Paris School of Economics

Université Paris-1 Panthéon-Sorbonne

Essays on Network Formation and Farsighted Behavior

A doctoral thesis written for the purpose

of obtaining a doctoral degree in economics by

Amit Dekel

December 14, 2023

Thesis Directors:	Francis Bloch
	Margherita Comola
Jury members:	Ana Mauleon
	Dominik Karos
	Marco Mantovani
	Frédéric Moisan

Summary

I study network formation and farsighted behavior theoretically and empirically using both observational and experimental data. The first chapter develops a structural estimation method for measuring network externalities. The second chapter develops a solution concept aiming to capture farsighted behavior in abstract games, which encompass network formation games as a special case. The third chapter tests the predictive power of various myopic and farsighted solution concepts in the context of a network formation game played in the lab.

Keywords: Social and Economic Networks, Game Theory, Experimental Economics, Econometric theory.

Résumé

J'étudie la formation des réseaux et le comportement prévoyant de manière théorique et empirique en utilisant des données d'observation et expérimentales. Le premier chapitre développe une méthode d'estimation structurelle pour mesurer les externalités de réseau. Le deuxième chapitre développe un concept de solution visant à capturer le comportement prévoyant dans les jeux abstraits, qui englobent les jeux de formation de réseaux comme un cas particulier. Le troisième chapitre teste le pouvoir prédictif de divers concepts de solution myopes et clairvoyants dans le contexte d'un jeu de formation de réseau joué en laboratoire.

Mots-clés: Réseaux sociaux et économiques, Théorie des jeux, Économie expérimentale, Théorie économétrique.

Acknowledgements

I've accumulated a great debt of gratitude over the past few years.

Dotan, much of this journey is due to you. Had I not crossed paths with you back when I was an undergraduate student, I probably would not have been drawn to research as much as I am. I also would not have been exposed to the field of networks economics, nor would I have discovered my affection (and respect) to economic theory and mathematical coding. Your constant support over the years has been invaluable to me.

Margherita, the origins of this thesis are rooted in your research. My Master thesis, which served as the starting point of the current thesis, was initially inspired by one of your working papers. Thank you for writing it. Your wisdom, advice and support have been of incredible help during these years.

Francis, thinking back on the evolution of this thesis, it seems that things could have gone very differently had you not introduced me to the literature on farsighted coalition formation during the "advanced game theory" course. Thank you for that. Your guidance was indispensable in helping me clarify, refine, and distill the contents of this thesis. I've learned a lot from you, and hope to keep doing so in the future.

Hillel, without making it look as if pursuing a Master's degree at PSE wasn't beyond reach, I might have ended up not even trying. Thank you for pushing me back then, and throughout.

Angelo, thank you for hosting me at Johns Hopkins. That was a wonderful experience.

Philipp, Liza, Agnieszka, thank you for participating in my thesis committees. You played a major role in helping me distinguish good ideas from bad ones.

Umberto, Kartik, Sonia, I am greatful for having shared this journey with you. Hopefully we'll still have things to complain about as time goes by.

Dose café, Tryst, Emissary, Back in Black, Le Barbouquin, Le Grand Breguet, Ten Belles, thank you for providing the work environment I needed, and for allowing me to meet Maria – the best coffee companion ever.

Mom, dad, Tom, Ofri, thank you for unconditionally supporting whatever it is I was doing all along.

Nina, words cannot describe. I love you so.

Contents

Introduction (in English)

Overview

In economics, a “social network” is defined by a set of nodes that represent agents (e.g. individuals, firms, states) and a set of links that represent connections between them (e.g. friendship, trade, diplomatic relations). The field of “network formation” studies the question “what determines networks’ structures?”. At the level of the individual agent, this question can be formulated as “what determines agents’ decisions regarding who to links with?”. The (somewhat cynical) approach taken by economists is to try to provide an answer in terms of a cost-benefit analysis: agent i chooses to link with agent j if the benefit she obtains from it outweighs her cost of maintaining it. Benefits may include elements such as emotional support (e.g. in friendship networks), information (e.g. in gossip networks) or increased national security (e.g. in diplomatic-relations networks). Costs may include elements such as emotional resources, time or effort. Those costs and benefits are aggregated in agents’ “utility functions”, which assign a level of utility to every linking decision profile.

The assumption underlying models of *strategic* network formation is that the benefit an agent i derives from a link with some other agent j depends on the existence of other links in the network. For instance, it may depend on the number of j’s links, or on whether or not i already maintains a link with some other agent k. These dependencies are termed “network externalities”. Chapter 1 deals with the question “do network externalities indeed play a role in determining agents’ linking decisions?”. The challenge in answering this question is that the existence of network externalities renders links interdependent, whereas statistical inference always requires some form of independence. The chapter develops an estimation technique that allows drawing inferences on the existence and magnitude of network externalities given a dataset containing information on a single network.

A description of a strategic network formation model, and indeed of any game-theoretic model, is like the introduction of tale, presenting the heroes, their interests and the setting in which they operate.[1] In a network formation game, this introduction may include elements like the agents’ utility functions and the conditions under which

[1] The equating of an economic model with a tale is borrowed from Rubinstein (2012).

they may form or delete links. An array of rules by which the model is "allowed" to develop from its beginning to its end is called a **solution concept**. The "end" of a network formation model, for instance, takes the form of a network structure. A solution concept applied to a network formation model yeilds an answer to the question "given the model's description, what network structure is expected to arise?".

Many solution concepts exist and most models can be analyzed by more than one. Thus, a natural question to ask once a model is fully described is "which solution concept should be applied to it?". The crucial aspect in which solution concepts differ from one another is the set of assumptions they reflect. The answer to the above question therefore depends on the set of assumptions that the modeler wishes to adopt. Chapter 2 deals with solution concepts reflecting the assumption that players are farsighted, i.e. that they take into account the entire chain of reactions their own actions might trigger. The chapter formulates several critiques on existing solution concepts attempting to capture farsighted behavior and proposes a new solution concept that is immune to them.

Since different solution concepts can be applied to the same model, it is often constructive to fix the game defined by the model and consider alternative "endings"\predictions produced by alternative solution concepts. Such games can sometimes be (approximately) implemented in a laboratory setting, making it possible to compare the theoretical predictions produced by various solution concepts against the one obtained when human subjects play the game. This allows answering the question "which solution concept is most predictive of actual human behavior?". Chapter 3 attempts to answer this question in the context of a network formation game. The theoretical predictions it considers are produced by the following three categories of solution concepts: *(i)* myopic solution concepts (where players are assumed to consider only the immediate consequences of their action and ignore any subsequent reactions they might trigger); *(ii)* the farsighted solution concepts that are criticized in Chapter 2; *(iii)* the new solution concept proposed in Chapter 2.

Summary of Chapter 1

The ample theoretical literature on strategic network formation presupposes the existence of network externalities, i.e. that the benefits agents derive from links depend on the existence of links elsewhere in the network. Empirical evidence for the existence (and magnitude) of such externalities is, however, sparse. This is in part because externalities render links interdependent, while causal inference always requires independence in some form or another. Chapter 1 proposes a method to consistently estimate (certain types of) network externalities using data on a single network. This allows answering questions that are hard to settle theoretically, such as "do agents prefer their partners to maintain many links, or few?". On the one hand, many indirect connections may provide benefits such as more information flow, better reach etc. On the other hand, when a potential partner maintains many links, she is likely to have less time to devote to each of her partners, implying connections of lower quality.[2]

The general estimation approach builds on Leung (2015) and relies on specifying a network formation model with incomplete information while maintaining three key assumptions: *(i)* **separability**: agent i's marginal utility from a link with j is independent of i's other links; *(ii)* **iid shocks**: the random components in marginal utilities are independent of one another; *(iii)* **symmetric BNE**: the observed data is generated from a symmetric Bayes Nash Equilibrium. Note that since the model is of incomplete information, agents are assumed to make linking decisions based on their *beliefs* about the emerging network. The first two assumptions guarantee that *conditional on beliefs* all linking decisions are independent. The assumption that agents play a BNE means that their beliefs about the linking probability of each pair are correct. The assumption that the BNE played in the data is *symmetric* allows estimating these beliefs from the data. Hence, the method boils down to a two-step procedure where beliefs are estimated on the first and the parameters in the utility function are estimated on the second.

The procedure proposed in this chapter departs from Leung (2015) in one important way: while in his model links are assumed to be directed, in this chapter links are assumed to be undirected. While in the directed case agents' decisions are simply interpreted as decisions to link, in the undirected case they are interpreted as "proposals to link". Hence, a rule that transforms "proposals" into undircted links is required. We consider a bilateral rule and a unilateral rule. In the bilateral rule an undirected link is assumed to exist if and only if both sides propose to one another.[3] In the unilateral rule, an undirected link is assumed to exist if and only if at least one side proposes to the other. These rules give pairs of players the *joint* power to form\delete links between them, which is somewhat incompatible with the fully non-cooperative nature

[2]These are the two intuitions underlying the canonical "connections model" and "co-authors model". See Jackson and Wolinsky (1996).

[3]This idea can be traced back to Myerson (1991)'s link announcement game.

of the BNE solution concept. We show that this can be solved by assuming that the BNE played in the data is admissible, meaning that it does not prescribe the use of weakly dominated strategies.

After showing that the proposed estimator is consistent and asymptotically normal, we apply it to real-world data. The dataset we use contains self-declared information on risk-sharing arrangements in the Tanzanian village of Nyakatoke, from which we draw the undirected village network. We then investigate whether agents choose between risk-sharing partners on the basis of their characteristics alone or whether indirect connections (i.e. friends of friends) also play a role in these decisions. Our estimates suggest that for a given pair of potential partners i and j, the probability that i proposes a link to j increases on average by 0.016 for any additional indirect connection j provides, which corresponds to approximately 9% of the average fitted probability of link proposal. This provides evidence for the existence of positive externalities from indirect connections.

Summary of Chapter 2

If an economic model is a tale, a solution concept is what brings about its ending. Solution concepts take as input the introduction of the tale, which includes the agents' interests and the setting in which they operate, and produce as output the ending of the tale, more commonly referred to as a "prediction". Solution concepts differ in the assumptions they reflect and in the range of models on which they can be applied. Chapter 2 proposes a new solution concept that reflects the assumption that agents are perfectly farsighted (i.e. take into account the entire chain of reaction that their own actions might trigger), and that can be applied to a wide range of models.

The set of models that the proposed solution can be applied to are those that can be described as "abstract games" (sometimes referred to as "games in effectivity function form"). An abstract game is defined by set of states, players' preferences over these states, and an effectivity correspondence specifying which subsets of players ("coalitions") are allowed to move from one state to another. Models that can be described as abstract games include games in characteristic function form, games in partition function form, strategic form games, extensive form games (with perfect information), network formation games, voting games, matching games and others. Given that some of these games belong to the cooperative branch of game theory and others to the non-cooperative one, abstract games can be viewed as a unifying framework that breaks the traditional dichotomy between the two.

The motivation for the development of this new solution concept stems from two critiques on existing solutions. The first relates to reliance on the notion of "farsighted improving paths" (a.k.a. "farsighted objections", or "indirect dominance"), initially due to Harsanyi (1974).[4] A *farsighted improving path* is a finite sequence of states and coalitions $\{z^0, S^1, z^1, ..., S^K, z^K\}$ such that for all $1 \leq k \leq K$: *(i)* the coalition S^k has the ability to replace state z^{k-1} by state z^k; and, *(ii)* all players in S^k prefer the final state in the sequence z^K over the status quo state z^{k-1}. It is attractive to deploy this notion in order to describe farsighted behavior because it assumes players make decisions based on the *final* state that will be reached z^K. Its drawback, however, is that the final state z^K is compared against the *status quo* z^{k-1}, rather than against the final state of some alternative continuation path that could take place *had* S^k *decided to remain at the status quo* z^{k-1}. Following Chwe (1994) and Karos and Robles (2021), we refer to this critique as the *counterfactual critique*.

The second critique relates to the (recently deployed) "rational expectations" approach, initially due to Jordan (2006).[5] Under this approach, players are assumed to hold endogenous expectations about the continuation path that would follow each state. The main benefit of following this approach is that it allows incorporating a "maximality condition" which ensures that players make moves that are optimal for them, rather than just improving.[6]. A drawback common to all existing solutions adopting this approach, however, is that the expected continuation path from each state is assumed to be unique. This means that the uncertainty about the order of play embedded in the definition of abstract games, i.e. the fact that no such order is defined (even stochastically), is ignored. Following Granot and Hanany (2022), we refer to this critique as the *overconfidence critique*.

The solution concepts we propose, entitled *Set-Valued Rational Expectations* (SVRE), is immune to these critiques. Our solution departs from existing solutions within the rational expectations approach in one crucial aspect: it allows expectations to be *set-*

[4]Examples of solution concepts for abstract games relying on the notion of farsighted improving paths include, for instance, the Farsighted Stable Set, the Largest Consistent Set (Chwe, 1994) and the Rational Expectations Farsighted Stable Set (Dutta and Vohra, 2017).

[5]Examples of solution concepts for abstract games utilizing the rational expectations approach can be found in Dutta and Vohra (2017), Dutta and Vartiainen (2020), Bloch and van den Nouweland (2020), Kimya (2020), Karos and Robles (2021).

[6]This allows tackling the "maximality critique". See Dutta and Vohra (2017) for more details.

valued. A *set-valued expectation*, generically denoted by m, is a subset of all possible moves. Moves that are included in the set-valued expectation are interpreted as moves that are *intended* to be executed, if ever a chance arises.

Fixing a set-valued expectation m, a state is said to be *stationary under* m if no move away from it is included in m. For every state z, we use $Y(z, m)$ to denote the set of stationary states under m that are reachable from z via moves in m. A move from z to z' is said to *eventually lead to* $Y(z', m)$. Preferences over moves are determined based on comparisons of the sets of states they eventually lead to. Given these preferences, roughly speaking, a set-valued expectation m is said to be *rational* (SVRE) if: *(i)* it is dynamically consistent, i.e. all players can commit to the prescriptions of m; *(ii)* it is optimal, i.e. no coalition can deviate to an alternative m' such that the moves it intends to execute under m' are preferred over those that it intends to execute under m.

Our baseline results show that the SVRE concept generalizes some well-established solution concepts. In particular, we show that when players are restricted to consider only one step ahead (i.e. are myopic), the SVRE concept coincides with the core of an abstract game, which in turn, depending on how the effectivity correspondence is defined, can be shown to coincide with Nash, strong Nash, pairwise stability, pairwise-Nash, stable matching, Condorcet winner, and others. When expectations are restricted to contain only one move away from each state, the SVRE concept boils down to standard (single-valued) expectation functions satisfying the conditions proposed by Ray and Vohra (2019). Our general results include sufficient conditions for existence, uniqueness, and absorption. When considering applications to specific classes of games we find the following. In perfect information extensive form games the SVRE concept boils down to subgame perfection. In strategic form games, a state (i.e. an actions profile) is supported as stationary by some SVRE if and only if it is Pareto efficient. In partition function form games, a state is supported as stationary by some SVRE if and only if it is immune to myopically beneficial deviaitons of coalitions that include either all players, or all players but one.

Summary of Chapter 3

Game-theoretic solution concepts differ from one another in the set of assumptions they reflect. Some reflect myopic behavior in the sense that players are assumed to take into account only the immediate consequences of their actions (e.g. the Core). Others reflect farsighted behavior in the sense that players are assumed to take into account the long-term consequences of their actions (e.g. the Farsighted Stable Set). Yet others assume players take into account not only the long-term consequences of their actions, but also of *lack of actions* on their part (e.g. the Set-Valued Rational Expectations solution proposed in Chapter 2). Chapter 3 tests the predictive power of each of these categories of solution concepts in the context of a network formation lab experiment.

The experiment is composed of two parts. In the first, participants perform a cognitive task based on the "Hit 15" game. This provides a measure of participants ability to reason farsightedly in a non-strategic environment and allows assigning each a "farsighted reasoning score". In the second, participants play a dynamic network formation game in groups of four. The game starts at the empty network and proceeds by sequentially offering randomly selected group members to form\delete links with other group members. Following any such offer, all group members are presented with the new state of the network (and associated payoffs) and are asked whether they want to stop the formation process at the current network. If they all say YES, the game ends and the current network is declared "final". Otherwise another group member is randomly offered to form\delete links with other group members. The payoffs associated with network structures are designed so that solutions concepts in each of the three categories above produce mutually exclusive predictions.

Each experimental session is randomly assigned a matching treatment and an information treatment. In the "random" matching treatment groups are composed at random, while in the "homogeneous" matching treatment groups are composed to minimize the variance of their members' farsighted reasoning scores (as measured by the questionnaire administered in the first part of the experiment). In the "private" information treatment farsighted reasoning scores remains private information, while in the "public" information treatment they are publicly disclosed to all group members. These treatments allow examining how are the observed dynamics of play and final outcomes affected by groups' compositions in terms of the farsighted reasoning scores of their constituent members, as well as the disclosure of information on others' farsighted reasoning scores.

We find that in 87% of the rounds played in the lab groups converge to the predictions made by the third category of solution concepts. This lends strong empirical support to solution concepts belonging to this category, which include the "Set-Valued

Rational Expectations" solution proposed in Chapter 2. The interpretation is that participants not only take into account others' reactions to their actions, but also to lack of action on their part. This in turn pushes them towards taking actions that cannot be rationalized but on preemtive grounds, i.e. that are only sensible insofar as they prevent others from taking certain actions.

With respect to the level of farsighted reasoning scores, we find that they are strongly negatively correlated with taking myopically rational actions. This suggests that the devised Hit-15 questionairre is a useful tool for measuring the ability to reason farsightedly. In addition, we find that participants with higher farsighted reasoning scores tend to insist harder on ahcieving the predictions made by the third category, characterized by being Pareto efficient. This suggests that high-score individuals are less prone to coordination failures. There is limited evidence that their insisitence on acheiving Pareto efficiency is further amplified by the provision of information on others' scores. With respect to the dispersion of farsighted reasoning scores, we find that groups with low dispersion comprised of high-score individuals tend to converge to a final (Pareto efficient) outcome faster than groups with high dispersion. This further illustrates high-score individuals' strategic competence.

Introduction (en français)

Aperçu Général

En économie, un "réseau social" est défini par un ensemble de nœuds qui représentent des agents (par exemple, des individus, des entreprises, des états) et un ensemble de liens qui représentent des connexions entre les nœuds (par exemple, l'amitié, le commerce, les relations diplomatiques). Le domaine de la "formation des réseaux" étudie la question suivante : "' Comment les structures des réseaux sont-elles déterminées ?" Au niveau de l'agent individuel, cette question peut être formulée comme suit : " Comment les agents décident-ils avec qui ils vont tisser des liens ?" L'approche (quelque peu cynique) adoptée par les économistes consiste à tenter d'apporter une réponse en termes d'analyse coût-bénéfice : l'agent i choisit d'établir un lien avec l'agent j si les avantages qu'il en retire l'emportent sur le coût qu'il doit assumer pour l'entretenir. Les avantages peuvent inclure des éléments tels que le soutien émotionnel (par exemple dans les réseaux d'amitié), l'information (par exemple dans les réseaux de bavardage) ou le renforcement de la sécurité nationale (par exemple dans les réseaux de relations diplomatiques). Les coûts peuvent inclure des éléments tels que des ressources émotionnelles, du temps ou des efforts. Ces coûts et avantages sont agrégés dans les "fonctions d'utilité" des agents, qui attribuent un niveau d'utilité à chaque profil de décision de liaison.

L'hypothèse qui sous-tend les modèles de formation de réseaux *stratégiques* consiste à dire que l'avantage qu'un agent i tire d'un lien avec un autre agent j dépend des autres liens du réseau. Par exemple, cet avantage peut dépendre du nombre de liens de j, ou du fait que i entretient déjà un lien avec un autre agent k. Ces dépendances sont appelées "externalités de réseaux". Le chapitre 1 traite de la question suivante : "Les externalités de réseaux jouent-elles un rôle dans les décisions des agents en matière de création de liens ?" Le défi à relever pour répondre à cette question est que l'existence d'externalités de réseau rend les liens interdépendants, alors que l'inférence statistique exige toujours une certaine forme d'indépendance. Ce chapitre développe une technique d'estimation qui permet, à partir de données provenants d'un seul réseau, de tirer des conclusions sur l'existence et l'ampleur des externalités de réseaux.

La description d'un modèle de formation de réseaux stratégiques, et par ailleurs de tout modèle de théorie des jeux, s'apparente à l'introduction d'un conte, présentant les héros, leurs intérêts et le cadre dans lequel ils évoluent.[7] Dans un jeu de formation de réseau, cette introduction peut inclure des éléments tels que les fonctions d'utilité des agents et les conditions dans lesquelles ils peuvent former ou supprimer des liens. Un ensemble de règles permettant au modèle de se développer du début à la fin est appelé **concept de solution**. La "fin" d'un modèle de formation de réseau, par exemple, prend la forme d'une structure de réseau. Un concept de solution appliqué à un modèle de formation de réseau apporte une réponse à la question suivante : "Compte tenu de la description du modèle, quelle structure de réseau devrait apparaître ?"

Il existe de nombreux concepts de solution et la plupart des modèles peuvent être analysés par plusieurs d'entre eux. Ainsi, une fois qu'un modèle est entièrement décrit, il est naturel de se demander "quel concept de solution devrait lui être appliqué". L'aspect crucial par lequel les concepts de solution diffèrent les uns des autres est l'ensemble des hypothèses qu'ils reflètent. La réponse à la question ci-dessus dépend donc de l'ensemble des hypothèses que le modélisateur souhaite adopter. Le chapitre 2 traite des concepts de solution reflétant l'hypothèse selon laquelle les joueurs sont prévoyants, c'est-à-dire selon laquelle ils prennent en compte l'ensemble de la chaîne de réactions que leurs propres actions pourraient déclencher. Ce chapitre formule plusieurs critiques sur les concepts de solution existants qui tentent d'appréhender le comportement prévoyant et propose un nouveau concept de solution qui est à l'abri de ces critiques.

