WORKING PAPERS

Stability in Network Formation Games with Streams of Payoffs: An Experimental Study

Mariya Teteryatnikova
James Tremewan

July 2015

Working Paper No: 1508

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Stability in Network Formation Games with Streams of Payoffs: An Experimental Study*

Mariya Teteryatnikova[†] James Tremewan[‡]

July 29, 2015

Abstract

We run a novel network formation experiment with a stream of payoffs and relatively unstructured link formation process, and test the performance of a number of theoretical stability concepts in this environment. We focus especially on the issue of myopic versus farsighted behaviour in network formation. A subtle treatment variation demonstrates clearly the power of myopic stability concepts in identifying the most stable networks. However, we also find support for farsighted concepts of stability, especially those that assume players are pessimistic about the eventual outcome of a deviation.

KEY WORDS: network formation; myopic and farsighted stability; cautious behaviour; laboratory experiment

JEL CLASSIFICATION: A14, C71, C92, D85

*We acknowledge financial support from the Vienna Science and Technology Fund (WWTF) under project fund MA 09-017.

[†]University of Vienna. Email: mariya.teteryatnikova@univie.ac.at.

[‡]University of Vienna, Vienna Center for Experimental Economics. Email: james.tremewan@univie.ac.at.

1 Introduction

Network interactions involving a regular flow of payoffs feature in many social and economic environments. For example, the yearly investment opportunities of business-people, output of journalists, and the number of papers published by academic researchers all depend on their networks of associates, contacts, and co-authors. The importance attached to such networks is shown by the popularity of websites such as LinkedIn, Xing, and ResearchGate. The prevalence of situations in which people are concerned with benefits and losses that they derive from their ongoing relationships makes it desirable to understand which network structures in these environments are stable and thus likely to be observed.

Many theories have been proposed to predict the stable outcomes of network formation games. As the relationships we are modelling are those where "it takes two to tango", we restrict attention to theories that assume bilateral, or *pairwise* link formation, where links require the consent of both parties to form, and can be broken unilaterally. Within this approach we focus on the large and prominent group of *cooperative* stability concepts, which analyze network stability by describing the conditions that stable networks should satisfy and do not rely on a particular network formation protocol.[1] As we describe in more detail later, these stability concepts can be broadly divided into three classes, assuming myopic, farsighted, or cautious farsighted behaviour.[2] The purpose of this study is to identify which of these classes of concepts best describes the outcomes of network formation in an environment with ongoing payoffs.

We test the theoretical predictions of pairwise myopic, farsighted and cautious farsighted stability concepts in a laboratory experiment, implementing two different three-player network formation games. The experiment is novel in several important ways. First, reflecting the real-world applications mentioned above, subjects in our experiment are paid for all intermediate steps of network formation rather than only for the final network. Secondly, our subjects interact in an essentially unstructured manner. In particular, they are able to suggest and agree to links at any point in time. This is also a closer representation of the situations we are interested in, which do not involve strict timing or ordering of actions.[3] Finally, our experiment is the first we are aware of to enable the clear identification of farsighted

[1]Theories that make alternative assumptions to pairwise and/or cooperative behaviour are briefly discussed in footnote 9.

[2]An extended discussion of these theoretical stability concepts and their predictions for our experimental games are provided in Section 2.

[3]Berninghaus et al. (2006) also implement an experiment with free-timing of moves and a flow of payoffs, but with unilateral link formation. Callander and Plott (2005) have treatments with free timing of moves, but again in a unilateral link formation game. Their motivation was to overcome coordination problems that arise from the non-cooperative game

behaviour by having networks which are predicted to be farsightedly but not myopically stable.

Overall, our experimental results suggest that if one is concerned with only the most strongly stable networks, myopic concepts perform best, however, if one wishes to identify all networks which have the potential to be stable, then cautious farsighted concepts should be considered. We now describe these findings in more detail.

The most strongly stable networks are those predicted by myopic stability concepts, that is, concepts in which players are assumed to consider only the immediate payoff consequences of adding or deleting a link they may be involved in and do not take into account possible chains of other players' reactions.[4] Networks that are identified as myopically stable exhibit strong empirical stability according to two measures: they often lasted the full duration of a game, and once they were in place, on average they remained in place for at least one more time period with very high probability.

We also find clear support for stability concepts that assume farsighted behaviour. Farsightedness means that players take into account the chains of reactions that might follow after their own initial deviation.[5] Although the networks identified by farsighted concepts do exhibit as high a degree of stability as those identified by myopic concepts, they are the only other networks that were regularly observed to last entire games, and once in place, to be significantly more likely to remain than not.

Furthermore, considering which concepts most precisely identify the set of stable networks, we find that this depends on the strictness of one's empirical definition of stability. We define the strictness of a stability definition according to the minimum probability with which a type of network must remain in place in order to be classed as stable. We find a range of probabilities where myopic concepts precisely identify the set of stable networks, however, for a range of lower probabilities, it is two of the cautious farsighted concepts – the set of farsightedly consistent networks (Page Jr et al., 2005) and the cautious path stable set (Teteryatnikova, 2015) – that predict exactly the empirically stable networks. Cautious farsighted concepts assume that players will not add or delete a link if there is any possibility of eventually ending up worse off than in the status quo. All other considered stability concepts either predicted stability for networks that were not empirically stable, or classed as unstable

they implement.

[4]The myopic stability concepts we consider are pairwise stability (Jackson and Wolinsky, 1996) and the pairwise myopically stable set (Herings et al., 2009).

[5]Theoretically, farsighted behavior is identified by such concepts of stability as the pairwise farsightedly stable set, von Neumann-Morgenstern pairwise farsightedly stable set, largest pairwise consistent set (Herings et al. (2009), von Neumann and Morgenstern (1944), Chwe (1994)), the set of farsightedly consistent networks (Page Jr et al., 2005) and the cautious path stable set (Teteryatnikova, 2015).

networks that were, in fact, empirically stable. Thus, if one wants to identify only networks which have a high likelihood of showing strong stability, myopic concepts are best, but if one wishes to identify all networks which have the potential to be stable, cautious farsighted concepts are preferable.

The relatively stronger stability of myopically stable networks is highlighted by a subtle difference between the two network formation games used in the experiment. In the first game, the complete network is, in theory, myopically stable, while the one-link networks are farsightedly but not myopically stable. In the second game, we change only the payoffs of the unstable two-link networks, but in such a way that the one-link networks become myopically stable and the complete network only (cautious) farsightedly stable. This change in payoffs of an unstable intermediate network leads to a complete reversal in observed stability: the vast majority of stable networks in the first game are complete networks, with a small but non-negligible number of stable one-link networks; in the second game, the bulk of stable outcomes are one-link networks, with relatively few stable complete networks.

The primary advantage of using a lab experiment rather than a field study is that we know the precise payoffs associated with each network; in an empirical study, it may be difficult to observe not only payoffs, but even which networks are in place at a given time. Furthermore, our relatively simple experimental setting gives theory its best chance by maximising the possibility for subjects to fully understand the environment and the likely consequences of their actions. Although external validity may be questioned, testing theoretical predictions in such a simplified setting is useful as the stability concepts we consider are general, and should apply equally in the laboratory as in the outside world. Thus, if a concept is not useful in explaining behaviour in the simplified laboratory environment, it is also unlikely to have predictive power in more complex "real life" applications.

There is a small but growing experimental literature on "pure" network formation games such as ours, where payoffs are derived directly from the network structure rather than from further interactions between linked players.[6] These kinds of games are used to focus on network stability and not risk confounds with behaviour in unrelated interactions. We are aware of only three experiments examining cooperative pairwise stability in pure network formation games: Pantz (2006), Carrillo and Gaduh (2012), Kirchsteiger et al. (2013).[7] Unlike our experiment, in all of these only one final network is paid, and the games are played with a strict timing of moves. Only Kirchsteiger et al. (2013) considers

[6]Experiments where payoffs are derived from games played between individuals who had chosen to form a link in an earlier stage include Hauk and Nagel (2001) and Corbae and Duffy (2008).

[7]Experiments testing (Nash) stability of networks in a non-cooperative framework include Goeree et al. (2009), Falk and Kosfeld (2012), Callander and Plott (2005), Berninghaus et al. (2006).

farsighted behaviour, but in their design, all farsightedly stable networks are also myopically stable, so identification of farsighted behaviour is problematic.

The paper proceeds as follows: Section 2 discusses the theoretical stability concepts we consider and their predictions for the two network formation games we implement in the experiment; Sections 3 and 4 describe our laboratory experiment and the results; finally, Section 5 concludes with a discussion of the implications of our results regarding desirable features of network stability concepts. Additional tables, formal theoretical definitions and experiment instructions are provided in the Appendix.

2 Experimental Games and Theoretical Stability Concepts

In this section we first describe the two network formation games played in the experiment. We then introduce the relevant theoretical stability concepts, and explain their different predictions in the experimental games. Because we discuss a large number of concepts, some of which involve lengthy technical definitions, in the main text we provide only a rough intuition of why each prediction arises. Formal definitions are provided in the Appendix. We emphasise here that the purpose of this experiment is not to differentiate between *each* of the myriad theoretical stability concepts that have been proposed in the literature – this would require a huge number of treatments given the number of concepts and the overlap in many of their predictions. We attempt only to differentiate between three broad classes of concepts, namely those based on myopic, farsighted, and cautious farsighted behaviour.

