for N=50 in Baseline, and 0.58 for N=8, 0.58 for N=50 in PayInfo. Mean effort of the hub: 8.18 for N=8, 16.16 for N=50 in Baseline, and 7.79 for N=8, 5.52 for N=50 in PayInfo; mean effort of the non-hubs: 3.06 for N=8, 2.85 for N=50 in Baseline, and 2.97 for N=8, 2.28 for N=50 in PayInfo.

Table A10: Treatment effects of group size and payoff information

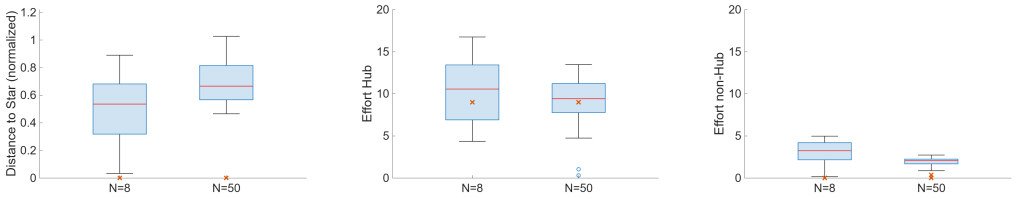
	Distance to star	Effort hub		Effort Non-hub	
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	Median	OLS	Median
$N = 50$	-0.06	7.98***	7.97***	-0.21	-0.21
	(0.08)	(1.14)	(1.45)	(0.32)	(0.51)
PayInfo	-0.07	-0.39	-0.47	-0.09	-0.16
	(0.09)	(0.55)	(0.78)	(0.34)	(0.57)
PayInfo $\times$ $(N = 50)$	0.07	-10.25***	-11.39***	-0.48	-0.49
	(0.11)	(1.63)	(2.13)	(0.42)	(0.63)
$N = 8$	0.65	8.18	8.37	3.06	3.12
# obs.	120	120	120	120	120

Notes: In each treatment, a data point corresponds to the mean of all values in the last 300 seconds (5 minutes) of a given round (30 data points per treatment). The reference group size is $N = 8$ baseline (mean for OLS and Median for median regression). Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Treatment effects of group size and payoff information on payoffs

	Hub's payoffs		Non-hubs' payoffs	
	OLS	Median	OLS	Median
$N = 50$	-298.07	-61.20*	24.31**	24.31**
	(193.09)	(34.49)	(8.82)	(8.82)
PayInfo	11.52	10.18	0.01	0.01
	(8.39)	(10.97)	(4.38)	(4.38)
PayInfo $\times$ $(N = 50)$	278.67	111.29**	-19.63*	-19.63*
	(201.78)	(48.55)	(9.77)	(9.77)
$N = 8$	93.28	94.07	73.40	75.89
# obs.	120	120	120	120

Notes: In each treatment, a data point corresponds to the mean of all values in the last 300 seconds (5 minutes) of a given round (30 data points per treatment). The reference group size is $N = 8$ baseline (mean for OLS and Median for median regression). Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.



(a) Distance to star (b) Effort of the hub (c) Effort of non-hubs

Figure A16: Networks and Efforts in the counterfactual treatment

Notes: Box plots follow the same description as Figure A15. Mean normalized distance to star: 0.51 for N=8, 0.70 for N=50. Mean effort of the hub: 10.38 for N=8, 8.79 for N=50; mean effort of the non-hubs: 3.04 for N=8, 1.96 for N=50.