Étant donné que différents concepts de solution peuvent être appliqués au même modèle, il est souvent constructif de fixer le jeu défini par le modèle et d'envisager d'autres "fins", c'est-à-dire des prédictions produites par d'autres concepts de solution. De tels jeux peuvent parfois être (approximativement) mis en œuvre en laboratoire, ce qui permet de comparer les prédictions théoriques produites par divers concepts de solution à celles obtenues lorsque des sujets humains jouent au jeu. Cela permet de répondre à la question suivante : "Quel est le concept de solution qui prédit le mieux le comportement humain réel ?" Le chapitre 3 tente de répondre à cette question dans le contexte d'un jeu de formation de réseau. Les prédictions théoriques qu'il examine sont produites par les trois catégories suivantes de concepts de solution : *(i)* les concepts de solution myopes (dans lesquels les joueurs sont supposés ne prendre en compte que les conséquences immédiates de leur action et ignorer les réactions ultérieures qu'ils pourraient déclencher) ; *(ii)* les concepts de solution clairvoyants qui sont critiqués dans le chapitre 2 ; *(iii)* le nouveau concept de solution proposé dans le chapitre 2.

[7] L'assimilation d'un modèle économique à un conte est empruntée à Rubinstein (2012).

Résumé du Chapitre 1

L'abondante littérature théorique sur la formation de réseaux stratégiques présuppose l'existence d'externalités de réseaux, c'est-à-dire que les avantages que les agents tirent de leurs liens dépendent de l'existence de liens ailleurs dans le réseau. Les preuves empiriques de l'existence (et de l'ampleur) de ces externalités sont toutefois rares. Cela s'explique en partie par le fait que les externalités rendent les liens interdépendants, alors que l'inférence causale nécessite toujours une certaine forme d'indépendance. Le chapitre 1 propose une méthode pour estimer de manière cohérente (certains types) d'externalités de réseau en utilisant des données sur un seul réseau. Cela permet notamment de répondre à des questions difficiles à résoudre sur le plan théorique. Par exemple, "les agents préfèrent-ils que leurs partenaires entretiennent de nombreux liens ou peu de liens ?". Théoriquement, les deux réponses peuvent être justifiées. D'une part, de nombreux liens indirects peuvent présenter des avantages tels qu'un plus grand flux d'informations, une meilleure portée, etc. D'autre part, lorsqu'un partenaire potentiel entretient de nombreux liens, il est probable qu'il ait moins de temps à consacrer à chacun de ses partenaires, ce qui implique des connexions de moindre qualité. [8]

L'approche générale s'appuie sur Leung (2015) et repose sur la spécification d'un modèle de formation de réseau avec des informations incomplètes tout en maintenant trois hypothèses clés : *(i)* **séparabilité** : l'utilité marginale que retire l'agent i d'un lien avec j est indépendante des autres liens de i ; *(ii)* **iid shocks** : les composantes aléatoires des utilités marginales sont indépendantes les unes des autres ; *(iii)* **EBN symétrique** : les données observées sont générées à partir d'un équilibre de Bayes Nash symétrique. Il convient de noter qu'étant donné que le modèle est fondé sur des informations incomplètes, les agents sont supposés se comporter en fonction de leurs *croyances* sur le réseau émergeant. Les deux premières hypothèses garantissent que *conditionnellement aux croyances* toutes les décisions de liaison sont indépendantes. L'hypothèse selon laquelle les agents jouent un EBN signifie que leurs croyances concernant la probabilité de liaison de chaque paire sont correctes. L'hypothèse selon laquelle l'EBN joué dans les données est *symétrique* permet d'estimer ces croyances à partir des données. Par conséquent, la méthode se résume à une procédure en deux étapes où les croyances sont estimées dans un premier temps et les paramètres de la fonction d'utilité dans un second temps.

[8] Ce sont les deux intuitions qui sous-tendent le "connections model" et le "co-authors model" canoniques. Voir Jackson and Wolinsky (1996).

La procédure proposée dans ce chapitre se distingue de celle de Leung (2015) sur un point important : alors que dans son modèle les liens sont supposés être dirigés, dans ce chapitre les liens sont supposés être non dirigés. Cela soulève immédiatement la question de la règle à utiliser pour transformer les décisions des agents en liens non dirigés. Nous traitons cette question en considérant à la fois une règle bilatérale et une règle unilatérale. Dans la règle bilatérale, les décisions des agents sont interprétées comme des "propositions de liens" et un lien non dirigé est présumé exister si et seulement si les deux parties se proposent l'une à l'autre de créer un lien.[9] Dans la règle unilatérale, un lien non dirigé est supposé exister si et seulement si au moins une des parties fait une proposition à l'autre. Ces règles donnent aux paires d'acteurs le pouvoir *commun* de créer ou de supprimer des liens entre eux, mais cela est quelque peu incompatible avec la nature totalement non coopérative du concept de solution de l'EBN. Nous montrons que ce problème peut être résolu en supposant que l'EBN joué dans les données est admissible.

Enfin, nous appliquons la procédure d'estimation proposée à des données réelles sur le réseau de partage de risques dans le village tanzanien de Nyakatoke. L'ensemble de données que nous utilisons contient des informations détaillées et auto-déclarées sur les accords de partage de risques dans ce village. Nous utilisons ces données pour dessiner le réseau villageois non dirigé et pour déterminer si les agents choisissent leurs partenaires de partage de risques uniquement en fonction de leurs caractéristiques ou si les relations indirectes (c'est-à-dire les amis des amis) jouent également un rôle dans ces décisions. Nos estimations suggèrent que pour une paire de partenaires potentiels i et j, la probabilité que i propose un lien à j augmente en moyenne de $0,016$ pour chaque connexion indirecte supplémentaire fournie par j, ce qui correspond à environ 9% de la probabilité moyenne de la probabilité moyenne de proposition de lien. Cela atteste de l'existence d'externalités positives dues aux connexions indirectes.

Résumé du Chapitre 2

Si un modèle économique est un conte, un concept de solution est ce qui permet de le conclure. Les concepts de solution prennent en entrée l'introduction du conte, qui comprend les intérêts des agents et le cadre dans lequel ils opèrent, et produisent en sortie la fin du conte, plus communément appelée "prédiction". Les concepts de solution

[9] Cette idée peut être retracée jusqu'au jeu d'annonce de liens de Myerson (1991).

diffèrent dans les hypothèses qu'ils reflètent et dans la gamme de modèles sur lesquels ils peuvent être appliqués. Le chapitre 2 propose un nouveau concept de solution qui reflète l'hypothèse selon laquelle les agents sont parfaitement prévoyants (c'est-à-dire qu'ils prennent en compte l'ensemble de la chaîne de réactions que leurs propres actions pourraient déclencher), et qui peut être appliqué à un large éventail de modèles.

L'ensemble des modèles auxquels la solution proposée peut être appliquée sont ceux qui peuvent être décrits comme des "jeux abstraits" (parfois appelés "games in effectivity function form"). Un jeu abstrait est défini par un ensemble d'états, les préférences des joueurs sur ces états et une correspondance d'effectivité spécifiant quels sous-ensembles de joueurs (les "coalitions") sont autorisés à passer d'un état à l'autre. Les modèles qui peuvent être décrits comme des jeux abstraits comprennent les types de jeux suivants: "games in characteristic function form", "games in partition function form", "strategic form games", les jeux en forme extensive (avec une information parfaite), les jeux de formation de réseaux, les jeux de vote, les jeux de matching et d'autres encore. Étant donné que certains de ces jeux appartiennent à la branche coopérative de la théorie des jeux et d'autres à la branche non coopérative, les jeux abstraits peuvent être considérés comme un cadre unificateur qui rompt la dichotomie traditionnelle entre les deux.

La motivation pour le développement d'un nouveau concept de solution provient de deux critiques sur les solutions existantes. La première concerne le recours à la notion de "farsighted improving paths" (chemin d'amélioration clairvoyant, ou "farsighted objections", ou "dominance indirecte"), initialement due à l'équilibre de **?**. Parmi les concepts de solution pour les jeux abstraits reposant sur la notion de "farsighted improving paths" figurent, par exemple, le Farsighted Stable Set, le Largest Consistent Set (Chwe, 1994) et le Rational Expectations Farsighted Stable Set (Dutta and Vohra, 2017). Un "chemin d'amélioration prévoyant" est une séquence finie d'états et de coalitions $\{z^0, S^1, z^1, ..., S^K, z^K\}$ telle que pour tout $1 \leq k \leq K$, *(i)* la coalition S^k a la capacité de remplacer l'état z^{k-1} par l'état z^k : *(i)* la coalition S^k a la capacité de remplacer l'état z^{k-1} par l'état z^k ; et, *(ii)* tous les joueurs de S^k préfèrent l'état final de la séquence z^K à l'état de statu quo z^{k-1}. Il est intéressant d'utiliser cette notion pour décrire un comportement prévoyant, car elle suppose que les joueurs prennent des décisions en fonction de l'état *final* qui sera atteint z^K. Son inconvénient, cependant, est que l'état final z^K est comparé au *statu quo* z^{k-1}, plutôt qu'à l'état final d'une trajectoire de continuation alternative qui pourrait avoir lieu *si* S^k *avait décidé de rester au statu quo* z^{k-1}. Comme Chwe (1994) et Karos and Robles (2021), nous appelons cette critique *contrefactuelle*.

La deuxième critique concerne l'approche (récente) des "attentes rationnelles", initialement due à Jordan (2006). [10] Dans le cadre de cette approche, les joueurs sont supposés avoir des attentes endogènes quant à la trajectoire de continuation qui suivrait chaque état. Le principal avantage de cette approche est qu'elle permet d'incorporer une "condition de maximalité" qui garantit que les joueurs effectuent des mouvements qui sont optimaux pour eux, plutôt que des mouvements qui constituent simplement une amélioration.[11]. Un inconvénient commun à tous les concepts de solution existants adoptant cette approche, cependant, est que le chemin de continuation attendu à partir de chaque état est supposé être unique. Cela signifie que l'incertitude sur l'ordre de jeu intégrée dans la définition des jeux abstraits, c'est-à-dire le fait qu'un tel ordre n'est pas défini (même stochastiquement), est ignorée. Suivant Granot and Hanany (2022), nous nous référons à cette critique comme *critique de surconfiance*.

Nous proposons un concept de solution qui est à l'abri de ces critiques : le *Set-Valued Rational Expectations* (SVRE). Notre solution s'écarte des solutions existantes qui adoptent l'approche des anticipations rationnelles ("rational epectations") sur un point crucial : elle permet aux anticipations d'être *set-valued*. Une telle anticipation m (un équilibre potentiel) est un sous-ensemble de tous les coups possibles. Les coups inclus dans le "set-valued expectation" sont donc ceux que les joueurs ont textitl'intention d'exécuter si l'occasion se présente.

Fixons une "set-valued expectation" m. Un état est dit *stationnaire sous* m si aucun déplacement à partir de cet état n'est inclus dans m. Pour chaque état z, $Y(z, m)$ désigne l'ensemble des états stationnaires sous m qui sont atteignables à partir de z via des déplacements dans m. Un déplacement de z vers z' est dit *conduire éventuellement* à $Y(z', m)$. Les préférences sur les mouvements sont déterminées sur la base de comparaisons des ensembles d'états auxquels ils aboutissent. Compte tenu de ces préférences, une "set-valued expectation" m est *rationnelle* (SVRE) si : *(i)* elle est dynamiquement cohérente, c'est-à-dire que tous les joueurs peuvent s'engager à respecter les prescriptions de m ; *(ii)* elle est optimale, c'est-à-dire qu'aucune coalition ne peut dévier vers une alternative m' de sorte que les mouvements qu'elle a l'intention d'exécuter sous m' sont préférés à ceux qu'elle a l'intention d'exécuter sous m.

[10] Des exemples de concepts de solutions pour "jeux abstraits" utilisant l'approche des attentes rationnelles figurent dans Dutta and Vohra (2017), Dutta and Vartiainen (2020), Bloch and van den Nouweland (2020), Kimya (2020), Karos and Robles (2021).

[11] Ceci permet de s'attaquer à la "critique de maximalité" (voir Dutta and Vohra (2017) pour plus de détails).

Nos résultats de base montrent que le concept SVRE généralise certains concepts de solution bien établis. En particulier, nous montrons que lorsque les joueurs sont restreints à ne considérer qu'un pas en avant (autrement dit, sont myopes), le concept SVRE coïncide avec le noyau d'un jeu abstrait (qui à son tour, selon la façon dont la correspondance d'effectivité est définie, peut être montré comme coïncidant avec Nash, Strong Nash, pairwise stability, pairwise-Nash, stable matching, le vainqueur de Condorcet, et d'autres encore). En outre, nous montrons que lorsqu'on restreint les anticipations à une seule déviation de chaque état, le concept d'SVRE se résume à des fonctions d'anticipations standards (à valeur unique) satisfaisant aux conditions proposées par **?**. Nos résultats généraux comprennent des conditions suffisantes pour l'existence, l'unicité et l'absorption. Lorsque nous considérons les applications à des classes de jeux spécifiques, nous constatons ce qui suit. Dans les jeux de forme extensive à information parfaite, le concept SVRE se résume à l'équilibre parfait du sous-jeu. Dans les jeux à forme stratégique, un état (c'est-à-dire un profil d'actions) est considéré comme stationnaire par un SVRE si et seulement s'il est Pareto efficace. Dans les jeux de forme "partition function", un état est considéré comme stationnaire par un SVRE si et seulement s'il est immunisé contre les déviances myopiquement bénéfiques des coalitions qui incluent soit tous les joueurs, soit tous les joueurs sauf un.

Résumé du Chapitre 3

Les concepts de solution de la théorie des jeux diffèrent les uns des autres par l'ensemble des hypothèses qu'ils reflètent. Certains reflètent un comportement myope dans le sens où les joueurs sont supposés ne prendre en compte que les conséquences immédiates de leurs actions (par exemple, le noyau ("Core")). D'autres reflètent un comportement prévoyant dans le sens où les joueurs sont supposés prendre en compte les conséquences à long terme de leurs actions (par exemple, le Farsighted Stable Set). D'autres encore supposent que les joueurs tiennent compte non seulement des conséquences à long terme de leurs actions, mais aussi de l'absence d'actions de leur part (par exemple, le Set-Valued Rational Expectations défini au chapitre 2). Le chapitre 3 teste le pouvoir prédictif de chacune de ces catégories de concepts de solution dans le contexte d'une expérience de formation de réseau en laboratoire.

L'expérience se compose de deux parties. Dans la première, les participants effectuent une tâche cognitive basée sur le jeu "Hit 15". Cette tâche permet de mesurer la capacité des participants à raisonner avec clairvoyance dans un environnement non stratégique et d'attribuer à chacun un "score de raisonnement clairvoyant". Dans le second, les participants jouent à un jeu de formation de réseau dynamique par groupes de quatre. Le jeu commence par un réseau vide et se poursuit en proposant de manière séquentielle à des membres du groupe choisis au hasard de former ou de supprimer des liens dans lesquels ils sont impliqués. Après chaque proposition, tous les membres du groupe sont informés du nouvel état du réseau (et des gains associés) et il leur est demandé s'ils souhaitent arrêter le processus de formation au niveau du réseau actuel. S'ils répondent tous OUI, le jeu se termine et le réseau actuel est déclaré "final". Dans le cas contraire, un autre membre du groupe est choisi au hasard et se voit proposé de changer le statut des liens dans lesquels il est impliqué, et ainsi de suite. Les gains associés à chaque réseau sont conçus de manière à ce que les concepts de solutions dans chacune des trois catégories ci-dessus produisent des prédictions mutuellement exclusives.

Chaque session expérimentale se voit attribuer de manière aléatoire un traitement de matching et un traitement d'information. Dans le traitement de matching "aléatoire", les groupes sont composés au hasard, tandis que dans le traitement de matching "homogène", les groupes sont composés de manière à minimiser la variance des scores de raisonnement clairvoyant de leurs membres (tels que mesurés par le questionnaire administré dans la première partie de l'expérience). Dans le traitement de l'information "privée", les scores de raisonnement clairvoyant restent confidentiels, tandis que dans le traitement de l'information "publique", ils sont divulgués publiquement à tous les membres du groupe. Ces traitements permettent d'examiner comment la dynamique observée du jeu et les résultats finaux sont affectés par la composition des groupes en termes de scores de raisonnement clairvoyant de leurs membres constitutifs, ainsi que par la divulgation d'informations sur les scores de raisonnement clairvoyant des autres.

Nous constatons que dans 87% des tours joués en laboratoire, les groupes convergent vers les prédictions de la troisième catégorie de concepts de solution. Ceci constitue une base empirique solide pour les concepts de solution appartenant à cette catégorie, qui incluent la solution de "Set-Valued Rational Expectations" proposée au chapitre 2. L'interprétation est que les participants ne prennent pas seulement en compte les réactions des autres à leurs actions, mais aussi à l'absence d'action de leur part. Cela les pousse à entreprendre des actions qui ne peuvent être rationalisées que sur des bases prééminentes, c'est-à-dire qui ne font sens que dans la mesure où elles empêchent les autres d'entreprendre certaines actions.

En ce qui concerne le niveau des scores de raisonnement clairvoyant, nous constatons qu'ils sont fortement corrélés négativement avec le fait d'entreprendre des actions myopiquement rationnelles. Cela suggère que le questionnaire développé autour de Hit-15 est un outil utile pour mesurer la capacité à raisonner avec clairvoyance. En outre, nous constatons que les participants ayant des scores élevés en matière de raisonnement clairvoyant ont tendance à insister davantage sur la réalisation des prédictions de la troisième catégorie. Les preuves que cette insistance est encore amplifiée par la fourniture d'informations sur les scores des autres participants sont limitées. En ce qui concerne la dispersion des scores de raisonnement clairvoyant, nous constatons que plus celle-ci est faible, plus les groupes convergent rapidement vers un résultat final. Une faible dispersion semble donc être liée à la mesure dans laquelle les membres du groupe sont d'accord sur l'ensemble des structures du réseau qui sont acceptables comme points d'arrêt du processus de formation du réseau.

Chapter 1

Estimating Network Externalities in Undirected Link Formation Games

(joint with Margherita Comola[1])

Abstract

This paper explores the existence of externalities from network architecture (so-called network externalities) in link formation games of incomplete information. It extends the structural estimation method by Leung (2015) to games where links are undirected and proposals are only partially observable. We provide an econometric characterization of the proposed two-step estimator, and we document its performance through a simulation exercise. When the estimation method is applied to data on risk-sharing arrangements in a Tanzanian village, results indicate that indirect connections matter. Assuming that link formation follows a bilateral process, the estimated probability of proposing a link to a potential partner increases by 9% for any additional indirect connection provided.

Keywords: Undirected Networks, Network Externalities, Incomplete Information, Risk-sharing.

JEL codes: C45, D85, O12.

[1]University Paris-Saclay and Paris School of Economics.

1.1 Introduction

From its very first steps network theory has claimed that the formation of links may depend strategically on the entire graph (Jackson and Wolinsky, 1996; Bala and Goyal, 2000). However, evidence-based on experimental and observational data still lags behind, and empirical questions about the value of indirect connections in real-life situations remain largely unexplored.[2] This paper proposes an estimation procedure for undirected link formation games where link proposals are not fully observed and depend on network architecture. In particular, we investigate whether agents choose potential partners on the basis of their individual characteristics only, or whether indirect connections also play a role in these decisions. To answer this question, we develop a simple estimation protocol that extends Leung (2015) to the class of undirected network formation models. In our setting, agents play a simultaneous game of incomplete information where they form undirected links on the basis of their beliefs. Assuming that these beliefs satisfy a number of regularity conditions (discussed in Section 1.2), the estimation strategy boils down to a two-step procedure where the first stage consistently estimates agents' beliefs about the emerging network, and the second stage estimates the role of network externalities.[3] This procedure is flexible enough to accommodate both bilateral and unilateral link formation rules. We provide existence, consistency and asymptotic normality results for the two-step estimator, and we conduct a comprehensive set of simulation exercises to investigate its performance as sample size grows.

We illustrate the procedure using data on risk-sharing arrangements from the Tanzanian village of Nyakatoke. Lacking access to formal insurance, most households in developing countries rely on informal risk-sharing arrangements in face of shocks such as health-related expenses, injuries, funerals and job losses. These arrangements have long captured the attention of economists. On the one hand, the prevalence of the phenomenon makes it of paramount importance for economic development.[4] On the other hand, most arrangements do not take place at the level of the entire community

[2] Most of the available evidence relates to specific settings. For instance, the study of cross-firm collaborative networks suggests that information flows are insignificant for indirect neighbors (Breschi and Lissoni, 2005; Singh, 2005). On the other hand, experimental evidence with dictator games shows that further-away connections are relevant and decay with the inverse of distance (Goeree et al., 2010). Graham and Pelican (2019) provide a test for interdependencies in link-formation preferences and conclude for the presence of externalities in the same data we use here.

[3] A two-step approach is also taken by König et al. (2019).

[4] Coate and Ravallion (1993), Townsend (1994), Udry (1994), Fafchamps and Lund (2003).

but among pairs of households.[5] By aggregating all declared links, we obtain a graph which is among the most compelling applications of networks in economics.[6]

We use the self-declared information in Nyakatoke data to draw the undirected village network and to investigate the role of network architecture. Specifically, we test whether agents choose between risk-sharing partners on the basis of their individual characteristics only or whether indirect connections also play a role in these decisions. Much of the economic literature assumes that informal risk-sharing arrangements require the consent of the two parties involved, which implies that link formation follows a bilateral process.[7] Following this literature, in the empirical illustration of Section 1.5 we assume that the links are formed bilaterally. Appendix 1.8 shows how our estimation strategy also accommodates undirected networks issued by unilateral link formation rules. Risk-sharing arrangements constitute an intriguing setting where network externalities may combine positive and negative components: indirect connections are beneficial if they broaden social interactions but detrimental if there is competition for scarce resources. Results from Section 5 indicate that Nyakatoke villagers do evaluate potential partners' connections and that the positive component prevails. Our estimates suggest that for a given pair of potential partners ij, the probability that i proposes a link to j increases on average by 0.016 for any additional indirect connection j provides. This increase is sizeable, as it corresponds to approximately 9% of the average fitted probability of link proposal.