The network formation games we implement in our experiment are shown in Figures 1 and 2. Each game consists of the set of all possible networks for the three-player case. These are the *empty* network g_0 (where no links have formed), three 1-link networks g_1, g_2, g_3, three 2-link networks g_4, g_5, g_6, and the *complete* network g_7 (where all three links are formed). For convenience of exposition, we define G_j^i as the network, or set of networks, in Game i with j links, for $i \in \{I, II\}$ and $j \in \{0, ..., 3\}$. Nodes of a network represent players and links indicate bilateral relationships between players. The payoff of a player in each network is represented by a number next to the corresponding node. In both games these payoffs are anonymous, in the sense that all players in a symmetric position receive the same payoff – both within each network and across networks of the same type. Moreover, in all networks, apart from the empty network, the sum of players' payoffs is identical, which makes these networks equally efficient and hence, rules out efficiency concerns as an explanation of behaviour.[8]

[8]The empty network was assigned very low payoffs in the hope that it would almost never occur, allowing us to focus on three networks which are identified in different combinations by different stability concepts.

We chose this setting with three players and anonymous allocation of payoffs to keep things as simple as possible and maximize the chances that subjects in our experiment fully understand the environment. This is particularly important in view of our interest in testing farsighted behaviour in network formation, which requires subjects to understand not only the payoff structure of the games, but also chains of others' reactions. This simple setting with a small number of players suffices to identify the relative importance of myopic and different types of farsighted stability.

As is most relevant to our experiment, we confine attention to the *cooperative pairwise* theory of network formation, where creation of a link requires the consent of both involved players, while severance of a link is a unilateral decision of either player involved in the link.[9] This theory focuses on characterizing the outcome rather than the process of network formation, leaving the process itself largely undefined or unstructured.[10] Among the cooperative pairwise concepts, two main approaches have been proposed in the literature.

The first approach assumes that players behave myopically, in the sense that their decision to add or delete a link is guided completely by payoffs that can be obtained immediately after making the change. In particular, players do not take into account that others might react to their actions by adding or deleting some other links, which might eventually lower or increase the payoffs of the original individual(s). The most well-known concept within this approach is pairwise stability, or PWS (Jackson and Wolinsky, 1996), and its set-valued version, pairwise myopically stable set, or PWMS (Herings et al., 2009).[11] According to Jackson and Wolinsky (1996), a network is considered pairwise (myopically) stable if no individual player can immediately benefit from deleting one of her links, and no pair of individuals can benefit from forming a link. It is easy to see that this holds only for the complete network in Game I (G_3^I) and for each of the 1-link networks in Game II (G_1^{II}).

[9] Two alternative approaches are explicitly modeling a network formation game and using non-cooperative equilibrium concepts, or considering deviating coalitions of more than two players. Examples of the former include Myerson (1991), Jackson and Watts (2002b), Bala and Goyal (2000), Hojman and Szeidl (2008), Bloch (1996), Currarini and Morelli (2000), Galeotti and Goyal (2010). Examples of the latter, with considerations of farsightedness in network formation, include Aumann and Myerson (1988), Chwe (1994), Xue (1998), Dutta and Mutuswami (1997), Page Jr et al. (2005), Page Jr and Wooders (2009), Herings et al. (2004), Mauleon and Vannetelbosch (2004). Note that the limited cooperation between the two players involved in a link establishes an important distinction between cooperative pairwise stability and coalitional stability. That is, while in the pairwise approach, only special 2-player "coalitions" can form, the cooperation in such coalitions is only partial, and every player has a natural "unilateral" domain of action. See discussion in Dutta et al. (2005).

[10] This is a key difference from the approach in non-cooperative network formation models, or in "hybrid" models such as Dutta et al. (2005), where the protocol of network formation (order of moves, choice of players to make a move at every period, players' strategies, etc.) is specifically defined.

[11] A notion of the pairwise myopically stable set relies on the assumption that players make changes along a *myopic improving path* of networks, which was first defined by Jackson and Watts (2002a).

The second approach to cooperative pairwise stability assumes that players are farsighted and do take into account the chains of reactions that might follow after their own move. Almost all pairwise farsighted stability concepts rely on the idea that players form or sever links based on the improvement that an eventual stable network offers relative to the current network. These include the pairwise farsightedly stable set (PWFS), von Neumann-Morgenstern pairwise farsightedly stable set (vN-MFS), largest pairwise consistent set (LPWC), all introduced in Herings et al. (2009), and the set of farsightedly consistent networks (FCN) introduced in Page Jr et al. (2005).[12] Following an alternative approach, the cautious path stable set, or CPS (Teteryatnikova, 2015), assumes that players add or delete links taking into account not only the payoff associated with a final stable network, but also payoffs from intermediate networks.

The predictions of these farsighted stability concepts in our experimental games can be intuitively understood as follows. In Game I, all pairwise farsighted concepts identify the 1-link networks (G_1^I) and the complete network (G_3^I) as stable. For example, the 1-link networks are farsightedly stable because although either of the linked players could achieve a short-term gain by forming a link with the third player, they do not do so as they foresee that the other two players would then have an incentive to form the last link, leaving them with a payoff of 22 rather than 30.[13] Making a change in a 1-link network therefore "guarantees" that the eventual payoff of Player 1 will be worse. Similarly, in the complete network, cutting a link "guarantees" that the player's payoff will not improve: in this case even the short-term payoff of the player is worse than in the complete network ($17 < 22$).

The same certainty regarding the negative consequences of a link change clearly exists for each 1-link network in Game II (G_1^{II}), as 30 is the maximum payoff players can achieve, and so these networks are stable under all farsighted concepts. However, in the complete network of Game II (G_3^{II}), deleting a link *may or may not* eventually decrease a player's payoff, depending on the subsequent action chosen by the player with two links remaining, who now has an incentive to cut a further link: if the link that is deleted at this second step is with the original player, she will be worse off, however, if it is with the third player she will be better off. The *possibility* of an eventual payoff improvement as a result of cutting a link in the complete network makes this network unstable according to many farsighted

[12] All of these concepts apart from PWFS are based on earlier definitions in von Neumann and Morgenstern (1944) and Chwe (1994).

[13] This is also true if players care not only about payoffs in a final network but also about payoffs at intermediate steps of network formation, as is allowed by CPS. For example, if players weigh payoffs in all networks emerging in the process of network formation equally, then the simple average payoff of Player 1 from staying in 1-link network is 30, while the simple average from moving to 2-link network and then to the complete network is $\frac{1}{2}(32 + 22) = 27$, which is less than 30.

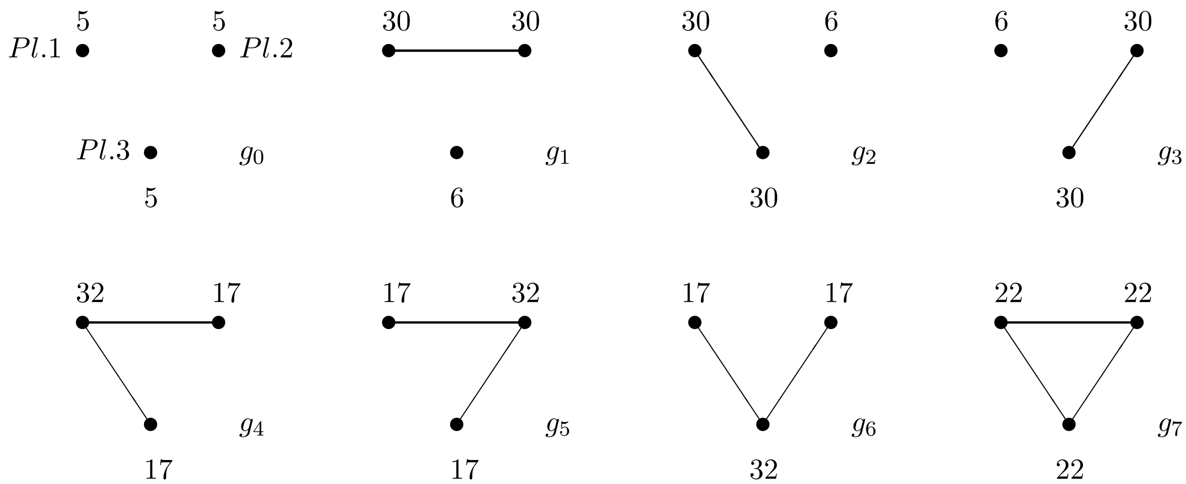


Figure 1: Game I.

concepts of stability (PWFS, vN-MFS). It is only when players are sufficiently cautious or pessimistic about future network changes that they prefer to avoid the risk and do not delete a link in the complete network.[14] This intuition is captured by such concepts of stability as LPWC, FCN and CPS, which identify the complete network as stable.[15]

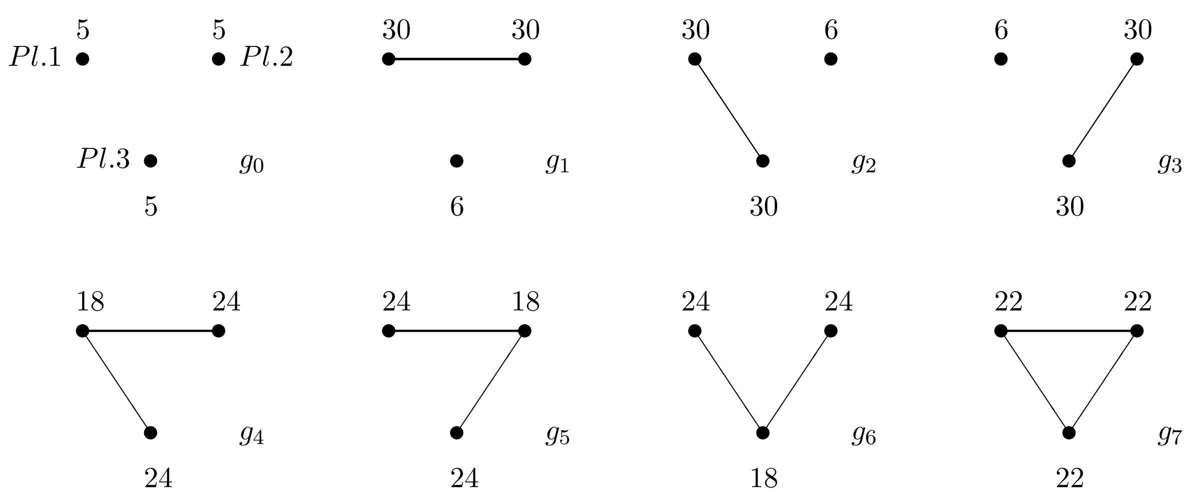


Figure 2: Game II.