Table A12: Treatment effects of group size and counterfactual payoff information on network and efforts

	Distance to star	Effort hub		Effort Non-hub	
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	Median	OLS	Median
$N = 50$	-0.06	7.98***	7.97***	-0.21	-0.21
	(0.08)	(1.14)	(1.61)	(0.32)	(0.51)
Counterfactual	-0.14	2.20	2.10	-0.02	-0.17
	(0.13)	(1.57)	(2.02)	(0.71)	(1.17)
Counterfactual $\times$ $(N = 50)$	0.25	-9.57***	-9.02***	-0.87	-0.65
	(0.16)	(2.17)	(2.65)	(0.77)	(1.23)
$N = 8$	0.65	8.18	8.37	3.06	3.12
# obs.	100	100	100	100	100

Notes: In each treatment, a data point corresponds to the mean of all values in the last 300 seconds (5 minutes) of a given round (30 data points per treatment). The reference group size is $N = 8$ baseline (mean for OLS and Median for median regression). Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Distribution of influencer and connector outcomes for $N = 50$

Treatments	# obs.	% influencer	% connector
Baseline	30	10	0
PayInfo	30	16.7	36.7
Counterfactual	20	75	10

Notes: A round is classified as "influencer" if the average effort of the most connected individual in the last 300 seconds (5 minutes) of the round is within the range $[7, 12]$. Similarly, a round is classified as "connector" if the average effort by the main hub in that period is lower than 3. Those cutoff effort values allow for the theoretical predictions with some noise.

F.2 Extra treatments

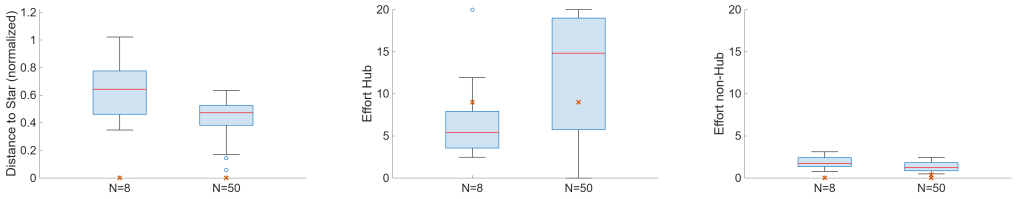
We here provide more details regarding the extra treatments used to check the robustness of our main results.

F.2.1 Lower linking cost

Figure A17 provides descriptive information about the distance of the observed networks in the experiment to the standardized star network, and the distribution of effort by the hub and non-hubs across both group sizes (similarly to Figures 3 and 5).

F.2.2 Larger group size $N = 100$

Table A14 describes the regression analysis related to the distance to the star network, and the effort of both the hub and non-hubs, comparing the Baseline and PayInfo treatment in the large scale setting ($N = 100$) with $N = 50$. Regarding efforts, we find no significant difference between $N = 100$ and $N = 50$. Regarding the distance to the star network, we find no effect in the Baseline treatment, but a slightly larger distance in PayInfo for the $N = 100$.



(a) Distance to star (b) Effort of the hub (c) Effort of non-hubs

Figure A17: Networks and Efforts in the low cost treatments

Notes: Same description as Figure 3 (20 data points per treatment). Mean normalized distance to star: 0.39 for N=4, 0.61 for N=8, 0.49 for N=50, 0.43 for N=100. Mean effort of the hub: 10.35 for N=4, 8.3 for N=8, 16.83 for N=50, 14.33 for N=100; mean effort of the non-hubs: 2.78 for N=4, 2.66 for N=8, 1.88 for N=50, 1.91 for N=100.

Table A14: Treatment effects of $N = 100$ and payoff information on network and efforts

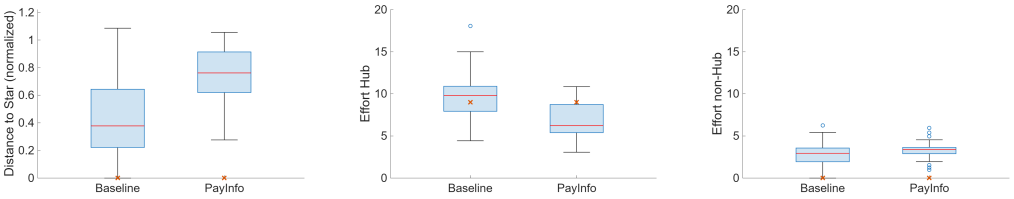
	Distance to star	Effort hub		Effort Non-hub
	(1)	(2)	(3)	(4)
	OLS	OLS	Median	OLS
$N = 100$	-0.07	-1.61	1.47	0.06
	(0.06)	(1.58)	(1.31)	(0.15)
PayInfo	0.00	-10.42***	-15.97***	-0.11
	(0.10)	(1.92)	(1.07)	(0.18)
PayInfo $\times$ ($N = 100$)	0.26**	1.50	-1.23	-0.16
	(0.10)	(3.28)	(3.18)	(0.21)
$N = 50$	0.5	16.18	18.2	1.88
# obs.	90	90	90	90