From an econometric standpoint, testing whether network architecture predicts link formation has proved to be a complex task. Our paper deals with the case where the researcher observes one single network at one single period and wants to include network covariates in the objective function of agents. In this scenario the structural equation can have multiple solutions (Bjorn and Vuong, 1984; Bresnahan and Reiss, 1991; Tamer, 2003), and the calculation may become intractable due to the combinatorial complexity of networks. One solution is provided by the exponential random graph models where a dynamic meeting protocol acts as an equilibrium selection mechanism (Hsieh and

[5] Alongside risk-sharing links, other types of financial arrangements in small groups (such as rotating savings and credit associations) have also been documented.

[6] Risk-sharing networks have been studied from multiple angles, including the efficiency and sustainability of the resulting arrangements, the determinants of link formation and the structural properties of the network architecture (Genicot and Ray, 2003; Bramoullé and Kranton, 2007; Bloch et al., 2008; Jackson et al., 2012; Banerjee et al., 2013; Ambrus et al., 2014; Ambrus and Elliott, 2020).

[7] Most models of risk sharing and favor exchange assume that agents can refuse transactions that are against their self-interest (Kimball, 1988; Coate and Ravallion, 1993; Kocherlakota, 1996; Bloch et al., 2008; Jackson et al., 2012).

Lee, 2016; König, 2016; Mele, 2017; Badev, 2020). Another solution is to condition on models that replicate some observed topological patterns or to limit the degree to which other players can affect one's utility.[8] Alternatively, one can simplify the estimation procedure by relying on incomplete information to induce symmetry and independence in agents' strategies (Leung, 2015; De Paula and Tang, 2012), which is the approach we take here.

This paper's main contribution is methodological: it develops an estimation protocol for a class of undirected link formation games with network externalities under the assumption of incomplete information. Our approach builds on Leung (2015) who also relies on incomplete information to estimate a simultaneous game of link formation. Our paper differs in one substantive aspect, however: while Leung (2015)'s procedure requires data on directed links, which are interpreted as observed proposals in a game of unilateral link formation, our's is suited for undirected link data, which we interpret as the equilibrium outcome of a link formation process where proposals are only partially observed. This in turn requires a new equilibrium selection rule that eliminates coordination failures (Section 1.2.2). Our work also relates to Ridder and Sheng (2020), who generalize Leung (2015) by relaxing the separability assumption to include additional non-linear network externalities. Their work, however, remains in the realm of directed networks.[9] As an additional contribution, our paper also advances the knowledge of risk-sharing arrangements in developing countries by providing first-hand evidence that indirect connections affect linking choices, while previous literature has focused mostly on documenting the number and characteristics of risk-sharing partners.[10]

Network formation models have proved difficult to estimate in presence of externalities because of multiplicity of equilibria. Most of the existing tools were developed for directed networks and expect two distinct reports per dyad (Leung, 2015; Mele,

[8] Some papers identify structural parameters by aggregating individuals into 'types' and assuming that agents have preferences only over the type of their partners (De Paula et al., 2018), or by the rate at which various sub-graphs are observed in the overall network (Chandrasekhar and Jackson, 2016). Along similar lines, Boucher and Mourifie (2017) study a setting where individual preferences display weak homophily.

[9] They do consider a scenario where the links' directionality is obscured due to issues of data collection and\or missreporting, but the underlying formation process remains directional in nature. This is conceptually different from our bilateral setting where undirected links are issued by mutual consent. For other papers estimating social interaction models under incomplete information see Gilleskie and Zhang (2009) and Hoshino (2019).

[10] An exception is Krishnan and Sciubba (2009), who identify the common features of all equilibrium configurations in a model with negative network externalities and test these predictions against data on labor exchange arrangements in Ethiopia.

2017; Badev, 2020). On the other hand, the available models of undirected network formation rely on complete information and achieve set identification (Miyauchi, 2016; Sheng, 2020; De Paula et al., 2018). The procedure we propose is computationally parsimonious, providing a convenient alternative to complete-information models. As such it can prove useful in a variety of applications where links are undirected by nature.

The paper is organized as follows. Section 1.2 introduces the theoretical setting. Section 1.3 presents the estimation method. Section 1.4 describes a simulation exercise. Section 1.5 applies the estimation method to risk-sharing data from rural Tanzania. Section 1.6 concludes. Appendix 1.7 discusses the inclusion of continuous attributes and the smoothing of discrete variables. Appendix 1.8 draws a comparison between our model and models of unilateral and\or directed link formation. Appendix 1.9 presents a second application that uses the same dataset as in Leung (2015), allowing for a direct comparison with his results. All proofs are relegated to Appendix 1.10.

1.2 The Model

In what follows we describe our model under the assumption that link formation follows a bilateral rule. The extension to unilateral link formation is discussed in Appendix 1.8.

1.2.1 The game

Let $N = \{1, 2, ..., n\}$ be a set of agents who play to form an undirected network. For agent i, let $X_i = [X_{i,1}, ..., X_{i,q}]$ be a vector of individual attributes of dimension $[1 \times q]$ and $X = \{X_1, ..., X_n\}$ denote the set of these vectors. For ease of exposition in this section we assume that X is composed of discrete attributes only (this assumption is relaxed in Appendix 1.7).

Assumption 1.1 (Discrete X)**.** *X has finite support.*

Let $\epsilon_i = [\epsilon_{i,1}, ..., \epsilon_{i,i-1}, 0, \epsilon_{i,i+1}, ..., \epsilon_{i,n}]$ be a $[1 \times n]$ vector of shocks of agent i with all other agents (ϵ_{ij} does not necessarily equal ϵ_{ji}), which are stochastically independent from X. ϵ denotes the collection ϵ_i over all $i \in N$.

Assumption 1.2 (i.i.d. Shocks). $\{\epsilon_{ij} \mid i, j \in N, i \neq j\}$ *are independently drawn from the standard normal distribution.*[11]

Thus, shocks are assumed to be uncorrelated within and across individuals.[12] The set of attributes vectors X is common knowledge, while the shocks are private information, i.e. only i knows ϵ_i.

Agents play a simultaneous-move game of link formation, where everyone announces independently the links they wish to form. Link formation follows a bilateral rule, and the resulting network is given by the mutually announced links (Myerson, 1991). The action of agent i is represented by a binary vector of length n, where the jth entry ($j \neq i$) equals 1 if i proposes j to form a link and 0 otherwise.[13] The actions of all agents stacked on top of each other, denoted S, can be interpreted as an adjacency matrix of a directed network of link proposals:

$$S = \begin{bmatrix} 0 & S_{1,2} & \dots & S_{1,n} \\ S_{2,1} & 0 & \dots & S_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n,1} & \dots & S_{n,n-1} & 0 \end{bmatrix} \tag{1.1}$$

In turn, this proposal network gives rise to an undirected network G, where a link between two agents exists if and only if both propose to each other: $G_{ij} = S_{ij} \cdot S_{ji}$.

For a given network G, the utility of agent i is given by:

$$u_i(X, G; \theta_0) = \sum_{j \neq i} G_{ij} \cdot (v_{ij}(X, G_{-i}; \theta_0) + \epsilon_{ij}) \tag{1.2}$$

where G_{-i} indicates G with the i^{th} row and column deleted, and $\theta_0 \in \Theta$ is a $[p \times 1]$ vector of parameters from a compact set Θ. Estimating the parameters in θ_0 is the goal of the procedure described in Section 1.3.

Assumption 1.3 (Linearity, Separability and Anonymity). *The* $v_{ij}(\cdot)$ *function: (i) is linear in* θ_0*; (ii) depends on* G *only through* G_{-i}*; (iii) is insensitive to a permutation of the agents' labels.*

[11] The standard normal distribution is chosen here for the sake of convenience, but our results hold for other full-support distributions.

[12] Independence within individuals serves to simplify our estimation framework, but it could be seen as implausible in some applications. A less restrictive approach is proposed by Graham (2017) who models agent-level unobserved heterogeneity as in a fixed-effect panel.

[13] Since an agent cannot form a link with herself, the ith entry always equals 0.

The separability condition, borrowed from Leung (2015), requires that the i's marginal utility from a link with j is independent from other links she may have.[14] In Section 1.2.4 below we discuss which types of externalities from indirect connections this assumption is compatible with.

1.2.2 Equilibrium

Let i's (pure) strategy be a function from commonly observed attributes and privately observed shocks to an action: $S_i : (X, \epsilon_i) \to \{0,1\}^n$ (henceforth we omit the dependency on X). A Bayes Nash Equilibrium (BNE) is a strategy profile $[S_i(\epsilon_i), S_{-i}(\epsilon_{-i})]$ such that for all $i \in N$ and for all $S_i'(\epsilon_i)$:

$$\mathbb{E}_{\epsilon_{-i}}\left[u_i(X, G[S_i(\epsilon_i), S_{-i}(\epsilon_{-i})]; \theta_0)\right] \geq \mathbb{E}_{\epsilon_{-i}}\left[u_i(X, G[S_i'(\epsilon_i), S_{-i}(\epsilon_{-i})]; \theta_0)\right] \tag{1.3}$$

Due to the separability assumption, in any BNE agents consider proposal decisions separately. Hence, we can write $S_i(X, \epsilon_i) = [S_{ij}(X, \epsilon_{ij})]_{j \in N}$, where $S_{ij} : (X, \epsilon_{ij}) \to \{0, 1\}$. In addition, in any BNE, S_{ij} must prescribe i to propose to j whenever it strictly increases her expected utility and not to propose whenever it strictly reduces it. Formally:

$$S_{ij}(\epsilon_{ij}) = \begin{cases} 1 & \text{if } \mathbb{E}_{\epsilon_{ji}}[S_{ji}(\epsilon_{ji})] \cdot \Big(\mathbb{E}_{\epsilon_{-i}}[v_{ij}(X, G_{-i}[S_{-i}(\epsilon_{-i})]; \theta_0)] + \epsilon_{ij}\Big) > 0 \\ 0 & \text{if } \mathbb{E}_{\epsilon_{ji}}[S_{ji}(\epsilon_{ji})] \cdot \Big(\mathbb{E}_{\epsilon_{-i}}[v_{ij}(X, G_{-i}[S_{-i}(\epsilon_{-i})]; \theta_0)] + \epsilon_{ij}\Big) < 0 \end{cases} \tag{1.4}$$

Whenever proposing to j does not change i's expected utility, proposing and not proposing are both best-replies. This makes it clear that Bayes Nash equilibria do not exclude coordination failures. For instance, a pair $S_{ij}(\epsilon_{ij})$ and $S_{ji}(\epsilon_{ji})$ that prescribed i and j (respectively) not to propose for any ϵ_{ij} and ϵ_{ji} (respectively) may well be part of a BNE profile, *even if* both i and j stand to gain (in expectation) from forming a link. Since we are interested in modeling bilateral network formation, where pairs of agents are free to coordinate their actions, we wish to rule out such equilibria. We do so by restricting attention to *admissible* Bayes Nash equilibria in, i.e. equilibria were no player uses a (weakly) dominated strategy. In any *admissible* BNE, S_{ij} must prescribe i to propose to j whenever, *assuming j proposes to i*, her expected utility from proposing

[14] Separability is relaxed by Ridder and Sheng (2020) in the context of directed network formation.

is strictly positive, and not to propose if it is strictly negative. Formally:

$$S_{ij}(\epsilon_{ij}) = \begin{cases} 1 & \text{if } \mathbb{E}_{\epsilon_{-i}}[v_{ij}(X, G_{-i}[S_{-i}(\epsilon_{-i})]; \theta_0)] + \epsilon_{ij} > 0 \\ 0 & \text{if } \mathbb{E}_{\epsilon_{-i}}[v_{ij}(X, G_{-i}[S_{-i}(\epsilon_{-i})]; \theta_0)] + \epsilon_{ij} < 0 \end{cases} \tag{1.5}$$

Given this decision rule, one may reformulate the equilibrium condition in terms of beliefs over proposal probabilities. To that end, let $\sigma^{S_{-i}}$ be a $[(n-1) \times n]$ matrix representing i's beliefs about the probabilities that each agent $j \neq i$ proposes to another agent $k \neq j$ (including i herself). Given the decision rule in Equation (1.5), and letting Φ denote the CDF of the standard normal distribution, the ex-ante probability that i proposes to j is:

$$Pr(S_{ij} = 1 | X, \sigma^{S_{-i}}) = Pr\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0) | X, \sigma^{S_{-i}})] + \epsilon_{ij} > 0\right) \tag{1.6}$$

$$= \Phi\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0) | X, \sigma^{S_{-i}})]\right) \tag{1.7}$$

Note that since ϵ_{ij} is drawn from a continuous distribution, is makes no difference whether i's strategy prescribes to propose or not when $\mathbb{E}_{\epsilon_{-i}}[v_{ij}(X, G_{-i}[S_{-i}(\epsilon_{-i})]; \theta_0)] + \epsilon_{ij}$ is exactly zero. A belief matrix σ^S corresponds to an admissible BNE if and only if it satisfies the following equality for all i and j:

$$\sigma_{ij}^{S_{-i}} = Pr(S_{ij} = 1 | X, \sigma^{S_{-i}}) \tag{1.8}$$

Hence, such equilibrium beliefs are an alternative description of admissible Bayes Nash equilibria. The fact that $v_{ij}(\cdot)$ depends on G_{-i}, rather than S_{-i}, allows deriving yet a simpler representation: instead of conditioning its expected value on beliefs over proposal probabilities it suffices to condition on beliefs over *linking* probabilities. To that end, let σ^G denote a $[n \times n]$ matrix representing agents' common beliefs about linking probabilities among all pairs of agents, and $\sigma^{G_{-i}}$ denote the same matrix but with its i^{th} row and column deleted. Lastly, note that the independence among ϵ_{ij} values imply that the probability that a link exists is simply the product of the proposal probabilities of the two parties involved. Thus, we can formulate the equilibrium condition in terms of beliefs over linking probabilities. A belief matrix σ^G corresponds to an admissible BNE if and only if it satisfies the condition below for all i and j. We call such σ^G an "equilibrium belief".

$$\sigma_{ij}^G = \underbrace{\Phi\left(\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}]\right)}_{Pr(i \text{ proposes to } j)} \underbrace{\Phi\left(\mathbb{E}[v_{ji}(X, G_{-j}; \theta_0)|X, \sigma^{G_{-j}}]\right)}_{Pr(j \text{ proposes to } i)} \tag{1.9}$$

Given an equilibrium belief σ^G, a network G is said to be an "equilibrium" if the following holds for all i and j:

$$G_{ij} = \underbrace{\mathbb{1}\left\{\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0)|X, \sigma^{G_{-i}}] + \epsilon_{ij} > 0\right\}}_{i \text{ proposes to } j} \underbrace{\mathbb{1}\left\{\mathbb{E}[v_{ji}(X, G_{-j}; \theta_0)|X, \sigma^{G_{-j}}] + \epsilon_{ji} > 0\right\}}_{j \text{ proposes to } i} \tag{1.10}$$

Note that due to admissibility, an equilibrium network G is one that satisfies the pairwise stability conditions *in expectation*: *(i)* if i and j are linked in G then the marginal expected utilities this link provides each is positive; *(ii)* if i and j are not linked in G then then the marginal expected utility this link provides is negative for at least one of them. Hence, even though the solution concept we deploy is non-cooperative, in equilibrium no pair of players fail to coordinate on forming a link.

Following Leung (2015), from here on we restrict attention to symmetric equilibria. A symmetric equilibrium is an equilibrium in which all pairs of agents that are observationally equivalent have the same linking probabilities. Formally, an equilibrium σ^G is symmetric if for all $i, j \neq k, l \in N$:

$$(X_i = X_k \text{ } and \text{ } X_j = X_l) \text{ } or \text{ } (X_i = X_l \text{ } and \text{ } X_j = X_k) \implies \sigma_{ij}^G = \sigma_{kl}^G \tag{1.11}$$

Figure 1.1 illustrates this definition. Agents in this network have a single binary attribute – being either black or white – depicted by the colors of the nodes. Beliefs are depicted by weights on edges and their values by their color (i.e. all red beliefs equal each other, and all blue beliefs equal each other). All pairs consisting of two black agents have the same σ^G value (red), and the same holds for pairs of white and black agents (blue) and pairs of two white agents (green). The described beliefs are therefore symmetric.

For given X and θ_0, we let $\omega(X, \theta_0)$ denote the set of symmetric equilibria. The following proposition establishes that $\omega(X, \theta_0)$ is non-empty.

Proposition 1.1 (Existence). *Under assumptions 1.1-1.3, there exists a symmetric*

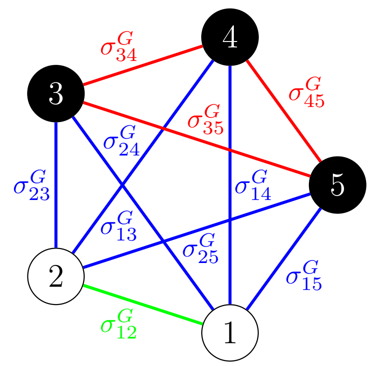


Figure 1.1: Example of a symmetric belief matrix

equilibrium, i.e. $\omega(X, \theta_0) \neq \emptyset$.

Assumption 1.4 (Symmetric Equilibrium)**.** *The observed network is generated according to Equation (1.10) where* $\sigma^G \in \omega(X, \theta_0)$.

Note that Assumption 1.4 does not impose any restrictions on the probability that a given symmetric equilibrium is selected. This stands in contrast to the "many markets asymptotics" setting where the econometrician observes many repetitions of the game and assumes that the probability distribution over (not necessarily symmetric) equilibria is degenerate. As a result, the equilibrium being played in all repetitions of the game is guaranteed to be the same one. Following Leung (2015), we are able to avoid this assumption and achieve point identification with one large network ("large market asymptotics") by allowing only symmetric equilibria to be selected.

1.2.3 Example

Consider the case where 3 agents have one binary attribute X_i, and their utility function is as follows:

$$v_{ij}(X, G_{-i}; \theta_0) = \theta_1 + \theta_2 X_i + \theta_3 |X_i - X_j| + \theta_4 \frac{1}{n-1} \sum_{k \neq i} G_{jk} \tag{1.12}$$

with $\theta_0 = [-1, 1, -0.5, 1]'$. The term $|X_i - X_j|$ represents a measure of similarity between i and j. It thus accounts for homophily. The term $\frac{1}{n-1} \sum_{k \neq i} G_{jk}$ represents the average number of indirect connections (i.e. paths of length 2) that i gains by forming a link with j. It thus accounts for externalities from the network topology.

Columns 1 and 2 in Table 1.1 present all possible ordered pairs in the 3-agent network. Columns 3 and 4 report the binary attributes of agents i and j respectively.

Column 5 reports $|X_i - X_j|$. The third term in the utility function $\frac{1}{n-1}\sum_{k\neq i} G_{jk}$ depends on the network structure G. Its expected value therefore depends on the beliefs about the network structure σ^G.

Let us consider a given set of beliefs which are reported in column 6. Column 7 uses these beliefs to compute $\frac{1}{n-1}\sum_{k\neq i} \sigma_{jk}^{G_{-i}}$. Using columns 3, 5 and 7 and the functional form we can now compute the expected value of v_{ij} for all pairs of agents. This is reported in column 8. Now, given that the ϵ_{ij} values are drawn independently from the standard normal distribution, the probability that i would propose to j (that is, that $\mathbb{E}[v_{ij}] + \epsilon_{ij} \geq 0$) is $\Phi(\mathbb{E}[v_{ij}])$. This is reported in column 9. Finally, the probability that a link exists in G is the product of the proposal probabilities of the two agents involved. This is reported in column 10.

1	2	3	4	5	6	7	8	9	10
i	j	X_i	X_j	$\|X_i - X_j\|$	σ^G	$\frac{1}{n-1}\sum_{k\neq i} \sigma_{jk}^{G_{-i}}$	$\mathbb{E}[v_{ij}]$	$\Phi(\mathbb{E}[v_{ij}])$	$\Phi(\mathbb{E}[v_{ij}]) \cdot \Phi(\mathbb{E}[v_{ji}])$
1	2	0	1	1	0.027	$0.5 \cdot 0.255$	-1.3725	0.0850	0.027
1	3	0	1	1	0.027	$0.5 \cdot 0.255$	-1.3725	0.0850	0.027
2	1	1	0	1	0.027	$0.5 \cdot 0.027$	-0.4865	0.3133	0.027
2	3	1	1	0	0.255	$0.5 \cdot 0.027$	0.0135	0.5054	0.255
3	1	1	0	1	0.027	$0.5 \cdot 0.027$	-0.4865	0.3133	0.027
3	2	1	1	0	0.255	$0.5 \cdot 0.027$	0.0135	0.5054	0.255

Table 1.1: Example

Note that in this example $\sigma_{ij}^G = \Phi(\mathbb{E}[v_{ij}])\Phi(\mathbb{E}[v_{ji}])$ for all i and $j \neq i$. This means that the beliefs σ^G in column 6 are equilibrium beliefs. Also note that all pairs of agents which are observationally equivalent have the same linking probabilities, e.g. the pairs $\{1, 2\}$ and $\{1, 3\}$ have the same linking probability under σ^G. This means that the beliefs σ^G are symmetric.

1.2.4 Separability and Externalities

The utility agents gain from the network might be related to different measures of their centrality in it. The assumptions we take on the form of agents' utility function, however, limits the type of centrality measures whose effect on proposal decision can be estimated. This subsection discusses what centrality measures are compatible with our assumptions.

Let $c_i(G)$ denote a generic centrality measure of player i in network G. The separability assumption requires it can be written in the form $c_i(G) = \sum_{j\neq i} G_{ij} \cdot f(G_{-i})$ for some function f. This condition can alternatively be written as $c_i(G+ij)-c_i(G-ij) = f(G_{-i})$, where $G+ij$ (respectively, $G-ij$) denote the network G with the link between i and j added (respectively, removed). Hence, our model allows for the marginal contribution of a link ij to i's centrality to be a function of all walks in G besides those that pass through i. Centrality measures compatible with this rule include "targeting centrality" "targeting centrality" (Bramoullé and Genicot, 2023) and "information centrality" (Stephenson and Zelen, 1989).[15]

Information centrality assign weights for every path emanating from i and sum those weights up. Since paths are walks are sequences of agents and links in which no agent appears twice, this measure is compatible with the separability assumption. While information centrality defines a spefic weighting scheme, one could generelize it by leaving the weighting scheme open. The externalities we use thoughout this paper is a special case where the weights on every paths of length larger then three are set to zero.