[14] Alternatively, the risk of ending up with a worse payoff could be avoided if the second player in the 2-link network could credibly commit to delete the "correct" link. In our experimental games and in many real-life applications that we are interested in such commitment is not possible.

[15] The idea behind these concepts is that no player has an incentive to deviate by adding or deleting a link in a network if there exist further deviations leading to some network (in the stable set) where the initially deviating players are not better off, and possibly worse off. CPS introduces an extra twist to this idea by allowing players to derive utility from intermediate steps, and not only from the final network. For example, in Game II, the complete network belongs to CPS set when players care about the simple average of their payoffs in all emerging networks because deleting a link may reduce the average payoff $\left(\frac{1}{2}(24+6) < 22\right)$.

Concept	Game I	Game II	Predictions
PWS	g_7	$g_1,\ g_2,\ g_3$	$G_3^I; G_1^{II}$
PWMS	$\{g_7\}$	$\{g_1, g_2, g_3\}$	$G_3^I; G_1^{II}$
PWFS	$\{g_1, g_2, g_3, g_7\}$, $\{g_1, g_6, g_7\}$, $\{g_2, g_5, g_7\}$, $\{g_3, g_4, g_7\}$, $\{g_4, g_5, g_7\}$, $\{g_4, g_6, g_7\}$, $\{g_5, g_6, g_7\}$	$\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$G_1^I, G_2^I, G_3^I; G_1^{II}$
vN-MFS	$\{g_1, g_2, g_3, g_7\}$	$\{g_1\}$, $\{g_2\}$, $\{g_3\}$	$G_1^I, G_3^I; G_1^{II}$
LPWC	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_4, g_5, g_6, g_7\}$	$G_1^I, G_3^I; G_1^{II}, G_2^{II}, G_3^{II}$
FCN	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$G_1^I, G_3^I; G_1^{II}, G_3^{II}$
CPS	$\{g_1, g_2, g_3, g_7\}$	$\{g_1, g_2, g_3, g_7\}$	$G_1^I, G_3^I; G_1^{II}, G_3^{II}$

Table 1: Summary of predictions

In summary, in Game I all pairwise myopic and farsighted concepts of stability identify the complete network as stable; in addition, all farsighted concepts identify 1-link networks as stable. In Game II, on the contrary, all myopic and farsighted concepts of stability identify 1-link networks as stable; in addition, the cautious farsighted concepts identify the complete network as stable. These predictions of different farsighted and myopic stability concepts in both games are summarized in Table 1.[16] For simplicity, in the following we will refer to the network which is stable only under cautious farsighted concepts (G_3^{II}) as cautious farsighted (cFS), those which are stable under all farsighted but *not* myopic concepts of stability (G_1^I) as farsightedly stable (FS), and we will refer to networks that are myopically stable (G_3^I and G_1^{II}) as pairwise stable (PWS), keeping in mind that they are also farsightedly stable.

In addition to the networks that we describe as FS and cFS, each of the 2-link networks is also identified as stable by one of the farsighted concepts: PWFS in Game I and LPWC in Game II. The reason has to do with the details of both concepts' definitions, which we leave for the Appendix. Inspecting the payoffs of the games, and as discussed further in the Appendix, it is intuitively unlikely that these networks will exhibit stability. Thus we do not refer to these networks as farsightedly stable.

3 Experimental Design

In this section we first describe the playing screen, the manner in which subjects interacted, and further procedural details. We then define precisely the hypotheses we are testing and their relationship to the theory described in the previous section.

[16]To be close to our experimental design, CPS predictions are calculated under the assumption that players' payoffs associated with a sequence of networks are defined by the arithmetic average of payoffs in all networks of the sequence.

3.1 Experimental Procedures

On the playing screen, subjects saw themselves represented as a green circle at the bottom of the screen and the other two players as blue circles at the top. Links could be formed between two players in the following way. A subject could indicate they were willing to form a link with another player by clicking on the appropriate blue dot, resulting in a pink arrow pointing towards the other player. Clicking again would undo this action. If two players had both clicked on each other then a link was formed and shown in red. Links could be broken by either of the parties clicking on the other, leaving a pink arrow pointing towards the player who had broken the link. All of these actions could be taken at any time.

Payoffs were made at one second intervals according to the network described by the red links. Each player's per-second payoffs were displayed next to their circle. Each game lasted 30 seconds: we chose this duration as a compromise between lasting long enough for stable networks to emerge, but short enough that boredom was unlikely to motivate subjects to leave otherwise stable networks. We refer to each 30 second game as a "round". The total points accrued were displayed and updated throughout the round, as was the number of seconds remaining.

The playing screen seen by subjects is shown in the instructions in the Appendix. A video of a sample round can be viewed at `http://homepage.univie.ac.at/Mariya.Teteryatnikova/Research/network.wmv`.

Before commencing the paid rounds, subjects completed a detailed tutorial familiarizing them with the interface, and played three practice rounds with payoffs different from those in the game of interest. Each session consisted of 20 incentivised rounds with groups randomly rematched after each.[17] The starting network was randomly determined at the beginning of each round.

The experiments were programmed in Z-tree (Fischbacher, 2007) and took place at the Vienna Center for Experimental Economics. For each game, four sessions were conducted, each consisting of 18 subjects divided into two matching groups, giving us a total of 16 independent observations. One randomly chosen period was paid, with every 45 points exchanged for 1 Euro. Sessions lasted approximately one hour.

[17] One advantage of our design is that stable networks can arise very quickly, with the brevity of each game allowing for a large number of repetitions and thus greater learning. In earlier experiments, subjects played only three or four times.

3.2 Predictions

We use the sets identified by the theories described in Section 2 as predictions about which networks will exhibit stability in our experiment. We are aware that many of these theories are not formally defined for environments with dynamic payoffs, and we do not claim that our experiment is a strict test of any theory. However we argue that it is reasonable to use them to provide predictions for our experiment, as the intuition underlying most of these concepts tends to be dynamic in nature. For example, networks that are PWS according to Jackson and Wolinsky (1996) are justified as being stable because any deviation would incur an immediate cost. Consider an alternative possible experimental design which implements a network formation game with two PWS networks, one Pareto-superior to the other, and where subjects are paid only for a final stable network: it is highly unlikely that the Pareto-inferior network will ever be paid out, because subjects could move to the Pareto-superior one at no cost, but this is hardly a fair test of the theory, as the original idea behind the concept, i.e. an immediate cost of deviation, is given no chance to play a role.

The identification of networks as empirically stable requires an absolute definition of stability. Strictly speaking, the theories suggest that a stable network, once entered, will remain in place with probability one. However, given the inevitable noise in human behaviour, this criterion is too strict. We define a type of network as being stable with respect to $\bar{q}$ if such a network that is in place for a given payment remains in place for the next payment with probability greater than some fixed probability $\bar{q}$:[18]

Definition 1 *A type of network is* stable with respect to probability $\bar{q}$ *if, conditional upon being paid in the current period, the probability of it being paid in the next is greater than* $\bar{q}$.

Let q be the proportion of paid networks of type G^i_j that remain in place for the next payment. Thus, there is statistical evidence that network G^i_j is stable according to Definition 1 if we can reject the null hypothesis $H_0 : q \leq \bar{q}$.

Which networks are determined to be stable clearly depends on the value chosen for $\bar{q}$. In the results section we first identify the networks which are stable for $\bar{q} = 0.5$. We view this as a minimal criterion for stability, under which networks are more likely to remain in place than not. We then show how this set of networks changes as the requirement for stability becomes more strict (i.e. as $\bar{q}$ increases to one).

[18] We cannot simply use the average duration of networks as the basis of a measure of stability because the final network in every game is censored.

Each theoretical stability concept divides the set of all networks into those which are stable and those which are unstable, as shown in Section 2. Rather than formally define a hypothesis for each concept, it is more straightforward to identify the networks which are stable, and compare this set of empirically stable networks to the sets that are predicted theoretically. For a concept to be validated, the empirically and theoretically stable sets should coincide precisely.[19] Thus, theories can be rejected either for failing to predict networks which turn out to be empirically stable, or predicting networks to be stable when they are not.

We are also interested in whether networks that are myopically stable are more stable than those that require farsighted behaviour. We therefore compare the relative stability of each type of network, predicting that networks identified by myopic concepts are more stable than those predicted only by farsighted concepts.