Notes: The reference group size is $N = 50$ baseline (mean for OLS and Median for median regression). Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A15: Distribution of influencer and connector outcomes for $N = 100$

Treatments	# obs.	% influencer	% connector
Baseline	15	0	0
PayInfo	15	20	40

Notes: A round is classified as "influencer" if the average effort of the most connected individual between 40 seconds and 10 seconds before the end of the round is within the range $[7, 12]$. Similarly, a round is classified as "connector" if the average effort by the main hub in that period is lower than 3. Those cutoff effort values allow for the theoretical predictions with some noise.



(a) Distance to star (b) Effort of the hub (c) Effort of non-hubs

Figure A18: Networks and Efforts for $N = 4$

Notes: Same description as Figure 3. Mean normalized distance to star: 0.42 for Baseline, 0.76 for PayInfo. Mean effort of the hub: 9.92 for Baseline, 6.92 for PayInfo; mean effort of the non-hubs: 2.87 for Baseline, 3.28 for PayInfo.

F.2.3 Smaller group size $N = 4$

Similarly to the previous treatments, we also tested the robustness of the small group results. For this purpose, we replicated the Baseline and PayInfo conditions in an even smaller scale setup, with groups of 4 subjects (6 groups per treatment with 48 participants overall). The corresponding experimental designs were identical to the Baseline and PayInfo treatments. For this N=4 treatment Baseline (PayInfo), 1 session (with 4 groups) was conducted at LINEEX at the University of Valencia in 2017, and 1 session (with 2 groups) was conducted at LEE at the University Jaume I of Castellón in 2025 (so 4 sessions overall).

Our results, summarized in Figure A18 and table A16, reveal similar patterns with $N = 8$ in the Baseline treatment. However, in the presence of others' payoff information, we find small but significant differences: the observed network are further away from the star network than for $N = 8$, and the effort of the most connected individual is significantly smaller than for $N = 8$ (about 3 units less).

Table A16: Treatment effects of $N = 4$ and payoff information on network and efforts

	Distance to star	Effort hub		Effort Non-hub
	(1)	(2)	(3)	(4)
	OLS	OLS	Median	OLS
$N = 4$	-0.19	1.85*	1.40	0.14
	(0.11)	(0.95)	(0.99)	(0.54)
PayInfo	-0.13	0.16	-0.00	-0.05
	(0.11)	(0.63)	(0.80)	(0.43)
PayInfo $\times$ ($N = 4$)	0.47***	-3.17**	-3.50**	0.46
	(0.13)	(1.16)	(1.43)	(0.69)
$N = 8$	0.61	8.08	8.68	2.73
# obs.	120	120	120	120

Notes: The reference group size is $N = 8$ baseline (mean for OLS and Median for median regression). Regressions are computed with robust standard errors clustered at the group level, allowing within-group correlation across rounds. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

J. Andreoni and J. Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

J. Barnes and P. Hut. A hierarchical o(n log n) force-calculation algorithm. *Nature*, 324: 446–449, 1986.

M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

P. Eades. A heuristic for graph drawing. *Congressus numerantium*, 42:149–160, 1984.

T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

E. Gallo and C. Yan. The effects of reputational and social knowledge on cooperation. *Proceedings of the National Academy of Sciences*, 112(12):3647–3652, 2015.

C. A. Holt and S. K. Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

Y. Hu. Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1): 37–71, 2005.

M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6), 2014.

M. Niederle and L. Vesterlund. Do women shy away from competition? do men compete too much? *The quarterly journal of economics*, 122(3):1067–1101, 2007.

M. Niederle and L. Vesterlund. Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630, 2011.

B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007.