To give the intuition behing taregting centrality, consider a dynamic process of information diffusion that takes place in discrete time. At time period $l = 0$ an agent i passes a message to each of her friends with some fixed probability p. At every subsequent period $l > 1$ any agent that received the message at period $l-1$ passes it to each of her friends with probability p. Now suppose that the message is targeted towards a specific agent j. j's taregting centrality measures the expected number of times she receives messages from others assuming she does not participate in the diffusion process (i.e. she never retrasmits the message). The idea that she does not retransmit the message makes this centrality measure compatible with our separability assumption.

While the discussion above presents centrality measures that are compatible with the separability assumption, some are clearly not. The following equation provides a generic way to construct a separable counterpart for *any* centrality measure $c_i(G)$: $\tilde{c}_i(G) = \sum_{j\neq i} G_{ij} \cdot c_i(G_{-i} + ij)$. As an illistration, suppose $c_i(G)$ denote i's diffusion centrality (Banerjee et al., 2013), which is based on the same information diffusion process described above. The interpretation of $\tilde{c}_i(G)$ is that i diffuses the message in period 1 and then never retransmits it again.

[15] Brandes and Fleischer (2005) show that information centrality is equivalent to current-flow closeness centrality.

1.3 Estimation

Imagine we observe a single network G and agents' attributes X. Let us assume that G is formed according to the model specified above, that is, the network results from all agents behaving optimally given the symmetric equilibrium belief σ^G and their realization of the error terms ϵ_i that we do not observe.[16] Our goal is to estimate and conduct inference on the true parameter vector θ_0. In what follows we describe the building blocks of our procedure.

1.3.1 Log-likelihood function

Let us denote by δ_{ij} a function that takes X_i, X_j and returns a vector of covariates of dimension $[1 \times (p-k)]$ (e.g. i's attributes and the distance between i and j's attributes, in the example above). Denote by γ_{ij} a function that takes i's beliefs about the emerging network (possibly together with X) and returns a vector of covariates of dimension $[1 \times k]$ (e.g. the number of length-two paths i gains from linking with j, in the example above). To facilitate an intercept, assume that δ_{ij} always returns 1 as a first element. We call the first type of covariates 'exogenous' as they do not depend on the network structure, and the second type 'endogenous', as they do. Using this terminology, while $\gamma_{ij}(X, G_{-i})$ represents the endogenous covariates associated with i's linking with j, $\gamma_{ij}(X, \sigma^{G_{-i}})$ represents their expected value. By Assumption 1.3 $v_{ij}(\cdot)$ is a linear function of the exogenous and endogenous covariates:

$$v_{ij}(X, G_{-i}; \theta_0) = [\delta_{ij}(X_i, X_j), \gamma_{ij}(X, G_{-i})] \cdot \theta_0 \tag{1.13}$$

The expected value of v_{ij} conditional on X and the event that σ^G is the selected equilibrium is therefore:

$$\mathbb{E}[v_{ij}(X, G_{-i}; \theta_0) | X, \sigma^{G_{-i}}] = [\delta_{ij}(X_i, X_j), \gamma_{ij}(X, \sigma^{G_{-i}})] \cdot \theta_0 \tag{1.14}$$

[16] Measurement error in the network topology is an important, yet largely unexplored issue that goes beyond the scope of this paper (De Paula, 2017; Advani and Malde, 2014; Bramoullé et al., 2020). Our results rely on the assumption that the network is measured in an accurate and complete manner, like other methods do (Leung, 2015; De Paula et al., 2018). In particular, the first-step estimates may not be consistent in presence of measurement errors.

Suppressing some of the input arguments, we can now rewrite Equation (1.9) as:

$$P(G_{ij} = 1|X, \sigma^G) = \Phi([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta_0) \cdot \Phi([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta_0) \tag{1.15}$$

Since $\{\epsilon_{ij}|i, j \in N, i \neq j\}$ are drawn independently from one another, conditional on X and the event that σ^G is selected, the likelihood of observing a network G is:

$$\begin{aligned} L(\theta, \sigma^G) = \prod_{i,j>i}^{n} \Bigg[&\bigg(\Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\Big) \cdot \Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\Big) \bigg)^{G_{ij}} \\ &\times \bigg(1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\Big) \cdot \Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\Big) \bigg)^{1-G_{ij}} \Bigg] \end{aligned} \tag{1.16}$$

By taking the log of this expression and dividing by the number of observations we obtain the following log-likelihood function:

$$\begin{aligned} l(\theta, \sigma^G) = \frac{2}{n(n-1)} \sum_{i,j>i}^{n} \Bigg[&\bigg(G_{ij} \cdot \log \Big(\Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\Big) \cdot \Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\Big) \Big) \bigg) \\ &+ \bigg(\Big(1 - G_{ij}\Big) \cdot \log \Big(1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta\Big) \cdot \Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta\Big) \Big) \bigg) \Bigg] \end{aligned} \tag{1.17}$$

This function depends on the unobserved beliefs σ^G. We therefore cannot directly proceed to maximize it with respect to θ. Instead, we follow a two-step procedure, where in the first stage we consistently estimate the symmetric equilibrium beliefs (Subsection 1.3.2), and in the second stage we plug the estimated beliefs into the log-likelihood function to recover the estimands (Subsection 1.3.3).

Two comments about the log-likelihood function are in place. First, note that if we rule out endogenous covariates from the marginal utility the model boils down to a bivariate probit with partial observability (Poirier, 1980). Partial observability occurs when a positive outcome for one response variable is only observed if the other response variable is also positive.[17,18] This model has been used to model undirected network

[17] In our context the decision rules of the two agents can be interpreted as two partially-observed latent response variables, where the θs are by construction the same across the two equations. For a discussion of how identification depends on the functional form of the payoff function, see Poirier (1980).

[18] Note that in our setting the two latent response variables are partially observed, but the equilibrium link is observed accurately. This stands in contrast with situations where links are measured

formation in the absence of externalities by Comola and Fafchamps (2014). Second, note that under uniqueness of equilibria, resorting to recovering σ^G from the data is not strictly necessary. Instead, we could analytically calculate the unique equilibrium beliefs for any candidate θ that is being considered by the optimization algorithm and evaluate the log-likelihood function at these beliefs.[19]

1.3.2 Estimating Beliefs

Under the assumption that beliefs satisfy the symmetric equilibrium condition, producing a consistent estimate of the beliefs $\hat{\sigma}^G$ is straightforward. Consider a set of observationally equivalent pairs of agents. In a symmetric equilibrium, the belief that any of these pairs are linked is identical (due to symmetry) and correct (since it is an equilibrium). Thus, the proportion of pairs within this set that are linked in the observed network is a consistent estimator for the belief that any of the pairs in the set are linked. In the case of discrete attributes, the estimator for the belief that i and j are linked $\hat{\sigma}^G_{ij}$ is defined as:

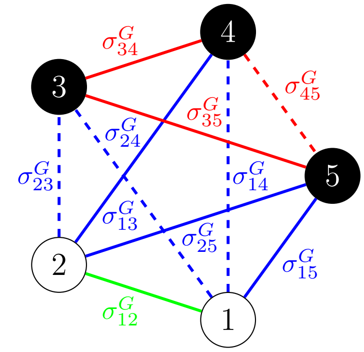
$$\hat{\sigma}^G_{ij} \equiv \frac{\sum_{l,k>l} \Big[G_{kl} \cdot \mathbb{1}\Big\{ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \Big\} \Big]}{\sum_{l,k>l} \Big[\mathbb{1}\Big\{ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \Big\} \Big]} \tag{1.18}$$

Proposition 1.2. *Under assumptions 1.1 and 1.4, $\hat{\sigma}^G_{ij}$ is consistent for σ^G_{ij} for all $i, j \in N$ such that $i \neq j$.*

Figure 1.2 provides an example of how this estimator is calculated. As in Figure 1.1, the colors of the agents depict their one-dimensional binary attribute (being either black or white) and the colors of the edges and weights illustrate which pairs of agents have identical ex-ante linking probabilities (due to symmetry). The type of the edges illustrate which links are realized in the observed network – full lines describe realized links and dashed lines describe unrealized ones. The $\hat{\sigma}^G$ matrix presents the estimated beliefs. Concentrating on the black pairs, for instance, since two out of the three potential links between this type of pairs are realized we estimate the belief that these pairs are linked to be $\frac{2}{3}$.

with error (Chandrasekhar and Lewis, 2012; Candelaria and Ura, 2018; Thirkettle, 2019).

[19] Under multiplicity, one could in principle calculate all equilibria for a candidate θ and compare their likelihood value. However, this approach could be difficult to implement (Aguirregabiria and



$$\hat{\sigma}^G = \begin{bmatrix} 0 & 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} & \frac{2}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

Figure 1.2: Example of beliefs estimation

To get a better understanding of the advantages of this estimation method, it is useful to contrast this "large-market" framework with an alternative "many-markets" framework. Assume we were to observe many repetitions of the game over a constant set of agents ("many-markets"). The same pairs of agents are expected to have the same ex-ante linking probabilities across games, regardless of anonymity of preferences or symmetry of beliefs. As mentioned in Subsection 1.2.2, this only holds when agents are guaranteed to play the same equilibrium across games, which can be obtained by assuming a degenerate equilibrium selection mechanism. Thus, the proportion of games in which a given pair is linked gives a consistent estimate for the belief that this pair would be linked as the number of games increases to infinity. In our context of "large-market" framework we can relax the assumption that the equilibrium selection mechanism is degenerate and estimate symmetric beliefs from one single network realization. This broadens the applicability of our estimator, since many network datasets depict a single network (Goyal et al., 2006; Mele, 2017).[20]

Two additional points are worth mentioning. First, since the denominator sums up pairs that are exactly identical, it is only applicable to cases where all attributes in X are discrete. Second, since the estimator divides the set of observations into bins of identical pairs of agents, we risk not having enough observations within each bin when the sample size is small, the number of attributes is high, and their support is large. Both of these concerns are formally addressed in Appendix 1.7. Subsection 1.7.1 allows for the inclusion of continuous attributes, thereby resolving the first concern.

Mira, 2007).

[20]Our estimation procedure also carries over to the case of multiple networks. In this case one should estimate beliefs separately for each network in the first stage, and then pool all observations together to estimate preferences in the second stage.

Subsection 1.7.2 discusses smoothing of discrete variables, which addresses the second.

1.3.3 Estimating Preferences

Once $\hat{\sigma}^G$ is computed, plugging it into Equation (1.17) and maximizing with respect to θ yields our estimates $\hat{\theta}$ of θ_0. Since $\hat{\sigma}^G$ is consistent $\hat{\theta}$ is also consistent under standard regularity conditions. Below we state the consistency and asymptotic normality results for the second-stage estimator.

Proposition 1.3 (Consistency)**.** *Under assumptions 1.1-1.4 and standard regularity conditions, $\hat{\theta}$ is consistent for θ_0.*

Since the endogenous covariates are computed based on the estimated beliefs rather than the true ones, standard errors should be adjusted. Proposition 1.4 shows how to do so provided that the aggregate values of the true endogenous covariates and the estimated ones are identical.

Proposition 1.4 (Asymptotic Normality)**.** *Assume the endogenous covariates satisfy*

$$\sum_{i,j\neq i} \gamma_{ij}(X, G_{-i}) = \sum_{i,j\neq i} \gamma_{ij}(X, \hat{\sigma}^{G_{-i}}). \tag{1.19}$$

Let γ_{ij}^0 denote the output of $\gamma_{ij}(X, \sigma^G)$ and γ_0 denote the set of γ_{ij}^0 for all i, j. Then, under assumptions 1.1-1.4:

$$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, [V(\gamma_0, \theta_0)]^{-1}\Psi(\gamma_0, \theta_0, G)[V(\gamma_0, \theta_0)]^{-1}) \tag{1.20}$$

where V and Ψ are defined as in Equations 1.69 and 1.88 in the Appendix.

As mentioned above, proposition 1.4 relies on the endogenous covariates satisfying condition (1.19).[21] Lemma 1.1 proves this property for endogenous covariates of the form $\frac{1}{n-1}\sum_{k\neq i} G_{jk} \cdot \mu(X_k)$, where $\mu(X_k)$ represents some weighting function of agent k's attributes, assuming the beliefs are estimated according to (1.18). $\mu(\cdot)$ captures any sort of observed attributes that agents might care about in their indirect contacts. For instance, when deciding to form a link with someone, they may care not only about the number of this potential partner's friends but also about their wealth. The illustration of Section 1.5 makes use of covariates of this form.

[21] If condition (1.19) does not hold, one could still compute standard errors with an appropriately-designed bootstrap test.

Lemma 1.1. *Let* $\gamma_{ij}(X, G_{-i}) \equiv \frac{1}{n-1}\sum_{k\neq i} G_{jk} \cdot \mu(X_k)$*, where* $\mu(X_k)$ *is some weighting function of the attributes of agent* k *and* $\hat{\sigma}^{G_{-i}}$ *be defined as in (1.18), then, for any* G_{-i}*, condition (1.19) holds.*

1.4 Simulations

We now describe the simulation exercise we designed to evaluate the asymptotic performance of the estimator in networks of increasing size (from $n = 100$ to $n = 900$). First we describe the data generating process, then the estimation results.

1.4.1 Data Generating Process

For a given number of agents n with a one-dimensional attribute vector X_i, we posit a data generating process of the form:

$$X_i \sim U\{0, 1, 2, 3, 4\}, \quad \forall i \tag{1.21}$$

$$\epsilon_{ij} \sim N(0, 1), \quad \forall i, j, \quad i \neq j \tag{1.22}$$

$$v_{ij} = \theta_1 + \theta_2 X_i + \theta_3 |X_i - X_j| + \theta_4 \frac{1}{n-1}\sum_{k\neq i} G_{jk}, \quad \forall i, j, \quad i \neq j \tag{1.23}$$

$$\theta_0 = [-1.6, 0.5, -0.1, 1]' \tag{1.24}$$

where $\frac{1}{n-1}\sum_{k\neq i} G_{jk}$ represents the average number of indirect friends that j grants access to, as in the example of Section 1.2.3. θ_0 is set so that the utility function is not dominated by its deterministic component, i.e. so that proposal decisions are sensitive to ϵ_{ij}.

The data generating process consists of three steps: first we draw the attribute X_i for all i. Second we find a corresponding symmetric equilibrium σ^G. We use an algorithm that starts from a randomly drawn belief matrix, computes the corresponding linking probabilities, and updates beliefs accordingly until convergence is achieved. Algorithm 1.1 describes the process in more detail.[22]

As a third step we draw the ϵ_{ij} values and construct a network realization G according to the following rule: a link in G exists if and only if the realization of ϵ_{ij} and ϵ_{ji} are such that $v_{ij}(X, \sigma^G, \theta_0) + \epsilon_{ij} \geq 0$ and $v_{ji}(X, \sigma^G, \theta_0) + \epsilon_{ji} \geq 0$.

[22] For further details on the convergence behaviour of the algorithm see Rabinovich et al. (2013).

Algorithm 1.1 Search Algorithm

1: Generate a random belief matrix σ^G
2: Calculate the matrix of linking probabilities L, given σ^G, X and θ_0:
$L_{ij} = L_{ji} = \Phi(\mathbb{E}[v_{ij}(X, \sigma^G, \theta_0)]) \cdot \Phi(\mathbb{E}[v_{ji}(X, \sigma^G, \theta_0)])$
3: **if** $\sigma^G \not\approx L$:
4: Re-assign $\sigma^G = L$ and go back to line 2
5: **else**
6: Return σ^G

For each $n \in \{100, 500, 900\}$ we generate 500 networks according to the procedure above. The networks that result from this process exhibit many commonly observed characteristics of real-world networks: the average geodesic distance between connected agents is low (≈ 2.1); the clustering coefficient is high compared to the linking probability of a comparable Poisson random network (≈ 0.25 vs. ≈ 0.1); and the degree distribution is positively skewed. The average degrees are approximately 10, 53 and 95 for $n \in \{100, 500, 900\}$ respectively.

1.4.2 Simulation Results

In the estimation step, for each simulation draw we use the realized network G and the agents' attributes X (but not the error terms and beliefs) to estimate σ^G (as explained in Section 1.3.2). Then we maximize Equation (1.17) by replacing σ^G with $\hat{\sigma}^G$ to obtain $\hat{\theta}$.

Table 1.2 presents histograms of the obtained $\hat{\theta}$ values. The values of the true coefficients are depicted by the vertical lines at the center of each sub-figure. As n increases the distributions of the estimated values become increasingly tight around the true values. This shows that the estimators are consistent.

Table 1.3 presents the fitted Kernel distributions of $\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0)$ over all 500 iterations (in dashed lines) as well as true normal distributions with mean zero and variance $V^{-1}\Psi V^{-1}$ (in full lines). As n increases, the dashed lines converge to the full lines. This shows that the estimators are asymptotically normal.

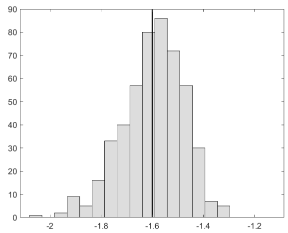
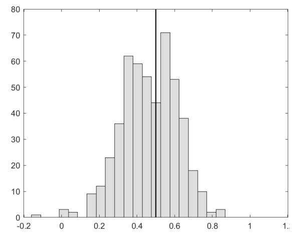
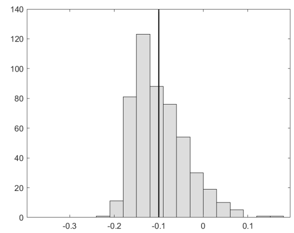
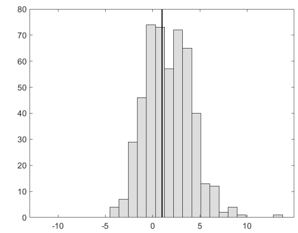
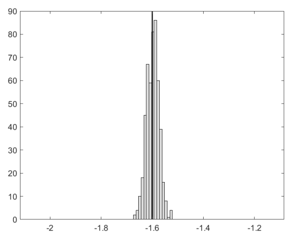
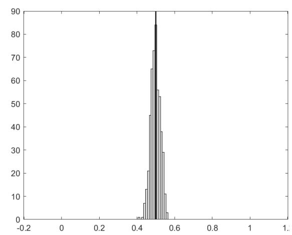
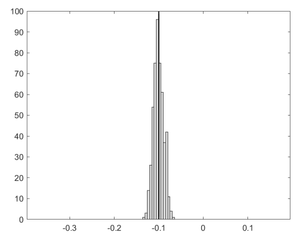
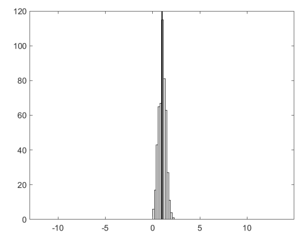
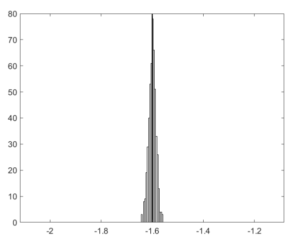
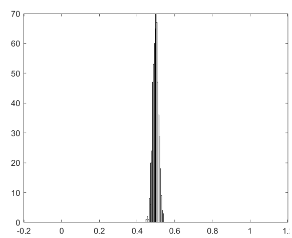
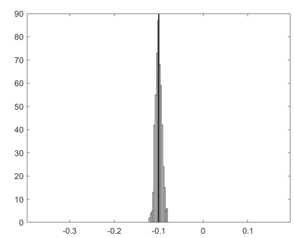
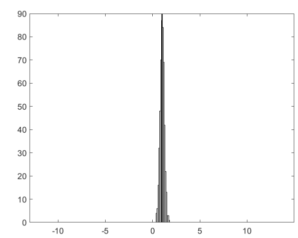
n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
100				
500				
900				

Table 1.2: Consistency

Note: The table reports histograms of estimated coefficients. The true values of the coefficients are depicted by the vertical line at the center of each sub-figure.

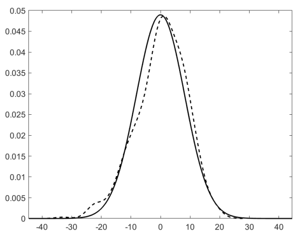
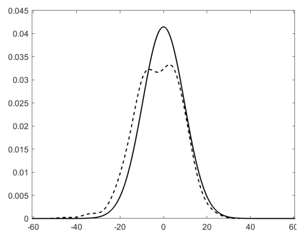
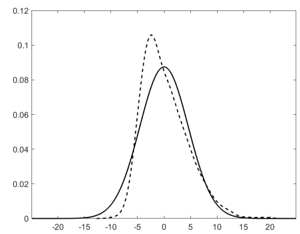
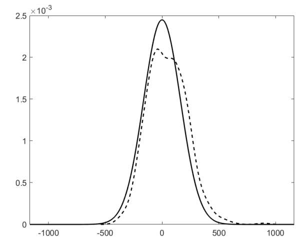
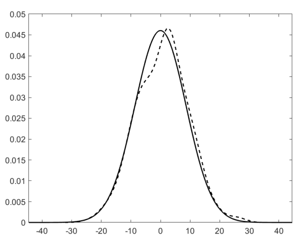
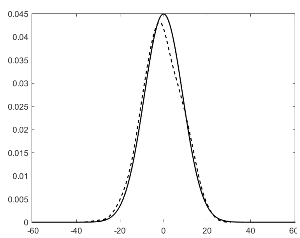
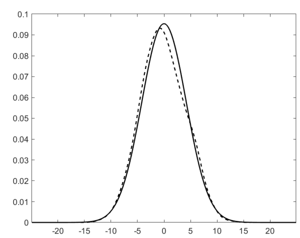
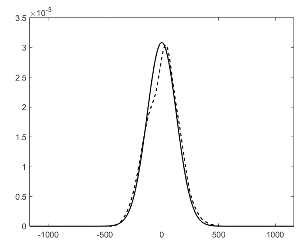
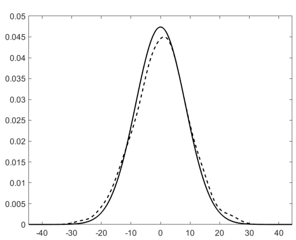
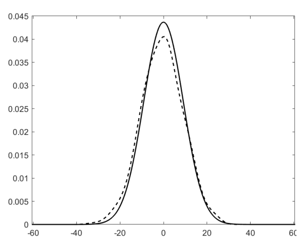
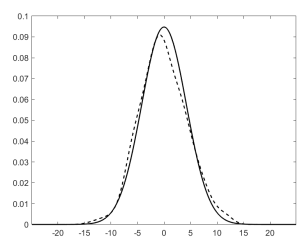
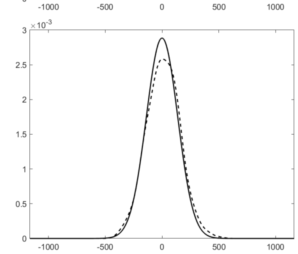
n	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_1 - \theta_1)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_2 - \theta_2)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_3 - \theta_3)$	$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta}_4 - \theta_4)$
100				
500				
900				

Table 1.3: Asymptotic normality.