4 Results

We begin this section by giving a descriptive overview of the data before proceeding to formal statistical tests. The first set of tests ask whether or not each type of network is "stable" using the absolute definition of stability described in the previous section (Definition 1). The second set of tests are relative, comparing the stability of different types of networks.

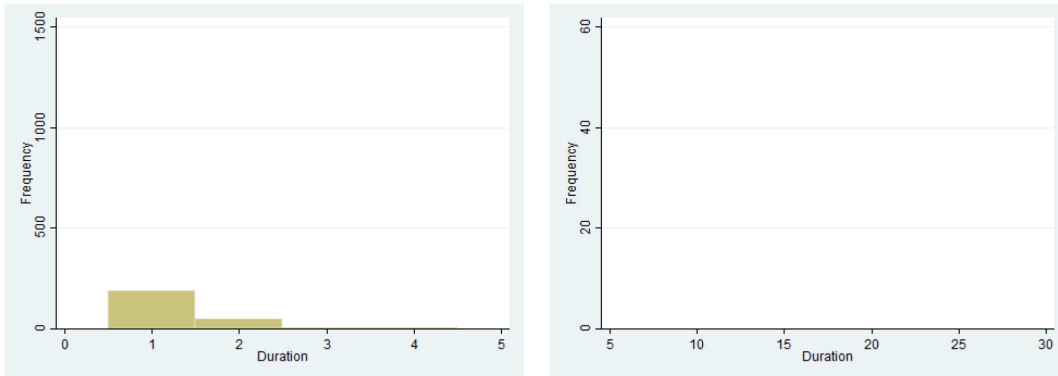
To give a full picture of the raw data, in the initial descriptive analysis we focus on the simplest statistic capturing the stability of a network, the *duration* of a network. We measure the duration of a network as the number of payments that occur between when the network is formed and when a different network becomes the basis for payment.[20] We regard the one-second period between payments as non-binding negotiation, so, for example, if a link is broken and reformed between payments, we do not consider the second network to be new. We refer to the number of consecutive payments of a given network as its duration in *periods*.

The duration and frequencies of each type of network in each game are displayed in Figures 3 and 4. To accommodate the large number of networks of short duration while keeping visible differences in the distributions of more stable networks, the data is split between durations of ≤ 5 and > 5 periods.

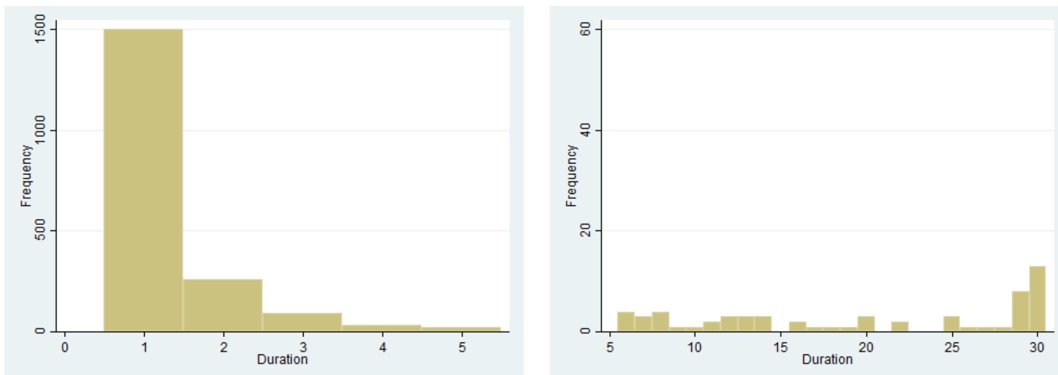
[19] Note that this is different from testing outcomes in a multi-equilibria environment, where failure to observe one of the equilibria does not invalidate the theory. We are not asking which networks will arise, but which networks are stable given that they have been entered, which is why *all* networks identified by a concept must be stable.

[20] While being a clear indicator of stability, as remarked earlier (see footnote 18) the average duration of networks cannot be used in a formal test of stability due to censoring of final networks in every game.

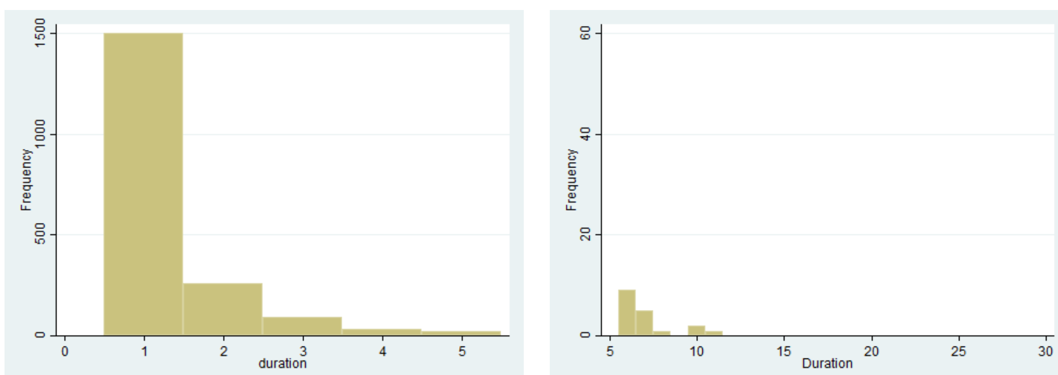
The PWS networks G_3^I and G_1^{II} clearly display the greatest stability, often lasting upwards of 15 seconds. The FS networks G_1^I and the cFS network G_3^{II} also often last more than half the periods of a game, but much less frequently than the PWS networks. By contrast, distributions of G_2^I (identified by PWFS) and G_2^{II} (identified by LPWC) do not possess these long tails, and there is only one round in which one of these two types of networks lasts longer than half the game. The empty networks occur rarely, and seldom last more than two seconds. As the only networks observed to last entire games, it is already clear that G_3^I, G_1^{II}, G_1^I, and G_3^{II} display a *potential* for strong stability, which the other networks do not.



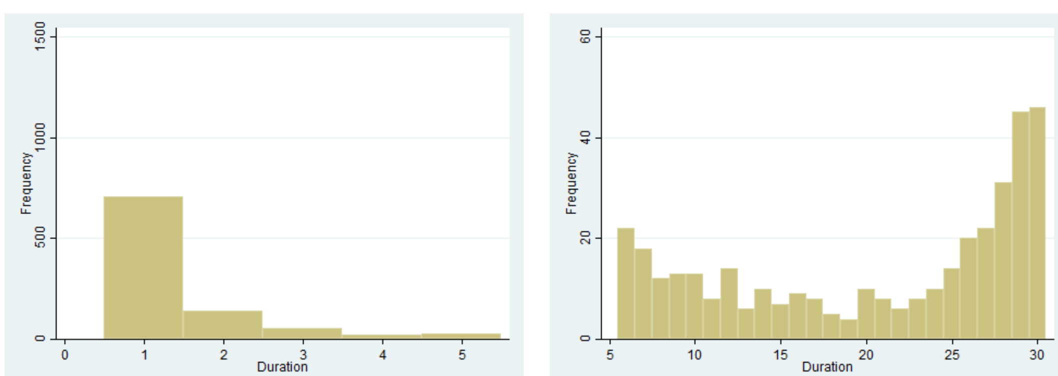
Empty (G_0^I)



$1 - Link\ \ (G_1^I)$

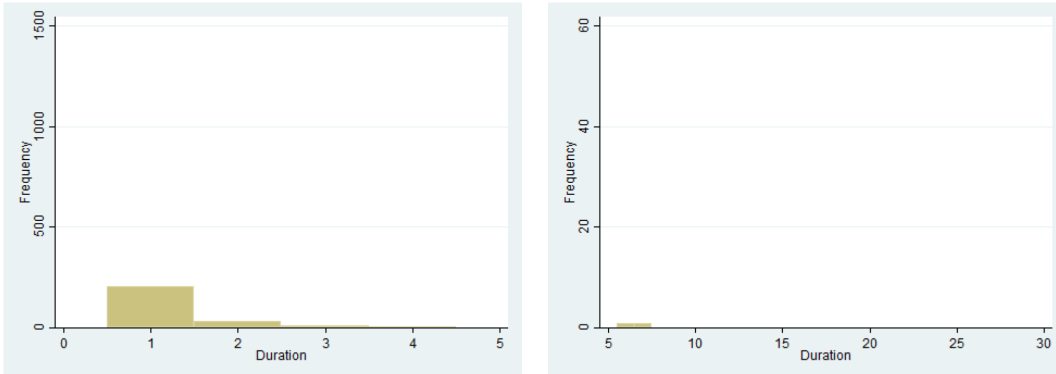


$2 - Link\ \ (G_2^I)$

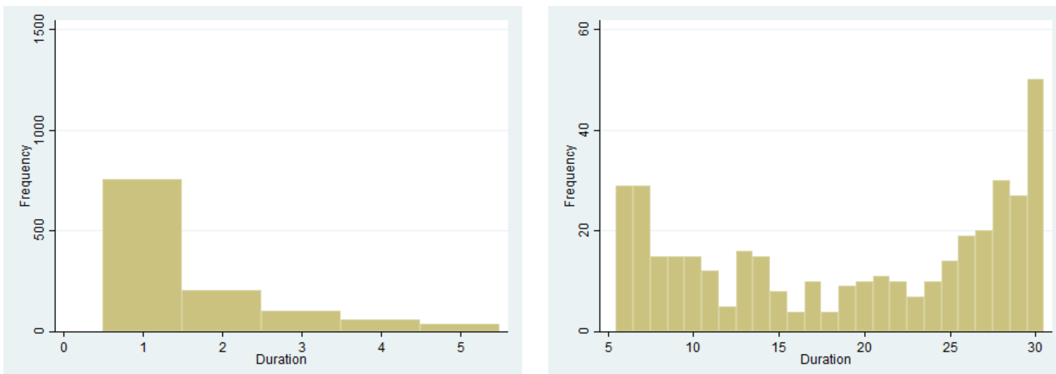


Complete (G_3^I)

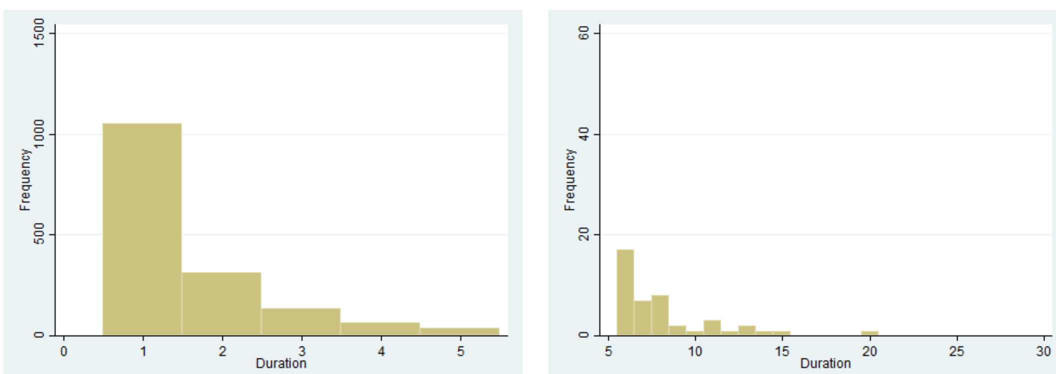
Figure 3: Duration and frequency of networks (Game I)



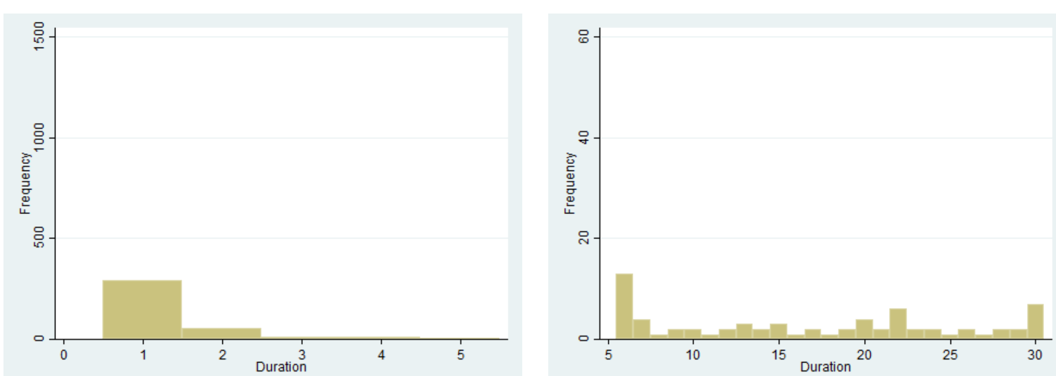
Empty (G_0^{II})



$1 - Link\ \ (G_1^{II})$



$2 - Link\ \ (G_2^{II})$



Complete (G_3^{II})

Figure 4: Duration and frequency of networks (Game II)

We turn now to formal statistical tests using the measure of stability proposed in Definition 1. Table 2 reports the proportions of paid networks of each type that are also paid in the subsequent period, along with standard errors clustered by matching group. These proportions disaggregated by matching group can be found in the Appendix (Table 4).

	Proportions (q)		
Network	Game 1	Game 2	Difference (Δ)
Empty	0.33	0.29	-0.04
	(0.027)	(0.031)	(0.040)
One-link	$0.68^{\circ\circ}$	$0.89^{\circ\circ\circ}$	0.21***
	(0.065)	(0.025)	(0.0.67)
Two-link	0.45	0.52	0.07
	(0.022)	(0.039)	(0.043)
Complete	$0.92^{\circ\circ\circ}$	$0.79^{\circ\circ\circ}$	-0.13**
	(0.026)	(0.048)	(0.053)

Table 2: Proportion of paid networks that are also paid in the subsequent period (q). Standard errors clustered by matching group in parentheses. $^{\circ\circ}(^{\circ\circ\circ})$ $H_0 : q \leq 0.5$ rejected, $p < 0.05$ ($p < 0.01$). **(***) $H_0 : \Delta = 0$ rejected, $p < 0.05$ ($p < 0.01$).

We begin by testing whether each type of network is stable according to Definition 1 with $\bar{q} = 0.5$ (Section 3.2). With empirical probabilities of remaining in the same network of 0.92 and 0.89 respectively, G_3^I and G_1^{II} are found to be stable according to one-tailed t-tests ($p < 0.01$). The equivalent probabilities for G_1^I and G_3^{II} are 0.68 and 0.79, both significantly greater than 0.5 ($p < 0.05$ and $p < 0.01$, respectively).

The empty networks in both games, and the 2-link networks in Game I are less likely to remain in place than be left, so cannot be stable according to our definition. The probability of remaining in the 2-link networks in Game II is 0.52, but this is not significantly greater than 0.5 ($p = 0.3$).

Comparing the set of networks that are empirically stable according to these tests to the sets identified by different theoretical stability concepts, we can see that only CPS and FCN identify them precisely: in Game I, PWS and PWMS do not include the stable G_1^I, and the PWFS contains the unstable G_2^I; in Game II, none of PWS, PWMS, PWFS, or vN-MFS contain the stable G_3^{II}, while the LPWC set identifies the unstable G_2^{II}.

Figure 5 shows how the set of stable networks varies as the criterion for stability becomes stricter. For each value of $\bar{q} \in [0.5, 1]$ we show the networks that are identified as stable at the 5% level, i.e. where we can reject the hypothesis that a type of network remains in place with probability less than

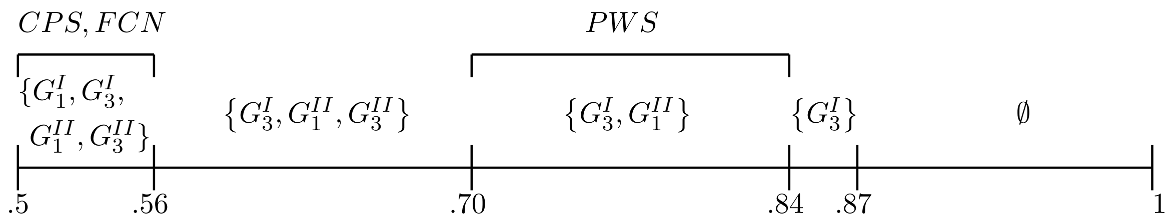


Figure 5: Sets of networks which are stable with respect to $\bar{q} \in [0.5, 1]$. A network G_j^i which remains in place with empirical probability q is considered stable if $H_0 : q \leq \bar{q}$ is rejected with $p < 0.05$.

or equal to $\bar{q}$.

For $\bar{q} \in [0.5, 0.56]$, there is no change in the set of stable networks. For $\bar{q} \in [0.56, 0.70]$, G_1^I is no longer stable, and none of the theoretical concepts we consider precisely identifies the remaining three networks. For $\bar{q} \in [0.70, 0.84]$, the only stable networks are those identified by PWS. For $\bar{q} \in [0.84, 0.87]$, only G_3^I is stable, which is again not the prediction of any theory, and for $\bar{q} > 0.87$, no network can be categorized as stable.

We turn now to comparative tests. First, comparing the stability of networks with equivalent structure and payoffs *across games*, there is a strong treatment effect. As can be seen in Table 2, changing the payoffs attached to the 2-link network from those in Game I to those in Game II has no significant impact on the stability of the empty network. However, it increases the probability of remaining in the 1-link network by 0.21 ($p < 0.01$) as it becomes PWS, and decreases the probability of remaining in the complete network by 0.13 ($p < 0.05$) as it becomes only FS. Despite the change in payoffs, there is no significant change in the stability of the 2-link networks.

To compare the stability of networks *within a game* we regress the probability of remaining in the current network on the different types of networks, with the 2-link network as the comparison group. Table 3 presents the results of probit regressions, the first column using data from only Game I, the second from only Game II, and the third pooling the data from both games. Because in the final regression we want a variable for *all* pairwise-stable networks, the dummy variable PWS takes the value one for networks in G_3^I and G_1^{II}. In line with the notation introduced earlier, we label the dummy variables indicating G_I^I and G_3^{II} as FS and cFS, respectively.

As can be seen from the first column of Table 3, the networks in Game I can be ranked from least to most stable as empty, 2-link, 1-link, complete, with all relationships significant ($p < 0.01$). The second column reports the results for Game II, ranking the networks from least to most stable as empty, 2-link, complete, 1-link, with all relationships strongly significant ($p < 0.01$) apart from the last, with

VARIABLES	(1) stay	(2) stay	(3) stay
Empty	-0.0902***	-0.197***	-0.143***
	(0.0246)	(0.0519)	(0.0314)
PWS	0.446***	0.351***	0.371***
	(0.0191)	(0.0311)	(0.0374)
FS	0.131***		0.110***
	(0.0189)		(0.0244)
cFS		0.150***	0.130***
		(0.0362)	(0.0421)
Complete			0.0470
			(0.0527)
Observations	13,920	13,920	27,840

PWS: Complete network in Game I and 1-link networks in Game II;
FS: 1-link networks in Game I; cFS: Complete network in Game II.
Standard errors clustered by matching group in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Probability of remaining in current network

the 1-link network being only weakly more stable than the complete ($p = 0.068$).