Note: The dashed lines depict the fitted Kernel distributions of $\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0)$. The full lines depict true normal distributions with mean 0 and variance $V^{-1}\Psi V^{-1}$.

1.5 Empirical Illustration

1.5.1 Data Description

We use data on the risk sharing network of Nyakatoke, a small village in the Buboka rural district of Tanzania.[23] Rural villages are an appropriate setting to study network formation, because the population can be entirely surveyed and several confounding effects (such as spatial and informational barriers) can be reasonably ruled out. The village of Nyakatoke consists of 119 households which have been interviewed in five regular intervals from February to December 2000. The data contains information on households' demographics, wealth, income sources and income shocks, transfers and risk-sharing links. At the time of the study, the village of Nyakatoke is isolated (the few unpaved roads leading to the village are hardly passable during rains), densely inhabited (90% of households live within a distance of 1 kilometer from each other) and relatively poor (consumption for adult equivalent is less than 2 US$ a week, and average food share in consumption is 77%). Households earn most of their income from agricultural activities, especially the cultivation of coffee and banana; other sources of income are rare and off-farming activities are mostly considered supplementary to farming.

During the first survey round all respondents were asked *'Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labour?'*.[24] The phrasing of this survey question was intended to capture undirected links of mutual assistance, and qualitative interviews and pilot tests suggested that respondents have understood it that way.[25] Our empirical exercise assumes that Nyakatoke survey responses represent undirected agreements of mutual help which could be activated if one of the partners is struck by an income shock. This is in line both with the survey design and with

[23]These data have been the object of numerous articles (De Weerdt and Dercon, 2006; De Weerdt and Fafchamps, 2011; Vandenbossche and Demuynck, 2013; Comola and Fafchamps, 2014).

[24]Respondents could list as many names as they wanted. They could also mention partners who live outside the village (this occurs in 34% of all declared partners). Since we have no information on the attributes of households outside the village we are obliged to omit them from the analysis.

[25]This phrasing was first piloted in the Philippines (Fafchamps and Lund, 2003) and subsequently adopted in the Nyakatoke survey, because respondents understand it and are willing to answer. Other survey questions on directed flows were tried during the pilots, for instance drawing a distinction between people which respondents would help and people which respondents would seek help from. But respondents were confused by this distinction, which they perceived as non-existent, and complained they are asked the same question twice. See also Comola and Fafchamps (2014).

theoretical work on the voluntary nature of risk-sharing arrangements (Bloch et al., 2008; Jackson et al., 2012).[26]

The resulting risk-sharing network of Nyakatoke is depicted in Figure 1.3. It consists of 490 links among $(119 \cdot 118)/2 = 7021$ household dyads. This network displays a mean geodesic distance of 2.5 steps and a maximum geodesic distance of 5 steps. No household is isolated, and the average degree is 8.2. The network exhibits all the empirical regularities of large social networks.[27]

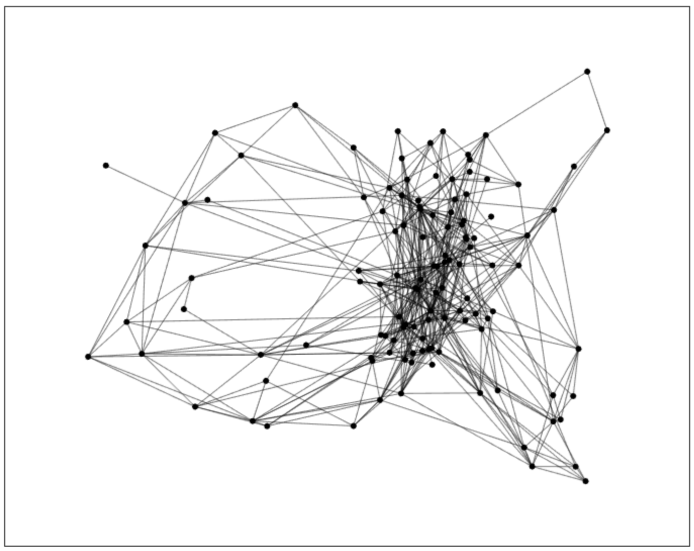


Figure 1.3: The risk-sharing network of Nyakatoke

1.5.2 Main Results

We now illustrate the estimation procedure described in Section 1.3 using the Nyakatoke data. We take the household as a unit of observation ($n = 119$) and we include as covariates: a constant, the geographical distance between households (in meters), the wealth

[26]In case of discordant reports, we assume that an undirected link exists whenever it is declared by at least one of the households involved. This is the most common stand in the empirical literature on risk-sharing links, and it is equivalent to assuming that all observed discordances are due to under-reporting.

[27]The Nyakatoke network has a unique component covering the entire population, the diameter is in the order of $\log(n)$ and the clustering coefficient is 7 times larger than in a randomly generated Poisson network with similar characteristics.

of j,[28] three types of homophily regressors, and two types of endogenous regressors. The homophily regressors are binary variables that take the value 1 if i and j belong to the same family,[29] same clan[30] or same religion[31] respectively. These exogenous covariates (i.e., distance, wealth and dummies for same family, clan and religion respectively) were identified by the previous literature as strong predictors of risk-sharing link formation in developing countries. The endogenous regressors are the number of j's friends ($\sum_{k\neq i} G_{jk}$) and the total wealth of j's friends ($\sum_{k\neq i} G_{jk} \cdot \text{Wealth}_k$).[32]

We run the first stage using the individual attributes that are used in the second stage (Wealth_j), as well as those implied by the relational attributes in the second stage (Family_i, Clan_i, Religion_i). Since the relational attribute "Distance_{ij}" does not imply a unique individual geographic location, we treat the entire vector of distances between i and the rest of the households as i's individual attribute.[33] The categorical variables (family, clan, religion) and continuous variables (distance, wealth) are combined as described in Appendix 1.7 (in particular, Equation (1.30)), with $\lambda = 0.1$ and h set according to the "normal reference rule-of-thumb") and a normal kernel function. Figure 1.4 presents a histogram of the resulting estimated beliefs.

The results of the second stage are reported in Table 1.4. Column 1 presents a specification without endogenous regressors, for reference. Columns 2 to 4 present different specifications including the endogenous regressors (number of j's friends only, total wealth of j's friends only, both). Column 5 presents the marginal effects that correspond to the most complete specification of column 4. Standard errors are computed according the expression given in Proposition 1.4 with the true parameters replaced by their estimates.

[28]The wealth of a household is defined as the total monetary value of its land and livestock assets (1 unit = $100,000$ Tanzanian shillings). Data on land were originally in acres and were transformed in monetary equivalent with a conversion rate of 300,000 *tzs* for 1 acre which reflects average local prices in 2000. For international comparisons, the exchange rate in 2000 was 1 US dollar for 800 *tzs*. Since land and livestock are publicly observable with a good degree of precision, we argue that the regressor satisfies the common-knowledge assumption (Section 2.1).

[29]Two households i and j are said to belong to the same family if there is some blood relation between at least one of the members of i and at least one of the members of j.

[30]There are 26 clans in Nyakatoke. 10 of them have only one household.

[31]There are three religions in Nyakatoke: Roman Catholic (49 households), Lutheran (46 households) and Muslim (24 households).

[32]For presentation purposes we do not re-scale these variables in the results of Table 4. In fact, the normalization is only needed to facilitate the asymptotic case where n approaches infinity.

[33]Consider a three-agent network in which agents 1 and 2 have the same geographic distances from (2,3) and (1,3), respectively. These distance profiles can be obtained by assuming various individual locations for agents 1 and 2, e.g. all location configurations in which all agents are located on a line and agents 1 and 2 are located symmetrically around agent 3.

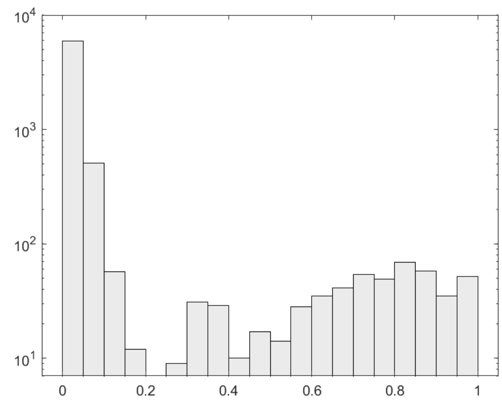


Figure 1.4: Histogram of the estimated beliefs in the Nyakatoke network. Note that the y-axis is on a logarithmic scale.

As for the endogenous regressors, the coefficient of the number of j's friends can be positive or negative depending on whether households prefer potential partners to have many or few other partners. In principle, both types of externalities are conceivable in the context of risk-sharing arrangements: if j has many friends she may have a rather limited amount of resources to devote to i, implying a negative coefficient. If j has many friends she is likely to be well-positioned to provide i with financial support in case of need, and is also less likely to rely heavily on i in case she herself is in need, implying a positive coefficient. The sum of wealth of j's friends is expected to be positive, as this grants j access to greater wealth which may indirectly benefit i.

The significance of the endogenous regressors' coefficients in Table 1.4 provides evidence for the existence of network externalities. Concentrating on the full specification in column 4, the positive sign of the coefficient of the number of j's friends suggests that the benefits from having a partner with many other partners (greater financial resilience) outweigh the costs (dilution of attention and/or resources). For the average pair i and j, an increment of one unit in the expected number of j's friends ($\approx$ 12% of the average expected number of j's friends) is associated with an increase of roughly 0.016 in the probability of a proposal ($\approx$ 9% of the average predicted proposal probability).

The signs of the other coefficients conform to our expectations. The constant appears negative, reflecting the idea that maintaining links is costly. The coefficient of the geographical distance between households is also negative, as distance is likely to render links harder to maintain. The coefficient of wealth is positive, as the wealthier a potential partner is the more helpful she could be in case of a negative income shock. The coefficients of the homophily regressors are all positive, in line with the large evidence that similarity between agents makes them more desirable to each other.

	G				ME
	(1)	(2)	(3)	(4)	(5)
Same family	0.8436***	0.8496***	0.8556***	0.8493***	0.2934***
	(0.0627)	(0.0643)	(0.0642)	(0.0644)	(0.0256)
Same clan	0.1661***	0.1483**	0.1487**	0.1485**	0.0415**
	(0.0579)	(0.0601)	(0.0605)	(0.0602)	(0.0177)
Same religion	0.1649***	0.1752***	0.1735***	0.1751***	0.0495***
	(0.0401)	(0.041)	(0.0411)	(0.041)	(0.0118)
Distance$_{ij}$	-0.0009***	-0.0009***	-0.0009***	-0.0009***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0)
Wealth$_j$	0.0586***	0.0358***	-0.021	0.0376**	0.0098**
	(0.0069)	(0.008)	(0.0192)	(0.0155)	(0.004)
Number of j's friends		0.0598***		0.0607***	0.0159***
		(0.0086)		(0.0113)	(0.003)
Wealth of j's friends			0.0075***	-0.0002	0
			(0.0017)	(0.0013)	(0.0003)
Constant	-0.6482***	-1.0888***	-0.6243***	-1.0967***	
	(0.0563)	(0.0863)	(0.0579)	(0.1063)	
# observations	7021	7021	7021	7021	

Table 1.4: Estimated coefficients.

Notes: Column 5 reports the marginal effects for the specification of column 4. Standard errors in parentheses. Significance level based on false discovery rate q-values (Benjamini and Hochberg, 1995): *q<10%, **q<5%, and ***q<1%.

In Appendix 1.8 we present estimates obtained under different hypotheses about mis-reporting and the data generation process. The scope of the exercise is to illustrate the use of our estimation protocol in the context of self-reported network data. In particular, we modify our estimator to accommodate for a unilateral link formation rule, and we show that it yields different results from the directed unilateral estimator by Leung (2015).

1.6 Concluding remarks

Data on network interactions were previously scarce but are now becoming more available to economists. The current enthusiasm for network data from digital interaction platforms (Vosoughi et al., 2018; Blumenstock, 2018) has refueled the research interest about how non-digital links are formed, and how they respond to strategic incentives. Models of link formation with network externalities are at the frontier of the econometric research, facing difficulties related to dimensionality and multiplicity of equilibria (Graham, 2015; Chandrasekhar, 2016; De Paula, 2017). Our paper fills a void in the literature by proposing a versatile method to estimate network externalities in a simultaneous-move game of undirected link formation. This method is naturally suited for bilateral link formation models, but it could also be applied to unilateral models where only the undirected link outcome (rather than the proposals) is observable. We provide existence, consistency and asymptotic normality results for the proposed estimator, and we test its asymptotic performance through a simulation exercise. In the context of bilateral link formation, this procedure provides a simpler alternative to methods exploiting pairwise stability under complete information (De Paula et al., 2018; Sheng, 2020). Importantly, it allows to make inferences about various aspects of agents' preferences over network topology when data on a single (and possibly large) network are available. For instance, our method could be paired with data issued from a randomized experiment, allowing the researcher to disentangle endogenous network externalities from other exogenous determinants (e.g., agent's randomly allocated treatment status).[34]

We illustrate the method using data on risk-sharing in a Tanzanian village named Nyakatoke. Risk-sharing links are commonly assumed to be mutually agreed upon and provide an intriguing case for the role of externalities from indirect connections. Results confirm that the network architecture has an explanatory value: households seem to take into consideration the number of indirect friends they stand to gain when making linking decisions. Our estimates suggest that an additional two-steps-away connection is associated with an average increase of roughly 9% in the predicted proposal probability.

[34]The assumption that the attributes of others are observable suits well the case of a medium-sized village community where randomization is implemented through a public lottery.

1.7 Appendix: Extensions

1.7.1 Continuous Attributes

The restriction that the selected equilibrium is symmetric requires that identical pairs of agents have identical ex-ante linking probabilities. In the case of continuous attributes no two pairs of agents are identical. The symmetric equilibrium condition is therefore non-restrictive - it is trivially fulfilled for any equilibrium. When attributes are continuous we substitute the symmetry condition with a continuity condition, requiring that *similar* pairs of agents have *similar* ex-ante linking probabilities. Formally, an equilibrium σ^G is continuous if for all $\varepsilon > 0$ there exists $\delta > 0$ such that for all $ij \neq kl \in N$:

$$\begin{gathered}(\|X_i - X_k\| < \delta \textit{ and } \|X_j - X_l\| < \delta) \textit{ or } (\|X_i - X_l\| < \delta \textit{ and } \|X_j - X_k\| < \delta) \\ \Downarrow \\ |\sigma_{ij} - \sigma_{kl}| < \varepsilon\end{gathered} \tag{1.25}$$

The following proposition establishes the existence of a continuous equilibrium.

Proposition 1.5 (Existence)**.** *For any continuous X and for any θ_0, there exists a continuous equilibrium.*

Under the assumption that the selected equilibrium is continuous, we can estimate σ^G_{ij} using Kernel methods. Letting $d(X_i, X_j, X_k, X_l)$ denote the vector of distances in attributes between the two unordered pairs,[35] $K(\cdot)$ denote a standard product kernel function, and h denote the bandwidth selection, the estimator is:

$$\hat{\sigma}^G_{ij} \equiv \frac{\sum\limits_{l,k>l} G_{kl} \cdot K\left(\dfrac{d(X_i, X_j, X_k, X_l)}{h}\right)}{\sum\limits_{l,k>l} K\left(\dfrac{d(X_i, X_j, X_k, X_l)}{h}\right)} \tag{1.26}$$

[35]Formally:

$$d(X_i, X_j, X_k, X_l) = \begin{cases} [X_i - X_k, X_j - X_l] & \text{if} \quad \|[X_i - X_k, X_j - X_l]\| \leq \|[X_i - X_l, X_j - X_k]\| \\ [X_i - X_l, X_j - X_k] & \text{otherwise} \end{cases}$$

Proposition 1.6. *When X is continuous and the selected equilibrium is continuous, $\hat{\sigma}^G_{ij}$ is consistent for σ^G_{ij} for all $i, j \in N$ such that $i \neq j$.*

While the estimator in equation (1.18) applies only to discrete attributes, the estimator in equation (1.26) applies only to continuous attributes. In various applications, however, X may contain a mix of both discrete and continuous attributes. One approach to deal with these cases is to weigh the discrete variables of an observation according to equation (1.18), the continuous variables according to equation (1.26), and define the weight of the observation as the product of the two. Formally, letting X^d_i be i's discrete attributes, X^c_i her continuous ones and $X_i = [X^d_i, X^c_i]$, this approach yields the following estimator:

$$\hat{\sigma}^G_{ij} \equiv \frac{\sum_{l,k>l} G_{kl} \mathbb{1}\Big\{(X^d_i = X^d_k \wedge X^d_j = X^d_l) \vee (X^d_i = X^d_l \wedge X^d_j = X^d_k)\Big\} K\left(\frac{d(X^c_i, X^c_j, X^c_k, X^c_l)}{h}\right)}{\sum_{l,k>l} \mathbb{1}\Big\{(X^d_i = X^d_k \wedge X^d_j = X^d_l) \vee (X^d_i = X^d_l \wedge X^d_j = X^d_k)\Big\} K\left(\frac{d(X^c_i, X^c_j, X^c_k, X^c_l)}{h}\right)} \tag{1.27}$$

1.7.2 Smoothing

A practical concern that may arise with respect to both the "mixed attributes" estimator and the "only discrete" estimator is that in a finite sample the number of observations with identical discrete attributes may be too small to allow for a meaningful estimation. This happens in particular when the sample size is small, the number of discrete variables is high and their support is large. Li and Racine (2007) suggest overcoming this problem by smoothing the discrete variables. Let $X^d_{i,s}$ be the sth component of the X^d_i vector and define

$$t_s(X^d_{i,s}, X^d_{j,s}, X^d_{k,s}, X^d_{l,s}, \lambda) \equiv \begin{cases} 0 & \text{if } (X^d_{i,s} = X^d_{k,s} \wedge X^d_{j,s} = X^d_{l,s}) \vee (X^d_{i,s} = X^d_{l,s} \wedge X^d_{j,s} = X^d_{k,s}) \\ 1 & \text{otherwise} \end{cases} \tag{1.28}$$

and

$$T(X^d_i, X^d_j, X^d_k, X^d_l, \lambda) \equiv \prod_s \lambda^{t_s(X^d_{i,s}, X^d_{j,s}, X^d_{k,s}, X^d_{l,s}, \lambda)} \tag{1.29}$$

Using $T(\cdot)$ as the product kernel function for the discrete variables, the mixed attributes estimator (1.27) becomes:

$$\hat{\sigma}_{ij}^{G} \equiv \frac{\sum_{l,k>l} G_{kl} \cdot T(X_i^d, X_j^d, X_k^d, X_l^d, \lambda) \cdot K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)}{\sum_{l,k>l} T(X_i^d, X_j^d, X_k^d, X_l^d, \lambda) \cdot K\left(\frac{d(X_i^c, X_j^c, X_k^c, X_l^c)}{h}\right)} \tag{1.30}$$

Note that when $\lambda = 0$, $T(X_i^d, X_j^d, X_k^d, X_l^d, 0)$ takes the value 1 if ij and kl are identical in their discrete attributes and 0 otherwise. (1.30) therefore reduces to (1.27) and no smoothing occurs. On the other extreme, when $\lambda = 1$, $T(X_i^d, X_j^d, X_k^d, X_l^d, 1) = 1$ for all ij and kl. The discrete attributes are therefore completely smoothed out. $\lambda \in (0, 1)$ corresponds to different levels of smoothing of the discrete variables.

1.8 Appendix: Auxiliary results

The data collected in Nyakatoke have been analyzed extensively (De Weerdt and Dercon, 2006; De Weerdt and Fafchamps, 2011; Vandenbossche and Demuynck, 2013). The survey question presented to respondents ('*Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help in cash, kind or labor?*') was originally intended to elicit undirected links. This interpretation is compatible with the phrasing of the question, and with the way respondents have understood it as suggested by qualitative interviews and pilot tests of the questionnaire. However, the interpretation of self-reported link data remains to some extent ambiguous. In what follows we use Nyakatoke data to explore the issue of mis-reporting, and to compare alternative models of link formation. Results are presented in Table 1.5 below.

1.8.1 Mis-reporting in undirected networks

We now discuss how to estimate undirected network formation models on the basis of multiple (conflicting) survey responses. In the scenario of reference, the link formation process is believed to be bilateral. However, the discussion also holds for models where link formation rule is unilateral, but only the undirected link outcome is observed by the econometrician (see below).

We let R denote the directed matrix of reports, with its generic element R_{ij} taking the value 1 if i reports having a link with j and 0 otherwise.[36] In the Nyakatoke data, we frequently observe that $R_{ij} \neq R_{ji}$.[37] These discrepancies do not come as a surprise. In fact, discrepancies are the rule rather than the exception even when link data depict supposedly mutual relationships such as risk-sharing, goods exchanges or friendship (Comola and Fafchamps, 2014, 2017). If Nyakatoke data truly represented undirected bilateral links (as assumed in Section 1.5) *and* they were perfectly measured, we would observe that R_{ij} and R_{ji} always coincide, but this is obviously not the case. One possible interpretation is that respondents provide information on undirected links, but their responses differ because of mis-reporting (which can be imputed to different factors such as unintentional errors, intentional omissions, data aggregation mistakes at the household level).