The last column of Table 3 addresses the possibility that the stability of the complete network is due in part to the fact that it represents an equal share of the surplus to all players, and thus may be focal, or appeal to subjects with fairness concerns. To test this we pool the data from the two games, and include as a regressor a dummy for being a complete network. This controls for fairness concerns because the coefficient on the dummy variable for the complete network in Game II (cFS) now represents extra stability given by factors other than factors shared with the complete network in Game I. As can be seen from the final column of Table 3, the coefficient on the complete network dummy is insignificant and close to zero, whereas all the findings from the previous two regressions are unchanged. We are therefore confident that the stability of the complete networks is not due to focality or fairness concerns.

5 Conclusion

We find that the most stable networks in two experimental network formation games with flows of payoffs are those identified by myopic stability concepts. The evidence comes in three forms: firstly, they remain in place for the whole duration of games with greatest frequency; secondly, comparing results across games we see that networks with identical structure and payoffs are significantly more stable when they are theoretically myopically stable; finally, within each game, they are the networks that when entered are most likely to remain in place for the next period.

While myopically stable networks are clearly the most stable, we also find convincing evidence of farsighed behaviour. Both farsighted and cautious farsighted networks exhibit stability, in that they have the potential to last the full 30 seconds of a round, and also in that once they are in place, they are significantly more likely to remain than not. They are also significantly more stable than all networks that do not satisfy these conditions.

Considering which concepts most precisely identify the set of stable networks, we find that this depends on the strictness of one's empirical definition of stability. The strictness of a definition depends on the probability with which a type of network must remain in place in order to be classed as stable, with stricter definitions requiring higher values of this probability. We find that myopic concepts precisely identify the set of stable networks for a range of high probabilities, but for definitions which are not so strict they fail to identify all stable networks. For a range of less strict definitions, it is the two cautious farsighted concepts that predict exactly the empirically stable networks.

We conclude that the appropriateness of myopic versus farsighted concepts of stability in predicting outcomes of network interactions which involve a regular flow of payoffs depends on the purpose of the prediction. If the aim is to predict the outcomes that are most likely to arise, the myopic concepts perform best, as they identify the networks that are most consistently and strongly stable. If on the other hand, the aim is to identify the full range of outcomes that may achieve stability, or if a weaker definition of stability is acceptable, then the concepts that perform best are farsighted and assume pessimism regarding final outcomes.

References

Aumann, R. and R. Myerson (1988). Endogenous formation of links between players and coalitions: an application of the shapley value. *The Shapley Value*, 175–191.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1229.

Berninghaus, S. K., K.-M. Ehrhart, and M. Ott (2006). A network experiment in continuous time: The influence of link costs. *Experimental Economics 9*(3), 237–251.

Bloch, F. (1996). Sequential formation of coalitions in games with externalities and fixed payoff division. *Games and Economic Behavior 14*(1), 90 – 123.

Callander, S. and C. R. Plott (2005). Principles of network development and evolution: An experimental study. *Journal of Public Economics 89*(8), 1469–1495.

Carrillo, J. and A. Gaduh (2012). The strategic formation of networks: Experimental evidence.

Chwe, M. S.-Y. (1994). Farsighted coalitional stability. *Journal of Economic Theory 63*(2), 299 – 325.

Corbae, D. and J. Duffy (2008). Experiments with network formation. *Games and Economic Behavior 64*(1), 81–120.

Currarini, S. and M. Morelli (2000). Network formation with sequential demands. *Review of Economic Design 5*(3), 229–249.

Dutta, B., S. Ghosal, and D. Ray (2005). Farsighted network formation. *Journal of Economic Theory 122*(2), 143–164.

Dutta, B. and S. Mutuswami (1997). Stable networks. *Journal of Economic Theory 76*(2), 322–344.

Falk, A. and M. Kosfeld (2012). It's all about connections: Evidence on network formation. *Review of Network Economics 11*(3).

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics 10*(2), 171–178.

Galeotti, A. and S. Goyal (2010). The law of the few. *The American Economic Review*, 1468–1492.

Goeree, J. K., A. Riedl, and A. Ule (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior 67*(2), 445–466.

Hauk, E. and R. Nagel (2001). Choice of partners in multiple two-person prisoner's dilemma games an experimental study. *Journal of conflict resolution 45*(6), 770–793.

Herings, P., A. Mauleon, and V. Vannetelbosch (2009). Farsightedly stable networks. *Games and Economic Behavior 67*(2), 526–541.

Herings, P., A. Mauleon, and V. J. Vannetelbosch (2004). Rationalizability for social environments. *Games and Economic Behavior 49*(1), 135–156.

Hojman, D. A. and A. Szeidl (2008). Core and periphery in networks. *Journal of Economic Theory 139*(1), 295–309.

Jackson, M. O. and A. Watts (2002a). The evolution of social and economic networks. *Journal of Economic Theory 106*(2), 265–295.

Jackson, M. O. and A. Watts (2002b). On the formation of interaction networks in social coordination games. *Games and Economic Behavior 41*(2), 265–291.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Kirchsteiger, G., M. Mantovani, A. Mauleon, V. Vannetelbosch, et al. (2013). Limited farsightedness in network formation. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Mauleon, A. and V. Vannetelbosch (2004). Farsightedness and cautiousness in coalition formation games with positive spillovers. *Theory and Decision 56*(3), 291–324.

Myerson, R. B. (1991). Game theory: analysis of conflict. *Harvard University*.

Page Jr, F. H. and M. Wooders (2009). Strategic basins of attraction, the path dominance core, and network formation games. *Games and Economic Behavior 66*(1), 462–487.

Page Jr, F. H., M. H. Wooders, and S. Kamat (2005). Networks and farsighted stability. *Journal of Economic Theory 120*(2), 257–269.

Pantz, K. (2006). *The strategic formation of social networks: experimental evidence*. Shaker Verlag gmbh, Germany.

Teteryatnikova, M. (2015). Cautious farsighted stability in network formation games with stream of payoffs. *Working paper, University of Vienna*.

von Neumann, J. and O. Morgenstern (1944). Theory of games and economic behavior. *Princeton University Press, Princeton*.

Xue, L. (1998). Coalitional stability under perfect foresight. *Economic Theory 11*(3), 603–627.

A Formal definitions

In this section we formally define theoretical concepts of stability discussed in the paper. While the application of the defined concepts to our experimental games is rather straightforward, technical details for some of the concepts are also explained below. The remaining details are available from the authors.

Let $Y_i(g)$ be a payoff that player i obtains in a network g, and let the set of all possible networks on n nodes be denoted by $\mathbb{G}$. If players i and j are linked in the network g, this is denoted by $ij \in g$; otherwise, $ij \notin g$. Networks obtained by adding or deleting a link ij to/from an existing network g are denoted by $g + ij$ and $g - ij$, respectively.

Following Jackson and Wolinsky (1996), a network g is defined to be *pairwise stable*, or PWS, if no player can immediately benefit from deleting one of her links, and no pair of players can benefit from forming a link.

Definition 2 *Network $g \in \mathbb{G}$ is pairwise stable if*

(i) for all $ij \in g$, $Y_i(g) \geq Y_i(g - ij)$ and $Y_j(g) \geq Y_j(g - ij)$, and

(ii) for all $ij \notin g$, if $Y_i(g) < Y_i(g + ij)$ then $Y_j(g) > Y_j(g + ij)$.

The definition of the *pairwise myopically stable set* of networks, or PWMS (Herings et al., 2009), requires introducing a *myopic improving path* first. It is a finite sequence of networks that can emerge when players form or sever links based on the improvement that the immediately resulting network offers them relative to the current network. Formally, the definition in Herings et al. (2009) states that a myopic improving path from a network g to a network $g' \neq g$ is a finite sequence of networks $g_1, .., g_K$ with $g_1 = g$ and $g_K = g'$ such that for any $1 \leq k \leq K - 1$ either

(i) $g_{k+1} = g_k - ij$ for some ij such that $Y_i(g_{k+1}) > Y_i(g_k)$ or $Y_j(g_{k+1}) > Y_j(g_k)$, or

(ii) $g_{k+1} = g_k + ij$ for some ij such that $Y_i(g_{k+1}) > Y_i(g_k)$ and $Y_j(g_{k+1}) \geq Y_j(g_k)$.

If there exists a myopic improving path from g to g', this is denoted by $g \mapsto g'$, and for any network g, $M(g) = \{g' \in \mathbb{G} | g \mapsto g'\}$. In terms of this notation, a pairwise myopically stable set can be defined as follows.

Definition 3 *A set of networks $G \subseteq \mathbb{G}$ is pairwise myopically stable if*

(i) $\forall\ g \in G$,

(ia) $\forall ij \notin g$ *such that* $g + ij \notin G$, $(Y_i(g+ij), Y_j(g+ij)) = (Y_i(g), Y_j(g))$ *or* $Y_i(g+ij) < Y_i(g)$ *or* $Y_j(g+ij) < Y_j(g)$,

(ib) $\forall ij \in g$ *such that* $g - ij \notin G$, $Y_i(g-ij) \leq Y_i(g)$ *and* $Y_j(g-ij) \leq Y_j(g)$,

(ii) $\forall\ g' \in \mathbb{G} \setminus G\ M(g') \bigcap G \neq \emptyset$,

(iii) $\forall G' \subsetneq G$ *at least one of conditions (ia), (ib), (ii) is violated by* G'.

Simply put, a set of networks G is PWMS if (i) all possible myopic pairwise deviations from any network $g \in G$ to a network outside the set are deterred by the threat of ending worse off or equally well off, (ii) there exists a myopic improving path from any network outside the set leading to some network in the set, and (iii) there is no proper subset of G satisfying conditions (i) and (ii).

The definition of the *pairwise farsightedly stable set* of networks, or PWFS (Herings et al., 2009), corresponds to the one of a pairwise myopically stable set with myopic deviations and myopic improving paths replaced by farsighted deviations and farsighted improving paths. A farsighted improving path is a sequence of networks, where in each network a player or players making a change may not gain immediately but they improve their payoff in the final network. Namely, a *farsighted improving path* from a network g to a network $g' \neq g$ is a finite sequence of networks $g_1, .., g_K$ with $g_1 = g$ and $g_K = g'$ such that for any $1 \leq k \leq K-1$ either

(i) $g_{k+1} = g_k - ij$ for some ij such that $Y_i(g_K) > Y_i(g_k)$ or $Y_j(g_K) > Y_j(g_k)$, or

(ii) $g_{k+1} = g_k + ij$ for some ij such that $Y_i(g_K) > Y_i(g_k)$ and $Y_j(g_K) \geq Y_j(g_k)$.

If there exists a farsighted improving path from g to g', this is denoted by $g \rightarrow g'$, and for a given network g, $F(g) = \{g' \in \mathbb{G} | g \rightarrow g'\}$. Using this notation, Herings et al. (2009) defines a pairwise farsightedly stable set of networks as follows.

Definition 4 *A set of networks* $G \subseteq \mathbb{G}$ *is pairwise farsightedly stable if*

(i) $\forall\ g \in G$,

(ia) $\forall ij \notin g$ *such that* $g + ij \notin G$, $\exists g' \in F(g+ij) \bigcap G$ *such that* $(Y_i(g'), Y_j(g')) = (Y_i(g), Y_j(g))$ *or* $Y_i(g') < Y_i(g)$ *or* $Y_j(g') < Y_j(g)$,

(ib) $\forall ij \in g$ *such that* $g - ij \notin G$, $\exists g', g'' \in F(g-ij) \bigcap G$ *such that* $Y_i(g') \leq Y_i(g)$ *and* $Y_j(g'') \leq Y_j(g)$,

(ii) $\forall\ g' \in \mathbb{G} \setminus G\ F(g') \bigcap G \neq \emptyset$,

(iii) $\forall G' \subsetneq G$ *at least one of conditions (ia), (ib), (ii) is violated by* G'.

Intuitively, and following Herings et al. (2009) on p. 532, a set of networks G is PWFS if (i) all possible pairwise deviations from any network $g \in G$ to a network outside G are deterred by a credible threat of ending worse off or equally well off, (ii) there exists a farsighted improving path from any network outside the set leading to some network in the set, and (iii) there is no proper subset of G satisfying conditions (i) and (ii). Applying this definition to the two network formation games that we implement in our experiment, we obtain multiple predictions. In particular, in Game I, apart from the PWFS set that includes all 1-link networks and the complete network, there are a number of PWFS sets that contain 2-link networks. For example, $G = \{g_4, g_5, g_7\}$ is PWFS because (i) all external pairwise deviations from any network in G are deterred by a possibility of returning to the starting network, (ii) from the empty and from each of the 1-link networks there exists a short farsighted improving path to either g_4 or g_5, and from the 2-link network g_6 there exists a one-step improving path to g_7, (iii) no proper subset of G satisfies (i) and (ii). The fact that 1-link networks are not stable, while 2-link networks are stable according to some of the PWFS predictions is counterintuitive. Indeed, even though there exists an improving path from any 1-link network to one of the 2-link networks in G, farsighted individuals should foresee that the process of network formation will not stop there, as from each of the 2-link networks another simple improving deviation leads to the complete network, where the payoff of every player is 22 rather than 30.

Another pair of farsighted stability concepts discussed by Herings et al. (2009) are the *von Neumann-Morgenstern pairwise farsightedly stable set*, or vN-MFS, and the *largest pairwise consistent set*, or LPWC. They are based on the original definition of the von Neumann-Morgenstern stable set (von Neumann and Morgenstern, 1944) and the largest consistent set (Chwe, 1994).

Definition 5 *A set of networks* $G \subseteq \mathbb{G}$ *is von Neumann-Morgenstern pairwise farsightedly stable if (i)* $\forall\ g \in G\ F(g) \bigcap G = \emptyset$ *and (ii)* $\forall\ g' \in \mathbb{G} \setminus G\ F(g') \bigcap G \neq \emptyset$.

Simply put, a set of networks G is vN-MFS if no farsighted improving path exists between any pair of networks in G, and from any network outside the set there is a farsighted improving path leading to some network in G.

The largest pairwise consistent set contains any *pairwise consistent set*. Here, rather than define the pairwise consistent set, we introduce the LPWC set directly via the iterative procedure that is commonly used to construct it.[21] The set is given by the intersection of sets $\{Z_k\}_{k\geq 1}$, where each Z^k ($k = 1, 2, ...$) is inductively defined as follows: let $Z^0 \equiv \mathbb{G}$ and $g \in Z^{k-1}$ belongs to Z^k with respect to Y if

(ia) $\forall ij \notin g \; \exists g' \in Z^{k-1}$, where $g' = g + ij$ or $g' \in F(g + ij)$ such that $(Y_i(g'), Y_j(g')) = (Y_i(g), Y_j(g))$ or $Y_i(g') < Y_i(g)$ or $Y_j(g') < Y_j(g)$,

(ib) $\forall ij \in g \; \exists g', g'' \in Z^{k-1}$, where $g' = g - ij$ or $g' \in F(g - ij)$, and $g'' = g - ij$ or $g'' \in F(g - ij)$, such that $Y_i(g') \leq Y_i(g)$ and $Y_j(g'') \leq Y_j(g)$.

The resulting LPWC set requires that both external and internal pairwise deviations are deterred. It assumes that players are sufficiently cautious and irrespective of whether they are in the network within or outside the stable set, consider all possible improving paths that might be followed after a deviation. Applying the above procedure to our experimental games, we find that in Game II it identifies all, apart from the empty network, as LPWC. Intuitively the reason why 1-link networks and the complete network are stable is the same as explained in the main text (Section 2). Furthermore, 2-link networks are stable because (ia) adding a link in a 2-link network reduces the payoffs of both involved players (or leaves them unchanged if the same link is deleted again), and (ib) deleting a link in a 2-link network may lead – via a certain farsighted improving path from 1-link to 2-link to another 1-link network – to the reduction of the initially deviating player's payoff ($6 < 18$). However, that specific improving path from 1-link to 2-link to another 1-link network requires that the intermediate network payoffs do not matter to players (and that when a link is added, only one player must *strictly* improve her final payoff, and the other only weakly). Indeed, a player with payoff 30 in a 1-link network who adds a link at the first step of that improving path obtains payoff 18 in the intermediate, 2-link network before regaining 30 in another 1-link network (after deleting the second link). We note that the same path is *not* improving according to alternative cautious farsighted stability concepts – CPS and FCN – which assume that payoffs in intermediate networks matter to players at least marginally (CPS) or that links can only be added when *both* players can strictly improve their payoff in the final network (FCN).

We now define these two alternative cautious farsighted stability concepts: *cautious path stable set*,

[21] This procedure was originally proposed by Chwe (1994) and is described in Herings et al. (2009) on p.539.

or CPS (Teteryatnikova, 2015), and the *set of farsightedly consistent networks*, or FCN (Page Jr et al., 2005). The former requires first defining *path payoffs* and *surely improving path*.

For any player i a path payoff is a function that specifies the payoff $\pi_i(P)$ that player i obtains on any path, or sequence of networks P. For example, this could be a simple arithmetic average of payoffs in all networks of the sequence, or a sum with exponential discounting. The path payoff function allows defining an improving path of networks as a sequence where every link is added or deleted based on the improvement that the *remainder of the path* offers to player(s) relative to staying in the status quo network for the same number of steps. Moreover, when players are sufficiently cautious, they change links in the network according to not just an improving but a surely improving path. An improving path is called surely improving relative to a (stable) set G if whenever a link is added or deleted, the involved player(s) prefer *any* improving path that starts after the link is added or deleted and leads to a network in G to staying in the current network for the same number of steps.[22] Then the CPS set is defined as follows.

Definition 6 *A set of networks $G \in \mathbb{G}$ is cautious path stable if (i) $\forall\ g \in G$ there does not exist a surely improving path relative to G that leads to $G \setminus \{g\}$; (ii) $\forall\ g' \in \mathbb{G} \setminus G$ there exists a surely improving path relative to G that leads to G; (iii) $\forall\ G' \subsetneq G$ at least one of conditions (1), (2) is violated by G'.*

So, similar to the vN-MFS set, but considering path payoffs instead of final-network payoffs and surely improving paths instead of "simple" improving paths, the CPS set G is the minimal set that satisfies two conditions: (i) there does *not* exist a surely improving path relative to G between any pair of networks in the set, and (ii) there exists a surely improving path relative to G from any network outside the set leading to some network in the set. Clearly, predictions of CPS depend on the exact specification of the path payoff function. In the two experimental games of this paper the predictions are calculated under the assumption that players' path payoffs are defined by the arithmetic average of payoffs in all networks of the path. The details of deriving these predictions are provided in Teteryatnikova (2015).