In this situation, the most common stand is to impute discrepancies to under-reporting, that is, to assume that an undirected link exists whenever it is declared by at least one of the parties involved. In practice, this means that for estimation purposes we set $G_{ij} = \max\{R_{ij}, R_{ji}\}$. This is the assumption behind our main results of Section 1.5 (Table 1.4, Column 4), which are reported again in column 1 of Table 1.5 below for reference. Another way to deal with discrepancies (possibly the most neutral one) is to assume that over-reporting and under-reporting are equally likely, which is the approach we take in column 2.[38,39] Note that for both bilateral models (columns 1 and 2) actual reports contained in R serve to build the undirected network matrix G, but they are *not* interpreted as a measure of the unobserved proposals S (see Section 2).

1.8.2 Undirected unilateral model

In what follows we show how our partial-observability framework could be modified to accommodate undirected links issued from a unilateral link formation model. Assume

[36] Each survey respondent in Nyakatoke could declare her links with all other individuals in the village. Since data are aggregated at the level of the household, $R_{ij} = 1$ if an adult member of i mentions an adult member of j.

[37] In particular, there are 6531 (93%) dyads such that $R_{ij} = R_{ji} = 0$, 140 (2%) dyads such that $R_{ij} = R_{ji} = 1$, and 350 (5%) dyads such that $R_{ij} \neq R_{ji}$.

[38] To do so we include two distinct observations per each undirected link ($G^1_{ij} = R_{ij}$ and $G^2_{ij} = R_{ji}$), and we correct the standard errors to account for the increased sample size.

[39] We have also estimated the bilateral model by assuming over-reporting, that is, setting $G_{ij} = \min\{R_{ij}, R_{ji}\}$. This assumption is rather drastic in the context of our illustration as it reduces the frequency of non-zero outcomes significantly. However, the results we obtain (available upon request) are comparable to the ones reported in columns (1) and (2) for sign and significance.

we observe an undirected network G: if we believe that the underlying link formation process is unilateral, we may want to estimate a model where a link *does not* exist if and only if both agents *do not* propose to each other, that is:

$$G_{ij} = S_{ij} + S_{ji} - S_{ij} \cdot S_{ji}. \tag{1.31}$$

This implies rewriting Equation (1.15) as:

$$P(G_{ij} = 0 | X, \sigma^G) = [1 - \Phi([\delta_{ij}, \gamma_{ij}(\sigma^{G_{-i}})]\theta_0)] \cdot [1 - \Phi([\delta_{ji}, \gamma_{ji}(\sigma^{G_{-j}})]\theta_0)] \tag{1.32}$$

and maximizing the associated log-likelihood function. This model is similar to the bilateral model discussed in Section 3 in that they both assume that proposals are not observed.

The choice between a bilateral or unilateral link formation process is ultimately at the discretion of the researcher, depending on the specific characteristics and knowledge of the data application. Given the phrasing of the Nyakatoke's survey question (*"people you can rely on and/or that can rely on you"*), this estimation strategy could in fact be an appealing alternative. For what concerns risk sharing, this choice boils down to whether agents can refuse links which are against their self-interest. Following a large literature in economic development, our main results of Section 1.5 assume that links are mutually agreed upon. However, in Column 3 of Table 1.5 we estimate Equation (1.32) by imputing discrepancies to under-reporting ($G_{ij} = \max\{R_{ij}, R_{ji}\}$), which sets a direct comparison with Column (1).

1.8.3 Directed unilateral model

The models estimated in Columns (1) to (3) assume that link data represent undirected (albeit mis-measured) links, and that link proposals are not separately observable. However, given the phrasing of the survey question it is also conceivable that Nyakatoke survey responses represent link proposals. In columns (4) and (5) we use Nyakatoke data to fit the model of unilateral link formation proposed by Leung (2015) by maximizing the following log-likelihood function

$$l(\theta, \sigma^Z) = \frac{1}{n(n-1)} \sum_{i,j\neq i}^{n} \Bigg[Z_{ij} \cdot \log\Big(\Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{Z_{-i}})]\theta\Big)\Big) + \Big(1 - Z_{ij}\Big) \cdot \log\Big(1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^{Z_{-i}})]\theta\Big)\Big) \Bigg] \tag{1.33}$$

where Z denotes a directed network and beliefs about Z are estimated according to[40]

$$\hat{\sigma}^Z_{ij} \equiv \frac{\sum_{l,k\neq l} Z_{kl} \cdot \mathbb{1}\Big\{X_i = X_k \wedge X_j = X_l\Big\}}{\sum_{l,k\neq l} \mathbb{1}\Big\{X_i = X_k \wedge X_j = X_l\Big\}} \tag{1.34}$$

The standard errors are computed as described in the proof of Proposition 1.4, with the log-likelihood function replaced by the one above.

We present two versions of this estimator. The difference between columns (4) and (5) is in the way we define the directed links Z_{ij}, Z_{ji} on the basis of Nyakatoke reports. In column (4) we force the undirected network obtained under the assumption of under-reporting into the unilateral model above by setting $Z_{ij} = Z_{ji} = \max\{R_{ij}, R_{ji}\}$.[41] This transformation may appear odd in the context of our illustration where survey data contain discordant reports, but it is meant to show that one can still force a unilateral link formation model when network data are undirected (i.e. contain one single measurement of the link outcome per pair). It also allows us to draw a straightforward comparison to columns (1) and (3). In column (5) we feed the unilateral model the actual survey reports as directed links, that is, we set $Z_{ij} = R_{ij}$, $Z_{ji} = R_{ji}$. This is unarguably the most sensible choice when two distinct reports are available for each dyad, as is the case for Nyakatoke.

1.8.4 Results

First, for what concerns the two bilateral models we remark that different assumptions regarding mis-reporting produce results which are qualitatively similar (columns 1 and

[40]We report here the estimator for the beliefs in the case of discrete X only for the sake of simplicity. In practice, since in our application X includes some continuous attributes, we use the directed-case equivalent of the "mixed attributes" estimator described in Appendix 1.7.

[41]Note that in this case Z is symmetric (i.e. $Z_{ij} = Z_{ji}$). It is nonetheless directed in the sense that all links are interpreted as being formed (or not) unilaterally.

Table 1.5: Ancillary Results

Directionality:	undirected			directed	
Formation Process:	bilateral $S_{ij}S_{ji}$	bilateral $S_{ij}S_{ji}$	unilateral $S_{ij}+S_{ji}-S_{ij}S_{ji}$	unilateral S_{ij}	unilateral S_{ij}
Dep. var.:	under reporting $\max\{R_{ij}, R_{ji}\}$	equal probabilities R_{ij}, R_{ji}	under reporting $\max\{R_{ij}, R_{ji}\}$	under reporting $\max\{R_{ij}, R_{ji}\}$	actual reports R_{ij}, R_{ji}
	(1)	(2)	(3)	(4)	(5)
Same family	0.8493***	0.7909***	0.912***	1.0722***	0.9752***
	(0.064)	(0.050)	(0.0681)	(0.020)	(0.013)
Same clan	0.1485**	0.1333***	0.1699**	0.2005***	0.1772***
	(0.060)	(0.048)	(0.0666)	(0.004)	(0.003)
Same religion	0.1751***	0.1522***	0.1903***	0.2015***	0.1727***
	(0.041)	(0.034)	(0.0461)	(0.002)	(0.001)
Distance$_{ij}$	-0.0009***	-0.0009***	-0.001***	-0.0011***	-0.001***
	(0.000)	(0.000)	(0.0001)	(0.000)	(0.000)
Wealth$_j$	0.0376**	0.0314***	0.0414***	0.0289***	0.0262***
	(0.015)	(0.010)	(0.0117)	(0.000)	(0.000)
Number of j's friends	0.0607***	0.0776***	0.0593***	0.0384***	0.0649***
	(0.011)	(0.007)	(0.0078)	(0.000)	(0.000)
Wealth of j's friends	-0.0002	-0.0002	-0.0015**	-0.001***	-0.0013***
	(0.001)	(0.001)	(0.0006)	(0.000)	(0.000)
Constant	-1.0967***	-1.2176***	-2.2353***	-1.6373***	-1.9177***
	(0.106)	(0.069)	(0.1011)	(0.003)	(0.002)
# observations	7021	14042	7021	14042	14042

Note: in Column 2 we have $N = 14042$ because we include two distinct observations per each undirected link ($G^1_{ij} = R_{ij}$ and $G^2_{ij} = R_{ji}$). Standard errors are adjusted accordingly.
*p<10%, **p<5%, and ***p<1%.

2). The two directed unilateral models (columns 4 and 5) display no significant differences in the pattern of results either. When comparing the bilateral models against the directed unilateral models, however, notable differences come to light regarding both size and significance of coefficients. For instance, if we focus on the endogenous covariates we notice that the estimated coefficient for the wealth of j's friends is much smaller in the bilateral models than in the directed unilateral models, and it loses significance. Unsurprisingly, results from the undirected unilateral model of Column (3) stand in-between the bilateral models and the directed unilateral model.[42] We also remark that

[42]The constant term in column (3) is smaller than the ones in columns (1) and (2), as the unilateral model requires lower proposal probabilities to observe a link.

the standard errors in the directed models are smaller. This difference stems from the fact that undirected models rely on partial observability, while directed models assume full observability: these alternative assumptions are not without consequences as they may lead towards different conclusions even when models are fed the same dependent variables.[43]

These results illustrate some important points. First, note that we only have to deal with mis-reporting when we fit undirected models on data with multiple (and largely discordant) reports, as is often the case for self-reported links from individual surveys. This is not the case when information on links is retrieved from administrative sources (e.g. registers of commercial or financial transactions, phone calls, traceable interactions on digital platforms) which usually contain one single link measurement per dyad. Whenever the link measurement is unique, it is natural (both from the point of view of the data and the interpretation) to estimate an undirected model, and no assumptions on mis-reporting are needed. Nonetheless, as column (4) illustrates, forcing a directed unilateral model over undirected network data is still possible. While this may be attractive (because the estimation procedure for the directed model is computationally simpler), the comparison between columns (1) and (4) shows that it may come at the cost of drawing different conclusions (e.g. regarding the impact of the wealth of j's friends). This illustrates our point that directed and undirected models are intrinsically different.

When link data contain two distinct reports per dyad, the researcher has the choice between an undirected link formation model (with some assumptions on mis-reporting) or a directed unilateral model (ideally using the two distinct reports directly, as in column 5). This choice depends on which data generation process appears to be most fit for the data at hand. We know from theory that unilateral and bilateral link formation rules result in fundamentally different network structures, which in turn has profound implications on the aggregate outcome that can be achieved (Bala and Goyal, 2000; Charness and Jackson, 2007). We have argued in Section 1.5 that Nyakatoke data are likely to result from a bilateral link formation process. However, in many other situations link data are likely to represent directed proposals. This is the case of Leung (2015), who illustrates his inference method using data on trust networks in rural India.

[43]Note that columns (1), (3) and (4), as well as columns (2) and (5) have the same dependent variables. This allows to perform a non-nested log-likelihood test on these two groups of models (Vuong, 1989). Still, the treatment of the data differs: for instance, the bilateral model of column (1) assumes that non-existing links could be due to the refusal of one part only, while the directed unilateral model of column (4) assumes that both sides must be non-interested.

He identifies directed links on the basis of a question where respondents could list names of people they trusted enough to lend a substantial amount of money.[44] Differently from the case of Nyakatoke, this question is phrased in a directed manner, and discrepancies in reports need not to be imputed to mis-reporting. Still, his results suggest that reciprocity is an important determinant of directed links of trust, which motivates the study of undirected link formation models that we pursue in this paper.

1.9 Appendix: Additional Empirical Application

This section applies our estimation method on the same dataset used in Leung (2015), providing a direct comparison with his results. We follow the same variable definition and data cleaning procedure, with the only exception that the network is assumed to be undirected. In particular, whenever a directed link between i and j is reported in the data, we assume that an undirected link between the two exists. Our specification differs from his only in that endogeneous variables that are irrelevant for undirected networks (e.g. "reciprocity") are dropped. For a description of the data, see Leung (2015).

The estimated coefficients of the homophily and endogeneous regressors are reported in Table 1.6. Columns 1 (respectively, 2) reports the results from our procedure assuming that the existence of an undirected link indicates that both sides are (respectively, one side is) interested in it. Column 3 reports the results from Leung's method. All columns use a smoothing parameter $\Gamma = 0.1$ (this parameter is denoted ω in Leung (2015)). Column 3 differs from the results reported in Leung (2015) due to two reasons. First, the specification is not entirely identical. Second, there seems to be a coding error in Leung's original code (see README.md file in his online replication package).

The endogeneous regressors include j's out-degree ($\frac{1}{n}\sum_{k\neq j} G_{jk}$), as well as j's out-degree weighted by either caste or religion ($\frac{1}{n}\sum_{k\neq j} G_{jk}\mathbb{1}\{\text{caste}_i = \text{caste}_k\}$, and $\frac{1}{n}\sum_{k\neq j} G_{jk}\mathbb{1}\{\text{religion}_i = \text{religion}_k\}$, respectively). Note that for the undirected cases "out-degree" and "number of j's friends" are interchangeable. Under all specifications externalities from indirect connections turn out positive and significant. However, the assumption one takes on the directionality of the links dramatically changes the magnitude of these externalities.

[44] *"Whom do you trust enough that if he or she needed to borrow Rs. 50 for a day you would lend it to him or her?"*

Table 1.6: Application to the data used in Leung (2015)

Directionality:	undirected		directed
Formation Process:	bilateral $S_{ij}S_{ji}$	unilateral $S_{ij}+S_{ji}-S_{ij}S_{ji}$	unilateral S_{ij}
Dep. var.:	under reporting $\max\{R_{ij},R_{ji}\}$	under reporting $\max\{R_{ij},R_{ji}\}$	actual reports R_{ij},R_{ji}
	(1)	(2)	(3)
Same caste	0.2294***	0.2650***	0.3286***
	(0.0308)	(0.0352)	(0.0001)
Same religion	0.4061***	0.4646***	0.4674***
	(0.0457)	(0.0588)	(0.0001)
Same sex	0.4977***	0.7105***	0.7375***
	(0.0778)	(0.0468)	(0.0001)
Same language	0.0330	0.0387	0.0576***
	(0.0233)	(0.0266)	(0.0000)
Same family	1.5572***	1.4959***	1.5955***
	(0.0905)	(0.0717)	(0.0001)
Out-degree	52.4109***	43.2457***	26.0504***
	(8.1185)	(5.7632)	(0.0156)
Out-degree, caste	-13.6293**	-9.6498**	-11.5319***
	(5.6137)	(4.9925)	(0.0119)
Out-degree, religion	-20.6551**	-13.4883*	-2.8915***
	(9.3352)	(7.2454)	(0.0169)
Constant	-2.5287***	-4.1715***	-3.8536***
	(0.1407)	(0.1381)	(0.0002)
# observations	246345	246345	492690

Note: Standard errors in paretheses. *p<10%, **p<5%, and ***p<1%.

1.10 Appendix: Proofs

1.10.1 Proposition 1.1

Proof. Denote by Σ the set of all σ^G matrices such that:

1. $\forall i,j \in N,\ \sigma_{ij}^G \in [0,1]$

2. $\forall i \in N,\ \sigma_{ii}^G = 0$

3. $\forall i,j \in N,\ \sigma_{ij}^G = \sigma_{ji}^G$

4. $\forall i,j,k,l \in N,\ (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k) \implies \sigma^G_{ij} = \sigma^G_{kl}$

Denote by $\Gamma(\cdot)$ the function that maps belief matrices to linking probabilities:

$$\Gamma_{ij}(\sigma^G) \equiv \begin{cases} \Phi(\mathbb{E}[v_{ij}(X, G_{-i};\theta_0) \mid \sigma^{G_{-i}}]) \cdot \Phi(\mathbb{E}[v_{ji}(X, G_{-j};\theta_0) \mid \sigma^{G_{-j}}]) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{1.35}$$

By 1.9, an equilibrium is a fixed point of $\Gamma(\cdot)$, i.e. a σ^G such that for all $i, j \in N$:

$$\Gamma_{ij}(\sigma^G) = \sigma^G_{ij} \tag{1.36}$$

To prove that such σ^G exists we verify the conditions of Brouwer's fixed point theorem.

$\Gamma(\cdot)$ maps from Σ to Σ. First, since Γ_{ij} is either the product of two probabilities or 0, $\Gamma_{ij} \in [0,1]$ for all i, j. Second, by definition $\Gamma_{ii} = 0$ for all i. Third, since Γ_{ij} depends symmetrically on the expected utility of i from a link with j and of j from a link with i, $\Gamma_{ij} = \Gamma_{ji}$ for all i, j. Fourth, for any two agents i and k such that $X_i = X_k$, condition 4 above implies that for any third agent $j \neq i, k$ the input matrix must satisfy $\sigma^G_{ij} = \sigma^G_{kj}$. By conditions 2 and 3, $\sigma^G_{ii} = \sigma^G_{kk}$ and $\sigma^G_{ik} = \sigma^G_{ki}$. The ith and kth rows and columns in σ^G therefore contain the same elements, implying that $\sigma^{G_{-i}}$ and $\sigma^{G_{-k}}$ are identical up to a permutation of labels. Anonymity of $v_{ij}(\cdot)$ thus implies that $\Gamma_{ij} = \Gamma_{kj}$ for all i, j, k. Applying the same logic for an agent l such that $X_j = X_l$, we obtain also that $\Gamma_{ij} = \Gamma_{kl}$ for all i, j, k, l.

Γ is continuous in σ^G. The expected utilities are continuous in σ^G, and $\Phi(\cdot)$ is a continuous function. Therefore Γ is continuous in σ^G.

Σ is a convex subset of $[0,1]^{n\times n}$. Since any linear combination of any two matrices in Σ yields a matrix in Σ, it is a convex set.

Σ is a compact subset of $[0,1]^{n\times n}$. The sets of values that each entry in the matrices in Σ can obtain are bounded (by 0 and 1) and closed (for off-diagonal elements because the boundaries 0 and 1 are included and for diagonal elements because they are singletons). The Cartesian product of bounded and closed sets is also bounded and closed, so Σ is bounded and closed. By the Heine-Borel theorem, it follows that Σ is compact.

The existence of a symmetric Bayesian equilibrium thus follows from Brouwer's fixed point theorem. □

1.10.2 Proposition 1.2

Proof. First, note that the linking statuses of all pairs of agents which are observationally equivalent are independent and have the same expected value (due to symmetry). Thus, for any i and j we can apply a law of large numbers:

$$\hat{\sigma}_{ij}^G \equiv \frac{\sum_{l,k>l}\Big[G_{kl}\cdot \mathbb{1}\Big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\Big\}\Big]}{\sum_{l,k>l}\Big[\mathbb{1}\Big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\Big\}\Big]} \tag{1.37}$$

$$= \frac{\sum_{l,k>l\,:\,(X_i=X_k\wedge X_j=X_l)\vee(X_i=X_l\wedge X_j=X_k)} G_{kl}}{\sum_{l,k>l}\Big[\mathbb{1}\Big\{(X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k)\Big\}\Big]} \tag{1.38}$$

$$\xrightarrow{p} \mathbb{E}[G_{kl} \mid (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k), X, \sigma^G] \tag{1.39}$$

In addition:

$$\mathbb{E}[G_{kl} \mid (X_i = X_k \wedge X_j = X_l) \vee (X_i = X_l \wedge X_j = X_k), X, \sigma^G] = \mathbb{E}(G_{ij} \mid X, \sigma^G) \tag{1.40}$$

$$= P(G_{ij} = 1 \mid X, \sigma^G) \tag{1.41}$$

$$= \sigma_{ij}^G \tag{1.42}$$

Line 1.40 is true because the probabilities of observationally equivalent pairs to be linked are equal (due to symmetry), and line 1.42 is true because in equilibrium beliefs are correct. □

1.10.3 Proposition 1.3

Proof. To prove that $\hat{\theta}$ is consistent for θ we verify the conditions of Theorem 2.1 in Newey and McFadden (1994).

$\mathbb{E}[l(\theta, \sigma^G)]$ **is uniquely maximized at** θ_0**.** $\mathbb{E}[l(\theta, \sigma^G)]$ is uniquely maximized at θ_0 if for all $\theta \neq \theta_0$, $\mathbb{E}[l(\theta, \sigma^G)] - \mathbb{E}[l(\theta_0, \sigma^G)] < 0$. We now show that this is true.

$$\mathbb{E}[l(\theta, \sigma^G)] - \mathbb{E}[l(\theta_0, \sigma^G)] = \mathbb{E}\left[\frac{\log(L(\theta, \sigma^G))}{\frac{1}{2}n(n-1)} - \frac{\log(L(\theta_0, \sigma^G))}{\frac{1}{2}n(n-1)}\right] \tag{1.43}$$

$$= \frac{1}{\frac{1}{2}n(n-1)}\mathbb{E}\left[\log(L(\theta, \sigma^G) - \log(L(\theta_0, \sigma^G))\right] \tag{1.44}$$

$$= \frac{1}{\frac{1}{2}n(n-1)}\mathbb{E}\left[\log\left(\frac{L(\theta, \sigma^G)}{L(\theta_0, \sigma^G)}\right)\right] \tag{1.45}$$

$$\begin{aligned} = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i}^{n} \Bigg[& \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\Big)\times \\ & \log\left(\frac{\Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta\Big)}{\Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\Big)}\right) + \\ & \Big(1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\Big)\Big) \times \\ & \log\left(\frac{1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta\Big)}{1 - \Phi\Big([\delta_{ij}, \gamma_{ij}(\sigma^G)]\theta_0\Big)\Phi\Big([\delta_{ji}, \gamma_{ji}(\sigma^G)]\theta_0\Big)}\right)\Bigg] \end{aligned} \tag{1.46}$$

$$\leq \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i}^{n} \Bigg[\log(1)\Bigg] \tag{1.47}$$

$$= 0 \tag{1.48}$$

where line 1.47 is obtained by applying Jensen's inequality.
This establishes that θ_0 maximizes $\mathbb{E}[l(\theta, \sigma^G)]$. It remains to show that it is its *unique* maximizer. Consider line 1.46. Since $\Phi(\cdot)$ is strictly positive, the only way 1.46 would equal 0 is if the fractions inside the logs evaluate to 1, but this only happens when $\theta = \theta_0$. Thus, θ_0 is the unique maximizer of $\mathbb{E}[l(\theta, \sigma^G)]$.