Finally, due to the heavy use of new notation in the original definition of farsightedly consistent networks, FCN, here we provide only a semi-formal definition and refer the reader to Page Jr et al. (2005) for details.

Definition 7 *A set of networks $G \in \mathbb{G}$ is farsightedly consistent if given any network $g \in G$ and any deviation to network $g_1 \in \mathbb{G}$ by coalition S (via adding or deleting links) there exist further deviations*

[22] See Teteryatnikova (2015) for formal definitions.

(by a finite sequence of coalitions) leading to some network $g_2 \in G$ where the initially deviating coalition S is not better off and possibly worse off.

This definition employs the assumption that the coalition is better off at network g_2 than at network g_1 if and only if payoffs of all members of the coalition are strictly higher in g_2. In particular, with the pairwise approach to network formation, when a link is added by a coalition of two players, both players must strictly improve their payoff in a final network, and when a link is deleted by one player, this player's payoff must strictly improve.

B Experimental results

	Empty	One-link	Two-link	Complete
Game 1	0.33	0.68	0.45	0.92
Matching group 1	0.33	0.91	0.51	0.99
Matching group 2	0.27	0.66	0.44	0.83
Matching group 3	0.37	0.71	0.51	0.89
Matching group 4	0.28	0.91	0.40	0.98
Matching group 5	0.43	0.40	0.47	0.88
Matching group 6	0.37	0.70	0.46	0.98
Matching group 7	0.10	0.44	0.47	0.93
Matching group 8	0.33	0.55	0.38	0.65
Game 2	0.29	0.89	0.52	0.79
Matching group 9	0.40	0.92	0.51	0.42
Matching group 10	0.22	0.88	0.62	0.39
Matching group 11	0.31	0.91	0.58	0.51
Matching group 12	0.30	0.98	0.53	0.54
Matching group 13	0.37	0.77	0.49	0.86
Matching group 14	0.15	0.89	0.54	0.92
Matching group 15	0.22	0.89	0.59	0.79
Matching group 16	0.27	0.75	0.34	0.75

Table 4: Probability of remaining in current network

C Instructions

[ONSCREEN]

Before the experiment begins there will be a short tutorial and three practice games to make sure everybody understands how points can be earned.

The points that are distributed in these three practice rounds will not affect your final payment.

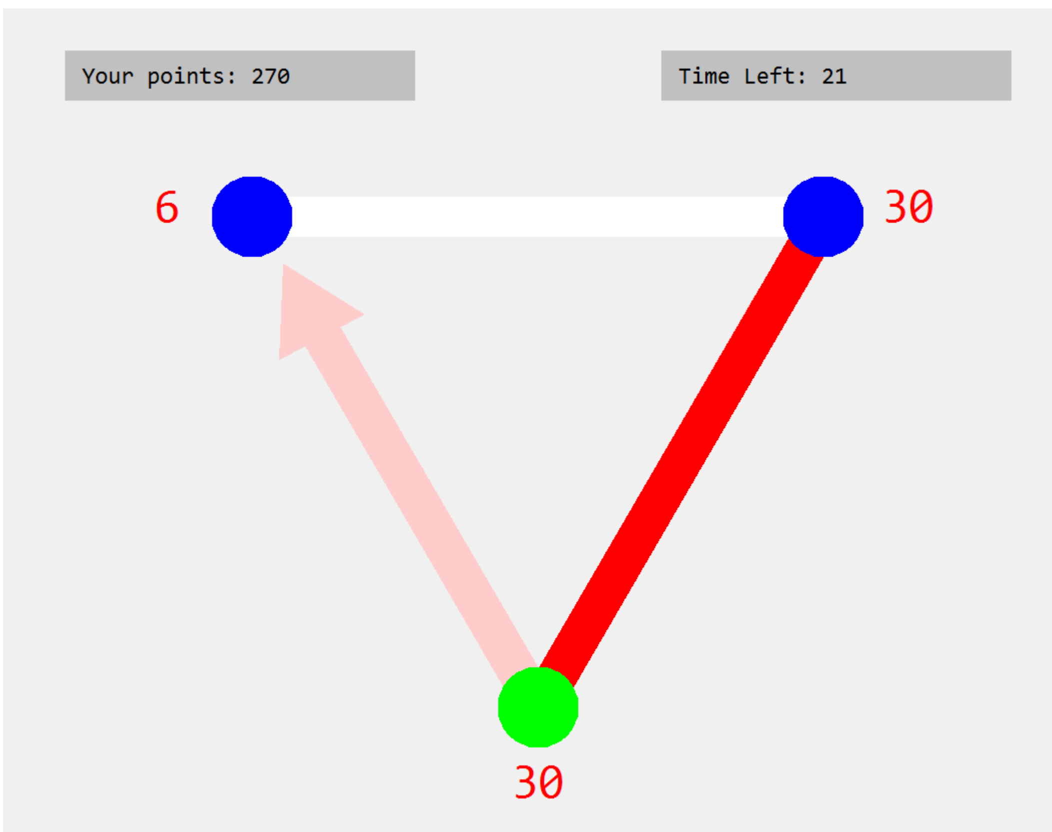
Please click "Continue" to proceed to the tutorial.

[PRINTED]

Please read and follow these instructions. Text in italics describes things you should do onscreen. If you have a question, raise your hand and someone will come to help you as soon as possible.

- In each round of this experiment you will be interacting with two other people using the screen you can see on your monitor.
- You are represented by the green dot, and the other two players by the blue dots.
- Links may be formed between two players in the following way:
  - You can indicate that you are willing to form a link with another person by clicking on their blue dot.
  - Clicking on them again indicates you are no longer willing to form a link with them.
  - You can click on a person as many times as you like, switching back and forth between being willing to form a link with them or not.
  - If two people have both clicked on each other then a link is formed and it is shown in red.
  - A link can be formed between to people only if both of them want it to be formed.
  - If only one of the people has shown they are willing to form a link then it is shown in pink.
- On the screen in front of you the two other people have formed a link, and the person on your right has indicated they are willing to form a link with you.
  - Click on the blue dot on the left and see how the line turns pink. Click again and see how it becomes white again.
  - Click on the blue dot on the right and see how the line turns red. Click again and see how it becomes pink again.
  - Notice that nothing you can do will change the colour of the line between the other two people. Whether or not that link is formed depends only on their decisions.

- Every second, you and the two other people will earn points. The number of points per second earned by each person is shown in red next two their dot.
- The number of points each person earns per second depends on which links are formed at that point in time.
- These numbers will vary from round to round and will be shown to you before each round begins. The numbers for the screen you see in front of you are described in the following diagram:
- Click on the two other people and see how the numbers in red change, and how they relate to the diagram. Notice that it doesn't make a difference if a line is white or pink; the numbers change only if a link is formed or broken (i.e. becomes red, or changes from red to pink).
- In the practice rounds and real interactions the screen will look slightly different. An example is shown below:



- As mentioned before, points will be earned every second. The total number of points you have earned so far in a round will be shown at the top left of the screen as shown in the picture "Your Point"). This number will increase every second by the red number below the green dot.
- Each round lasts for 30 seconds. The number of seconds left will shown at the top right.

- WHEN YOU HAVE UNDERSTOOD THESE INSTRUCTIONS, PLEASE CLICK THE BUTTON ON YOUR COMPUTER SCREEN.

[ONSCREEN]

Please answer the following questions relating to the picture shown in the handout.

Click "Continue" when you have answered all questions.

How many points are you earning per second?

How many points is the person on your left earning per second?

How many points have you earned so far this round?

How many seconds are left before the round ends?

[New Screen]

You have answered all the questions correctly.

Before the real experiment begins there will be three practice rounds.

These practice rounds will not affect your final payment.

The purpose of these practice rounds is for you to learn how these interactions work. You should use them to experiment and learn how links are formed and how they relate to the payoffs. Do not worry about earning a lot of points because they do not count!

The points associated with each practice round are shown on your handout.

If you have any questions, please raise your hand and someone will come to help you as soon as possible.

Otherwise please click "Continue" and wait for the other participants to finish the Tutorial.

The first practice round is about to begin.

Check the payoffs described in Figure 1 of your printed instructions. These are the payoffs that are relevant for the practice rounds.

Click OK when you are ready to begin.

[New Screen]

The first practice round is about to begin.

Check the payoffs described in Figure 1 of your printed instructions. These are the payoffs that are relevant for the practice rounds.

[New Screen]

Click OK when you are ready to begin.

The practice rounds are now over.

You will now be handed the diagrams which describe the payoffs for the first real rounds.

[New Screen]

You will now play a game similar to the one in the tutorial but with different payoffs.

Please look at the diagram you have just been given to see how the points you earn will depend on the links that are formed.

You will play this game 20 times. The links that have been already formed at the beginning of the game will be randomly determined each time.

After each time you will be randomly rematched with new participants. This means it is unlikely you will be playing with exactly the same people as in the previous round.

When all games have been played, one game will be randomly chosen to determine how much you will be paid. All participants will be paid for the same game. For every 90 points you earn in that game you will be paid 2 Euros.

When you are ready to start, please click "Continue".