Θ **is compact.** True by assumption.

$\mathbb{E}[l(\theta, \sigma^G)]$ **is continuous in** θ **and** $l(\theta, \hat{\sigma}^G)$ **converges uniformly in probability to** $\mathbb{E}[l(\theta, \sigma^G)]$**.** We show that this is true by verifying the conditions of Lemma 2.4 in Newey and McFadden (1994). First, Θ is compact, by assumption. Second, $l(\theta, \sigma^G)$ is continuous in θ because $\Phi(\cdot)$ and $\log(\cdot)$ are continuous. Third, we need to show that there exists a function $d(G, \delta, \hat{\gamma})$ such that $|l(\theta, \hat{\sigma}^G)| \leq d(G, \delta, \hat{\gamma})$ and $\mathbb{E}[d(G, \delta, \hat{\gamma})] < \infty$.

We start by considering the absolute value of the first part of the log-likelihood function:

$$
\begin{aligned}
& |\log\left(\Phi([\delta_{ij},\hat{\gamma}_{ij}]\theta)\Phi([\delta_{ji},\hat{\gamma}_{ji}]\theta)\right)| && (1.49)\\
&= |\log(\Phi(0)\Phi(0)) + \frac{\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ij},\hat{\gamma}_{ij}]\theta - 0) + \frac{\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ij},\hat{\gamma}_{ij}]\theta - 0)| && (1.50)\\
&\leq 2 + \frac{\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}|[\delta_{ij},\hat{\gamma}_{ij}]\theta| + \frac{\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}|[\delta_{ij},\hat{\gamma}_{ij}]\theta| && (1.51)\\
&\leq 2 + (1 + |[\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta}|)|[\delta_{ij},\hat{\gamma}_{ij}]\theta| + (1 + |[\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta}|)|[\delta_{ij},\hat{\gamma}_{ij}]\theta| && (1.52)\\
&\leq 2 + (1 + \|[\delta_{ij},\hat{\gamma}_{ij}]\| \cdot \|\tilde{\theta}\|) \cdot \|[\delta_{ij},\hat{\gamma}_{ij}]\| \cdot \|\theta\| + (1 + \|[\delta_{ji},\hat{\gamma}_{ji}]\| \cdot \|\tilde{\theta}|\|) \cdot \|[\delta_{ij},\hat{\gamma}_{ij}]\| \cdot \|\theta\| && (1.53)
\end{aligned}
$$

Where line 1.50 is true by the mean value theorem (recall that the derivative of $\log(\Phi(v)\Phi(u))$ w.r.t v is $\frac{\phi(v)}{\Phi(v)}$ and w.r.t u is $\frac{\phi(u)}{\Phi(u)}$), line 1.51 is true by the triangular inequality, line 1.52 is true because $\frac{\phi(v)}{\Phi(v)} \leq 1 + |v|$ for all v, and line 1.53 is true by the Cauchy-Schwartz inequality.
Consider now the absolute value of the second part of the log-likelihood function:

$$
\begin{aligned}
& |\log\left(1 - \Phi([\delta_{ij},\hat{\gamma}_{ij}]\theta)\Phi([\delta_{ji},\hat{\gamma}_{ji}]\theta)\right)| && (1.54)\\
&= |\log(1 - \Phi(0)\Phi(0)) + \frac{-\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})}{1 - \Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})}([\delta_{ij},\hat{\gamma}_{ij}]\theta - 0)\\
&\quad + \frac{-\phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{1 - \Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}([\delta_{ji},\hat{\gamma}_{ji}]\theta - 0)| && (1.55)\\
&\leq 2 + \frac{\phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})}{1 - \Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})\Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})} \cdot |[\delta_{ij},\hat{\gamma}_{ij}]\theta|\\
&\quad + \frac{\phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})}{1 - \Phi([\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta})\Phi([\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta})} \cdot |[\delta_{ji},\hat{\gamma}_{ji}]\theta| && (1.56)\\
&\leq 2 + (1 + |[\delta_{ij},\hat{\gamma}_{ij}]\tilde{\theta}|) \cdot |[\delta_{ij},\hat{\gamma}_{ij}]\theta| + (1 + |[\delta_{ji},\hat{\gamma}_{ji}]\tilde{\theta}|) \cdot |[\delta_{ji},\hat{\gamma}_{ji}]\theta| && (1.57)\\
&\leq 2 + (1 + \|[\delta_{ij},\hat{\gamma}_{ij}]\| \cdot \|\tilde{\theta}\|) \cdot \|[\delta_{ij},\hat{\gamma}_{ij}]\| \cdot \|\theta\| + (1 + \|[\delta_{ji},\hat{\gamma}_{ji}]\| \cdot \|\tilde{\theta}\|) \cdot \|[\delta_{ji},\hat{\gamma}_{ji}]\| \cdot \|\theta\| && (1.58)
\end{aligned}
$$

Where line 1.57 is true because $\frac{\phi(v)\Phi(u)}{1-\Phi(v)\Phi(u)} \leq \frac{\phi(v)}{1-\Phi(v)} = \frac{\phi(v)}{\Phi(-v)} \leq 1 + |v|$.
Letting $\theta_m = \sup_{\theta\in\Theta}\|\theta\|$, 1.53 and 1.58 imply that $|l(\theta, \hat{\sigma}^G)|$ is bounded from above by:

$$
2+(1+\|[\delta_{ij},\hat{\gamma}_{ij}]\|\cdot\|\theta_m\|)\cdot\|[\delta_{ij},\hat{\gamma}_{ij}]\|\cdot\|\theta_m\|+(1+\|[\delta_{ji},\hat{\gamma}_{ji}]\|\cdot\|\theta_m\|)\cdot\|[\delta_{ji},\hat{\gamma}_{ji}]\|\cdot\|\theta_m\| \quad (1.59)
$$

And a sufficient condition for the expected value of this function to be finite is that $\mathbb{E}\left[[\delta_{ij},\hat{\gamma}_{ij}][\delta_{ij},\hat{\gamma}_{ij}]'\right]$ and $\mathbb{E}\left[[\delta_{ji},\hat{\gamma}_{ji}][\delta_{ji},\hat{\gamma}_{ji}]'\right]$ exist and are finite. □

1.10.4 Proposition 1.4

Proof. Denote the score of the log-likelihood by S and its ijth summand by S_{ij}:

$$S(\gamma,\theta) \equiv \nabla_\theta l(\theta) \tag{1.60}$$
$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} S_{ij}(\gamma_{ij},\gamma_{ji},\theta) \tag{1.61}$$
$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{q_{ij}(G_{ij}-m_{ij})}{m_{ij}(1-m_{ij})} \tag{1.62}$$

Where:

$$m_{ij} \equiv \Phi([\delta_{ij},\gamma_{ij}]\theta)\cdot\Phi([\delta_{ji},\gamma_{ji}]\theta) \tag{1.63}$$
$$q_{ij} \equiv \phi([\delta_{ij},\gamma_{ij}]\theta)\cdot[\delta_{ij},\gamma_{ij}]\cdot\Phi([\delta_{ji},\gamma_{ji}]\theta)+\phi([\delta_{ji},\gamma_{ji}]\theta)\cdot[\delta_{ji},\gamma_{ji}]\cdot\Phi([\delta_{ij},\gamma_{ij}]\theta) \tag{1.64}$$

Let $\hat{\gamma}_{ij}$ denote the output of $\gamma_{ij}(X,\hat{\sigma}^G)$ and $\hat{\gamma}$ denote the set of $\hat{\gamma}_{ij}$ for all i,j. By first order condition:

$$S(\hat{\gamma},\hat{\theta}) = 0 \tag{1.65}$$

By the mean value theorem, there exists a θ^* between $\hat{\theta}$ and θ_0 such that:

$$S(\hat{\gamma},\hat{\theta}) = S(\hat{\gamma},\theta_0) + \nabla_\theta S(\hat{\gamma},\theta^*)(\hat{\theta}-\theta_0) \tag{1.66}$$

Combining 1.65 and 1.66, and solving for $(\hat{\theta}-\theta_0)$ gives:

$$\hat{\theta}-\theta_0 = -\left(\nabla_\theta S(\hat{\gamma},\theta^*)\right)^{-1} S(\hat{\gamma},\theta_0) \tag{1.67}$$

Since $\hat{\gamma}$ and $\hat{\theta}$ are consistent, and given that θ^* is "trapped" between $\hat{\theta}$ and θ_0 (which makes it also consistent):

$$\nabla_\theta S(\hat{\gamma},\theta^*) - \mathbb{E}[\nabla_\theta S(\gamma_0,\theta_0) \mid X,\sigma^G] \xrightarrow{p} 0 \tag{1.68}$$

Denote the expected value of the hessian, by V and its ijth summand by V_{ij}:

$$
\begin{aligned}
V(\gamma, \theta) &\equiv \mathbb{E}[\nabla_\theta S(\gamma, \theta) \mid X, \sigma^G] && (1.69)\\
&= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} V_{ij}(\gamma_{ij}, \gamma_{ji}, \theta) && (1.70)\\
&= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{-q_{ij}q'_{ij}}{m_{ij}(1-m_{ij})} && (1.71)
\end{aligned}
$$

We can thus rewrite 1.67 as:

$$
\hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} S(\hat{\gamma}, \theta_0) \tag{1.72}
$$

By adding and subtracting $\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]$ we obtain:

$$
\hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \left(\underbrace{S(\hat{\gamma}, \theta_0) - \mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]}_{a} + \underbrace{\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]}_{b} \right) \tag{1.73}
$$

By a second order Taylor expansion of $S(\hat{\gamma}, \theta_0)$:

$$
\begin{aligned}
S(\hat{\gamma}, \theta_0) = S(\gamma_0, \theta_0) &+ \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} [\nabla_{\gamma_{ij}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ij} - \gamma^0_{ij}) \\
&+ \nabla_{\gamma_{ji}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ji} - \gamma^0_{ji})] + o_p(1)
\end{aligned} \tag{1.74}
$$

By a second-order Taylor expansion of $\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G]$:

$$
\begin{aligned}
\mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G] = \underbrace{\mathbb{E}[S(\gamma_0, \theta_0) \mid X, \sigma^G]}_{0} &+ \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \mathbb{E}\Big[\nabla_{\gamma_{ij}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \\
&\cdot (\hat{\gamma}_{ij} - \gamma^0_{ij}) + \nabla_{\gamma_{ji}} S(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ji} - \gamma^0_{ji}) \mid X, \sigma^G\Big] + o_p(1)
\end{aligned} \tag{1.75}
$$

Since, by the law of large numbers, the middle part of 1.74 converges to the middle part of 1.75:

$$
a = S(\gamma_0, \theta_0) + o_p(1) \tag{1.76}
$$

Denote the expected value of $\nabla_{\gamma_{ij}} S(\gamma, \theta)$ by M and its ijth summand by M_{ij}:

$$M(\gamma, \theta) \equiv \mathbb{E}[\nabla_{\gamma_{ij}} S(\gamma, \theta) \mid X, \sigma^G] \tag{1.77}$$

$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} M_{ij}(\gamma_{ij}, \gamma_{ji}, \theta) \tag{1.78}$$

$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \frac{-q_{ij} p'_{ij}}{m_{ij}(1-m_{ij})} \tag{1.79}$$

Where:

$$p_{ij} \equiv \phi([\delta_{ij}, \gamma_{ij}]\theta) \cdot \theta^{\gamma} \cdot \Phi([\delta_{ji}, \gamma_{ji}]\theta) \tag{1.80}$$

and θ^{γ} denotes the elements in θ which correspond to the endogenous regressors γ.

Using this notation we can rewrite 1.75 as:

$$\begin{aligned} \mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G] = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \Big[& M_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ij} - \gamma^0_{ij}) \\ & + M_{ji}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ji} - \gamma^0_{ji}) \Big] + o_p(1) \end{aligned} \tag{1.81}$$

$$= \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j\neq i} \Big[M_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\hat{\gamma}_{ij} - \gamma^0_{ij}) \Big] + o_p(1) \tag{1.82}$$

Since we assume that $\sum_{i,j\neq i} \gamma_{ij}(X, G_{-i}) = \sum_{i,j\neq i} \gamma_{ij}(X, \hat{\sigma}^{G_{-i}})$, we can replace $\gamma_{ij}(\hat{\sigma}^G)$ by $\gamma_{ij}(G)$, which we denote here by α_{ij}:

$$\begin{aligned} \mathbb{E}[S(\hat{\gamma}, \theta_0) \mid X, \sigma^G] = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \Big[& M_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ij} - \gamma^0_{ij}) \\ & + M_{ji}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ji} - \gamma^0_{ji}) \Big] + o_p(1) \end{aligned} \tag{1.83}$$

We now rewrite 1.73 using our replacements for a and b:

$$\begin{aligned} \hat{\theta} - \theta_0 = -\left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \Big[& S_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \\ + M_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ij} - \gamma^0_{ij}) + M_{ji}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ji} - \gamma^0_{ji}) + o_p(1) \Big] & \end{aligned} \tag{1.84}$$

Denote the ijth summand in this equation by W_{ij}:

$$W_{ij} \equiv S_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) + M_{ij}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ij} - \gamma^0_{ij}) + M_{ji}(\gamma^0_{ij}, \gamma^0_{ji}, \theta_0) \cdot (\alpha_{ji} - \gamma^0_{ji}) \tag{1.85}$$

Using this notation and multiplying through by $\sqrt{\frac{1}{2}n(n-1)}$:

$$\begin{aligned}\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0) = &- \left(V(\gamma_0, \theta_0) + o_p(1)\right)^{-1} \\ &\cdot \sqrt{\frac{1}{2}n(n-1)} \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} \left[W_{ij} + o_p(1)\right]\end{aligned} \tag{1.86}$$

Since the summands under the summation sign are conditionally independent (because conditional on X and σ^G, the variation in G_{ij} comes only from ϵ_{ij} and ϵ_{ji}, which are all assumed to be i.i.d.), we can now apply a central limit theorem:

$$\sqrt{\frac{1}{2}n(n-1)}(\hat{\theta} - \theta_0) \sim N(0, V^{-1}\Psi V^{-1}) \tag{1.87}$$

Where:

$$\Psi \equiv \frac{1}{\frac{1}{2}n(n-1)} \sum_{i,j>i} W_{ij} W'_{ij} \tag{1.88}$$

□

1.10.5 Lemma 1.1

Proof. Plugging in the definition of $\gamma_{ij}(\cdot)$ into condition 1.19, we obtain:

$$\sum_i \sum_{j \neq i} \frac{1}{n-1} \sum_{k \neq i,j} G_{jk} \cdot \mu(X_k) = \sum_i \sum_{j \neq i} \frac{1}{n-1} \sum_{k \neq i,j} \hat{\sigma}^G_{jk} \cdot \mu(X_k) \tag{1.89}$$

Below, we show that this statement is true if and only if $\sum_{i,j>i} G_{ij} \cdot \mu(X_j) = \sum_{i,j>i} \hat{\sigma}^G_{ij} \cdot \mu(X_j)$. We then proceed to show that the latter is indeed true. From 1.89:

$$\sum_i \sum_{j \neq i} \sum_{k \neq j} G_{jk}\mu(X_k) - \sum_i \sum_{j \neq i} G_{ji}\mu(X_i) = \sum_i \sum_{j \neq i} \sum_{k \neq j} \hat{\sigma}^G_{jk}\mu(X_k) - \sum_i \sum_{j \neq i} \hat{\sigma}^G_{ji}\mu(X_j) \tag{1.90}$$

$$(n-1) \cdot \sum_{i;j \neq i} G_{ij}\mu(X_j) - \sum_{i;j \neq i} G_{ij}\mu(X_j) = (n-1) \cdot \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) - \sum_{i;j \neq i} \hat{\sigma}^G_{ij}\mu(X_j) \tag{1.91}$$

$$(n-2)\cdot\sum_{i;j\neq i} G_{ij}\mu(X_j) = (n-2)\cdot\sum_{i;j\neq i}\hat{\sigma}^G_{ij}\mu(X_j) \tag{1.92}$$

$$\sum_{i;j\neq i} G_{ij}\mu(X_j) = \sum_{i;j\neq i}\hat{\sigma}^G_{ij}\mu(X_j) \tag{1.93}$$

We now show that this is true:

$$\sum_{i;j\neq i}\hat{\sigma}^G_{ij}\mu(X_j) = \sum_{X_A,X_B\in X}\sum_{i;j>i}\mathbb{1}\Big\{(X_i = X_A \wedge X_j = X_B)\vee(X_i = X_B \wedge X_j = X_A)\Big\}\times \hat{\sigma}^G_{ij}\cdot\mu(X_j) \tag{1.94}$$

$$= \sum_{X_A,X_B\in X}\sum_{i;j\neq i}\mathbb{1}\Big\{(X_i = X_A \wedge X_j = X_B)\vee(X_i = X_B \wedge X_j = X_A)\Big\}\times \frac{\sum_{k,l\neq k}\Big[G_{kl}\cdot\mathbb{1}\Big\{(X_k = X_A \wedge X_l = X_B)\vee(X_k = X_B \wedge X_l = X_A)\Big\}\Big]}{\sum_{k,l\neq k}\mathbb{1}\Big\{(X_k = X_A \wedge X_l = X_B)\vee(X_k = X_B \wedge X_l = X_A)\Big\}}\times \mu(X_l) \tag{1.95}$$

$$= \sum_{X_A,X_B\in X}\sum_{k;l\neq k}\mathbb{1}\Big\{(X_k = X_A \wedge X_l = X_B)\vee(X_k = X_B \wedge X_l = X_A)\Big\}\times G_{kl}\cdot\mu(X_l) \tag{1.96}$$

$$= \sum_{X_A,X_B\in X}\sum_{i;j\neq i}\mathbb{1}\Big\{(X_i = X_A \wedge X_j = X_B)\vee(X_j = X_B \wedge X_i = X_A)\Big\}\times G_{ij}\cdot\mu(X_j) \tag{1.97}$$

$$= \sum_{i;j\neq i} G_{ij}\mu(X_j) \tag{1.98}$$

Intuitively, this result comes from the fact that when calculating $\hat{\sigma}^G$ we essentially partition the agents into mutually exclusive groups of observationally equivalent pairs, and for each group "redistribute" the total number of links within it among its pairs (uniformly). □

1.10.6 Proposition 1.5

Proof. The proof is identical to that of Proposition 1.1, only that condition 4 has to be replaced by the definition of a continuous equilibrium and the claim that $\Gamma(\cdot)$ **maps from** Σ **to** Σ has to be reestablished.

Denote by Σ the set of all σ^G matrices such that:

1. $\forall i, j \in N,\ \sigma^G_{ij} \in [0,1]$
2. $\forall i \in N,\ \sigma^G_{ii} = 0$
3. $\forall i, j \in N,\ \sigma^G_{ij} = \sigma^G_{ji}$
4. $\forall \varepsilon > 0\ \exists \delta > 0$ such that $\forall i, j \neq k, l \in N$:

$$(\|X_i - X_k\| < \delta \text{ and } \|X_j - X_l\| < \delta) \text{ or } (\|X_i - X_l\| < \delta \text{ and } \|X_j - X_k\| < \delta)$$
$$\Downarrow$$
$$|\sigma^G_{ij} - \sigma^G_{kl}| < \varepsilon$$

We need to show that $\Gamma(\cdot)$ (defined in 1.35) maps from Σ to Σ.

First, since Γ_{ij} is either the product of two probabilities or 0, $\Gamma_{ij} \in [0,1]$ for all i, j. Second, by definition $\Gamma_{ii} = 0$ for all i. Third, since Γ_{ij} depends symmetrically on the expected utility of i from a link with j and of j from a link with i, $\Gamma_{ij} = \Gamma_{ji}$ for all i, j.

It remains to show that $\Gamma(\cdot)$ maps into matrices that satisfy the 4th condition above, that is, that by choosing a small δ we can make $|\Gamma_{ij}(\sigma^G) - \Gamma_{kj}(\sigma^G)|$ arbitrarily small for all i, k such that $\|X_i - X_k\| < \delta$ and $j \neq i, k$ (by taking another agent $l \neq j$ such that $\|X_j - X_l\| < \delta$ it then follows that we can also make $|\Gamma_{ij}(\sigma^G) - \Gamma_{kl}(\sigma^G)|$ arbitrarily small). Since $|\Gamma_{ij}(\sigma^G) - \Gamma_{kj}(\sigma^G)| = |\Phi(\mathbb{E}[v_{ij}|X, \sigma^G])\Phi(\mathbb{E}[v_{ji}]) - \Phi(\mathbb{E}[v_{kj}])\Phi(\mathbb{E}[v_{jk}])|$ and $\Phi(\cdot)$ is continuous, it is sufficient to show that $|\mathbb{E}[v_{ij}] - \mathbb{E}[v_{kj}]|$ and $|\mathbb{E}[v_{ji}] - \mathbb{E}[v_{jk}]|$ can be made arbitrarily small. For $|\mathbb{E}[v_{ij}] - \mathbb{E}[v_{kj}]|$, by the triangle inequality:

$$\begin{aligned}\left|\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]\right| = {} & |\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] \\ & + \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]|\end{aligned} \tag{1.99}$$

$$\leq |\mathbb{E}[v_{ij}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}]| + |\mathbb{E}[v_{kj}(X)|\sigma^{G_{-i}}] - \mathbb{E}[v_{kj}(X)|\sigma^{G_{-k}}]| \tag{1.100}$$

The first part of 1.100 can be made arbitrarily small by choosing a small δ because the expected value of $v(\cdot)$ is continuous in X. The second part can be made arbitrarily small because by condition 4 the closer X_i and X_k are the closer $\sigma^{G_{-i}}$ and $\sigma^{G_{-k}}$ must be, and the expected value of $v(\cdot)$ is continuous in beliefs. By a similar argument, $|\mathbb{E}[v_{ji}] - \mathbb{E}[v_{jk}]|$ can also be made arbitrarily small: the closer X_i and X_k are the closer the exogenous variables of v_{ji} and v_{jk}, and, by condition 4 so are the ith and kth rows (and columns) of σ^G, and hence so are the endogenous variables of v_{ji} and v_{jk} (in expectancy). Therefore, $\Gamma(\cdot)$ maps from Σ to Σ and the existence of a continuous equilibrium follows from Brouwer's fixed point theorem.

□

1.10.7 Proposition 1.6

Proof. We show that $\left| \frac{\sum_{l,k>l} G_{kl} \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} - \mathbb{E}[G_{ij}|X,\sigma^G] \right| \xrightarrow{p} 0.$

By adding and subtracting $\frac{\sum_{l,k>l} \mathbb{E}[G_{kl}|X,\sigma^G] \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}$ to the left-hand side and applying the triangle inequality we obtain that the left-hand side is at most:

$$\left| \underbrace{\frac{\sum_{l,k>l}(G_{kl} - \mathbb{E}[G_{kl}|X,\sigma^G]) \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}}_{a} \right| + \underbrace{\left| \frac{\sum_{l,k>l} \mathbb{E}[G_{kl}|X,\sigma^G] \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} - \mathbb{E}[G_{ij}|X,\sigma^G] \right|}_{b} \tag{1.101}$$

We deal with a and b separately and show that as h goes to zero and nh^q goes to infinity each of them converges in probability to zero. Starting with a, note that it can be written as the sample average of the random variable $(G_{kl} - \mathbb{E}[G_{kl}|X,\sigma^G]) \cdot w_{kl}$, with

$w_{kl} = \frac{K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} \cdot \frac{1}{2}n(n-1)$:

$$a = \frac{1}{\frac{1}{2}n(n-1)} \sum_{l,k>l} (G_{kl} - \mathbb{E}[G_{kl}|X, \sigma^G]) \cdot w_{kl} \tag{1.102}$$

By the law of large numbers this average converges to the expectation of $(G_{kl} - \mathbb{E}[G_{kl}|X, \sigma^G]) \cdot w_{kl}$, which, by the law of iterated expectations is zero. $|a|$ therefore converges to zero.

For b, note that conditional on σ^G the expected value of G_{ij} is a function of X. We can thus write $\mathbb{E}[G_{ij}|X, \sigma^G] = \rho(X_i, X_j, X_{-ij})$. Similarly, $\mathbb{E}[G_{kl}|X, \sigma^G] = \rho(X_k, X_l, X_{-kl})$. Because of the undirected nature of the network, $\rho(\cdot)$ is invariant to the order of its first two arguments. In addition, by anonymity, $\rho(\cdot)$ is invariant to permutations of the components of its third argument. The only relevant difference between the inputs in the two cases above is therefore the difference in attributes of the unordered pairs ij and kl. Applying a mean value theorem, we therefore obtain:

$$\mathbb{E}[G_{kl}|X, \sigma^G] = \rho(X_k, X_l, X_{-kl}) = \underbrace{\rho(X_i, X_j, X_{-ij})}_{=\mathbb{E}[G_{ij}|X, \sigma^G]} + D\rho(C) \cdot q(X_k, X_l, X_i, X_j) \tag{1.103}$$

where $D\rho(\cdot)$ denotes the derivative of $\rho(\cdot)$ with respect to its first two arguments and C lies in between (X_k, X_l, X_{-kl}) and (X_i, X_j, X_{-ij}).

By plugging 1.103 in b:

$$b = \left| \frac{\sum_{l,k>l} D\rho(C) \cdot d(X_k, X_l, X_i, X_j) \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} \right| \tag{1.104}$$

$$= |D\rho(C)| \cdot \left| \frac{\sum_{l,k>l} d(X_k, X_l, X_i, X_j) \cdot K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)}{\sum_{l,k>l} K\left(\frac{d(X_i,X_j,X_k,X_l)}{h}\right)} \right| \tag{1.105}$$

The first term is a constant. The second term converges to zero because as h approaches zero, the larger the difference in attributes between ij and kl the smaller the weight ascribed to it. b therefore also converges to zero. □

Chapter 2

Set-Valued Rational Expectations and Farsighted Stability

Abstract

An abstract game consists of a set of states, preferences over states, and an effectivity correspondence specifying which coalitions are allowed to move from one state to another. Most existing solution concepts aiming to capture farsighted behavior in abstract games are susceptible to either the *counterfactual* or the *overconfidence* critiques (or both). The first refers to situations where a coalition that makes a move takes it for granted that in the *counterfactual* scenario where it does not make a move, no other coalition does either. The second refers to situations where coalitions behave as if they had *full confidence* regarding which coalition would move at each state, despite the fact that abstract games contain no information regarding the order of play. This paper proposes a new solution concept, entitled Set-Valued Rational Expectations (SVRE), that captures optimal farsighted behavior and is immune to these critiques. We take a "rational expectations" approach in the sense that players are assumed to hold common and endogenous expectations about the dynamics of play from each state. In contrast to existing literature, we specify expectations which are *set-valued*, meaning that players take into account a range of potential continuation paths from each state, rather than a single one. When applied to extensive form games, SVRE boils down to subgame perfection. When applied to strategic form games, a state is supported as stationary by some SVRE if and only if it is Pareto efficient. When applied to partition function form games, a state is supported as stationary by some SVRE if and only if it is immune to myopically beneficial moves by coalitions that include either all players, or all players but one.

Keywords: Abstract games, Farsighted stability, Expectation functions.

JEL codes: C70, C71, C72.

2.1 Introduction

Abstract games (a.k.a "games in effectivity function form") are defined by a set of states, players' preferences over these states, and effectivity correspondence specifying which subsets of players ("coalitions") are allowed to move from one state to another. This game form is general enough to describe both cooperative and non-cooperative games. For instance, it can be used to describe games in characteristic function form, games in partition function form, strategic form games, extensive form games (with perfect information), network formation games, voting games, matching games and others. As such, abstract games can be seen as a bridge between cooperative and non-cooperative game theory.

The high level of generality in the way abstract games are defined renders the task of formulating appropriate solution concepts a considerable challenge. This is particularly true when trying to capture farsighted behavior, that is, the idea that when players contemplate making a move they take into account the entire chain (or chains) of reactions that it might trigger. In this paper, we formulate two critiques of existing farsighted solution concepts for abstract games and propose a new one that is immune to these critiques.

The first critique relates to reliance on the notion of "farsighted improving paths" (a.k.a. "farsighted objections", or "indirect dominance"), initially due to Harsanyi (1974).[1] A *farsighted improving path* is a finite sequence of states and coalitions $\{z^0, S^1, z^1, ..., S^K, z^K\}$ such that for all $1 \leq k \leq K$: *(i)* the coalition S^k has the ability to replace state z^{k-1} by state z^k; and, *(ii)* all players in S^k prefer the final state in the sequence z^K over the status quo state z^{k-1}. It is attractive to deploy this notion in order to describe farsighted behavior because it assumes players make decisions based on the *final* state that will be reached z^K. Its drawback, however, is that the final state z^K is compared against the *status quo* z^{k-1}, rather than against the final state of some alternative continuation path that could take place *had* S^k *decided to remain at the status quo* z^{k-1}. Following Chwe (1994) and Karos and Robles (2021), we refer to this critique as the *counterfactual critique*. Example 2.1 illustrates it.

Example 2.1. *Consider the game in Figure 2.1.*[2] *Consider player 2's move from b*

[1]Examples of solution concepts for abstract games relying on the notion of farsighted improving paths include the Farsighted Stable Set, the Largest Consistent Set (Chwe, 1994), the Rational Expectations Farsighted Stable Set (Dutta and Vohra, 2017) and many others.

[2]This game appeared in a previous draft of Granot and Hanany (2022).

to c. Since c is necessarily the final state of any path, and $b \succ_2 c$*, this move does not belong to any farsighted improving path. As a consequence, any solution concept that restricts attention to farsighted improving paths predicts this move to never take place. According to these solution concepts, the reason player 2 refrains from moving to c is that she prefers b over c, as if she thought "if I were not to move away from b, no one else would". However, this counterfactual seems hard to justify. Clearly, if player 2 were to refrain from moving to c, player 1 would have executed the move to a, as* $a \succ_1 b$*. Once player 2 takes into account this* ***correct counterfactual****, she* ***does*** *find it profitable to move to c, even though this move does not belong to any farsighted improving path.*

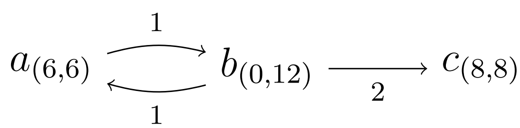


Figure 2.1

The second critique relates to the (recently deployed) "rational expectations" approach, initially due to Jordan (2006).[3] Under this approach, players are assumed to hold endogenous expectations about the continuation path that would follow each state. The main benefit of following this approach is that it allows incorporating a "maximality condition" which ensures that players make moves that are optimal for them, rather than just improving, thus tackling the "maximality critique" (see Dutta and Vohra (2017) for more details). A drawback common to all existing solution concepts adopting it, however, is that the expected continuation path from each state is assumed to be unique. This means that the uncertainty about the order of play embedded in the definition of abstract games, i.e. the fact that no such order is defined (even stochastically), is ignored. Following Granot and Hanany (2022), we refer to this critique as the *overconfidence critique*. Example 2.2 illustrates it.

Example 2.2. *Consider the game in Figure 2.2. At state b, a conflict arises: while player 2 wants to move to c, player 3 wants to move to d. The primitives of the abstract game provide no information (neither deterministic nor stochastic) on which of them will get the precedence to execute their preferred move. In the absence of such*

[3]Examples of solution concepts for abstract games utilizing the rational expectations approach can be found in Dutta and Vohra (2017), Dutta and Vartiainen (2020), Bloch and van den Nouweland (2020), Kimya (2020), Karos and Robles (2021).

information, it seems plausible that, when at a, player 1 will base her decision to move (or not) to b by comparing her payoff from state a against her payoffs from the final states of ***both*** *continuation paths (i.e. taking into account that after moving to b play might terminate at either c or d). However, solution concepts based on single-valued expectation functions assume player 1 either compares a to c alone (in case she is overly confident that the continuation path is the move to c) or to d alone (in case she is overly confident that the continuation path is the move to d).*

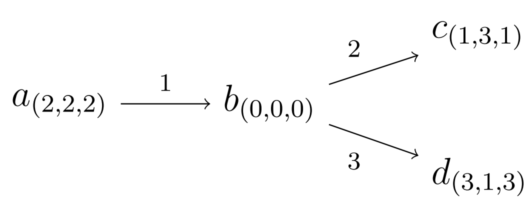


Figure 2.2

The current paper aims to propose a solution concept that captures optimal farsighted behavior and is immune to both the counterfactual and the overconfidence critiques. Our solution concept, entitled *Set-Valued Rational Expectations* (SVRE) is closely related to the above-mentioned rational expectations approach, however, in order to immunize it against the overconfidence critique, we depart from existing literature by specifying expectations which are *set-valued*. This means that from any state, players are allowed to expect *multiple* moves to be followed. In order to maintain immunity against the counterfactual critique we avoid the deployment of the notion of farsighted improving paths.

Roughly speaking, we define a *set-valued expectation* as a subset of all possible moves. Given a set-valued expectation, we define the set of *stationary states* as the set of states that no coalition is expected to move away from. For every state, we can then find the set of stationary states it might lead to by following all paths away from it that are included in the expectation and collecting all stationary states that are reached. Preferences over moves are determined based on comparisons of the sets of stationary states they lead to.[4] Given these preferences, a set-valued expectation is said to be *rational* if: *(i)* it is dynamically consistent, i.e. all players are prescribed a plan of action which they can commit to; *(ii)* it is optimal, i.e. at no state, no coalition can

[4] This requires us to specify how preferences over states are extended to preferences over *sets* of states. We impose weak conditions on this extension protocol and remain agnostic on whether players are optimistic\pessimistic\neither (i.e. whether they put more weight on good or bad outcomes).

deviate to an alternative (dynamically consistent) plan of action such that the moves it intends to execute under this alternative plan are preferred over those that it intends to execute under the original one.[5]

Our baseline results show that the SVRE concept generalizes some well-established solution concepts. In particular, we show that when players are restricted to consider only one step ahead (i.e. are myopic), the SVRE concept coincides with the core of an abstract game (which in turn, depending on how the effectivity correspondence is defined, can be shown to coincide with Nash, strong Nash, pairwise stability, pairwise-Nash, stable matching, Condorcet winner, and others). In addition, we show that when expectations are restricted to contain only one move away from each state, the SVRE concept boils down to standard (single-valued) expectation functions satisfying the conditions proposed by Ray and Vohra (2019). Our general results include sufficient conditions for existence, uniqueness, and absorption. When considering applications to specific classes of games we find the following. In perfect information extensive form games the SVRE concept boils down to subgame perfection. In strategic form games, a state (i.e. an actions profile) is supported as stationary by some SVRE if and only if it is Pareto efficient. In partition function form games, a state is supported as stationary by some SVRE if and only if it is immune to myopically beneficial deviations of coalitions that include either all players, or all players but one.

The paper proceeds as follows. Section 2.2 defines the proposed solution concept. Section 2.3 studies two benchmark cases: Myopic SVREs and "essentially single-valued" SVREs. Section 2.4 studies SVREs in their most general form. Section 2.5 applies the SVRE concept to extensive form games, strategic form games, and games partition function form games. Section 2.6 compares the SVRE concept to two closely related concepts: REEFS (Karos and Robles, 2021) and SPCS (Granot and Hanany, 2022). Section 2.7 concludes.

2.2 Defining Set-Valued Rational Expectations

2.2.1 Preliminaries

Formally, an abstract game is defined by $\Gamma = (N, Z, E, \{u_i\}_{i\in N})$, where:

- N is a set of n players.

[5] We use "plan of action" and "expectations" interchangeably.

- Z is a finite set of outcomes\states. Elements in this set are denoted by a, b, c... etc'.
- E is a correspondence from $Z \times Z$ to $\mathcal{N}$ (the set of all subsets of N) describing, for every ordered pair of states, which coalitions can replace the first by the second. If a coalition $S \in \mathcal{N}$ belongs to $E(a, b)$ we say that S "is effective in the move" (or "can move") from state a to state b.
- u_i is a function from Z to $\mathbb{R}$ describing player i's utility from each state.[6]

To define set-values rational expectations, we start by letting M be the set of all feasible moves in some abstract game Γ. Loosely speaking, a *set-valued expectation*, that is, a candidate for equilibrium, is simply a subset of all feasible moves $m \subseteq M$. To make these definitions precise, let (a, b, S) denote a move from $a \in Z$ to $b \in Z$ by coalition $S \in \mathcal{N}$. The set of all feasible moves M is defined as follows:

$$M = \left\{ (a, b, S) \;\middle|\; \begin{array}{l} S \in E(a, b), \text{ or,} \\ b = a, \; S = \{i\}, \; i \in N, \text{ or,} \\ b = a, \; \exists c \neq a \text{ s.t. } S \in E(a, c) \end{array} \right\} \tag{2.1}$$

The first clause ensures that M includes all moves specified by the effectivity correspondence. The second clause ensures that at any state it is feasible for all individual players to choose "inaction", which is formalized as a move from a state to itself. The third clause ensures that it is feasible for a coalition that can move away from a state *not* to do it, which, again, is formalized as a move by that coalition from the state to itself.[7]

To facilitate the formal definitions to come we introduce the following notation. Let $M(a) = \{(a, b, S) \in M\}$ (respectively, $m(a) = \{(a, b, S) \in m\}$) denote the set of moves from a that are in M (respectively, m), and $M_i(a) = \{(a, b, S) \in M \mid i \in S\}$ (respectively, $m_i(a) = \{(a, b, S) \in m \mid i \in S\}$) denote the set of moves from a that are in M (respectively, m) and i is involved in. m_i denotes the set of moves in m that i is involved in (from any state). $Y(m) = \{a \in Z \mid \nexists (a, b, S) \in m \text{ s.t. } b \neq a\}$ denotes

[6] An abstract game can be represented as a directed graph, where nodes represent states, and labeled directed edges represent the effectivity correspondence, i.e. a directed edge from z to z' labeled S signifies that S is effective in the move from z to z'. It is assumed that the directed graph induced by Z and E is weakly connected (otherwise we treat each graph component as a separate game).

[7] Clearly, the latter two clauses are only required when the effectivity correspondence does not specify "inaction moves" to begin with. Existing solutions for abstract games typically allow coalitions to choose "inaction" either way. Our definition of M merely formalizes this idea.

the set of stationary states under m and $Y(a,m)$ denotes the set of stationary states that are reachable from a via moves in m.[8] Players are assumed to have a preference relation between every two sets of states. These preferences are discussed in Subsection 2.2.2.

A *set-valued expectation* is a subset of M that describes all players' (intended) behavior at every state. The interpretation of the moves included in a set-valued expectation is that they are expected to be executed with some strictly positive, unknown, probability. In other words, if $(a,b,S) \in m$ then S has the *intention* to move from a to b, but this may or may not happen. This stands in contrast to standard expectation functions (as in Dutta and Vohra (2017) or Ray and Vohra (2019), for example), which specify moves that are expected to be executed for sure.

Definition 2.1. *$m \subseteq M$ is a set-valued expectation if for all $i \in N$ and $a \in Z$, $m_i(a) \neq \emptyset$. Let M^e denote the set of all set-valued expectations.*

A set-valued expectation is *dynamically consistent* if at all states $a \in Z$, for every player $i \in N$ we can find a single move in $m_i(a)$ such that i is indifferent between executing this move alone while supporting no other, and executing any of the moves in $m_i(a)$ while supporting all of them. As a consequence, in any dynamically consistent expectation, all players are indifferent among all moves they support. It can thus be interpreted as a plan of action that all players can commit to. Intuitively, if a player plans to support a set of moves among which she is not indifferent, at the moment of truth she would prefer not to adhere to the plan and instead withdraw her support from those moves that she prefers less.

Definition 2.2. *For any set-valued expectation $m \in M^e$, $m_i(a)$ is said to be dynamically consistent if for all $(a,b,S) \in m_i(a)$ there exists $(a,c,T) \in m_i(a)$ such that, letting $m' = [m \backslash m_i(a)] \cup \{(a,c,T)\}$, we have $Y(b,m) \sim_i Y(c,m')$. m_i is said to be dynamically consistent if $m_i(a)$ is dynamically consistent for all $a \in Z$. m is said to be dynamically consistent if m_i is dynamically consistent for all $i \in N$. Let M^d denote the set of all dynamically consistent set-valued expectations.*

We are now ready to state the definition of SVRE. A set-valued expectation is *rational* if it is dynamically consistent, and, no coalition can profitably deviate to any other set-valued expectation that is dynamically consistent.

[8] Formally, $Y(a,m) = \{a^K \in Z \mid \exists (a^0, ..., a^K) : a^0 = a,\ a^K \in Y(m), \text{and}, \forall k \in \{1, ..., K\},\ \exists (a^{k-1}, a^k, S^k) \in m\}$.

Definition 2.3. *A set-valued expectation $m \in M^e$ is rational (SVRE) if:*

(DC) it is dynamically consistent. Formally, $m \in M^d$;

(OP) it is optimal, i.e., no coalition T has a feasible and profitable deviation to an alternative set-valued expectation that is dynamically consistent. Formally, for any $a \in Z$, there does not exist $m' = [m \setminus R(a)] \cup \{(a, c, T)\}$, where $R(a) \subseteq m(a)$ and $(a, c, T) \in M \setminus R(a)$, such that:

(i) For all $(a, b, S) \in R(a)$, $T \cap S \neq \emptyset$;

(ii) For all $i \in T$, there exists $(a, b, S) \in m_i(a)$ such that $Y(c, m') \succ_i Y(b, m)$;

(iii) $m'_i(a)$ is dynamically consistent for all $i \in T$.

A technical remark is in place. Strictly speaking, for some choice of $R(a)$ and (a, c, T), m' may fail to belong to M^d simply because it is not a set-valued expectation, i.e. because there exists $i \in N$ for which $m'_i(a) = \emptyset$.[9] To avoid over-complicating the notation, we do not formally indicate, but nonetheless assume, that players are prescribed by default their individual inaction move. In other words, any instance of $m'_i(a) = \emptyset$ is implicitly and automatically replaced by $m'_i(a) = \{(a, a, \{i\})\}$. m' is therefore always treated as a set-valued expectation (which may or may not be dynamically consistent). We postpone interpreting the definition until after discussing extending preferences to sets of states.

2.2.2 Preferences Over Sets of States

The definitions above presuppose the existence of players' preferences over *sets* of states, while the primitives of the game only provide preferences over states. The problem of extending a preference relation over a set of states (or, more generally, "objects") to a preference relation over its power set is long-standing (see Barbera et al. (2004) for a review of this literature). Player i's preference relation over sets of states $\succcurlyeq_i$ is considered to be an extension of her preferences over states if for any two states $a, b \in Z$, $\{a\} \succcurlyeq_i \{b\}$ if and only if $u_i(a) \geq u_i(b)$.

A stronger restriction commonly imposed on $\succcurlyeq_i$ says that for any set of states $A \subseteq Z$ and state $b \in Z \setminus A$, if i prefers b over all states in A (respectively, all states in A over b) then she prefers $A \cup \{b\}$ over A (respectively, A over $A \cup \{b\}$). Intuitively, players always

[9] This happens, in particular, when $m_i(a) \subseteq R(a)$ and $i \notin T$.

prefer adding (removing) the possibility of ending up at a state that is better (worse) than those currently "on the table". This condition is usually referred to as either "dominance" or "the Gärdenfors principle". While the original Gärdenfors principle is silent about cases where i is indifferent between b and all states in A, in our formulation it is taken to imply $A \cup \{b\} \sim_i A$. The rationale is that in both cases i obtains the same level of utility for sure.[10]

In addition, we impose that any non-empty set is strictly preferred over the empty set. This corresponds to the idea that players only receive payoffs from stationary states. A scenario where no such state is reached, i.e. the set of reachable stationary states is empty, is deemed worse compared to any scenario in which *some* stationary states is reached, meaning that *some* payoff is obtained. Interpreting an abstract game as a negotiation process is one example that fits this assumption: any negotiator would prefer to put an end to a negotiation process that would have otherwise continued forever. While this assumption is not usually made explicit, it is shared by most of the existing solution concepts dealing with farsighted stability in abstract games.[11]

Assumption 2.1. *The following holds for all $i \in N$:*[12]

(i) For any non-empty set $\emptyset \neq A \subseteq Z$ and for all $b \in Z \setminus A$:

1. *If $u_i(b) = u_i(a)$ for all $a \in A$, then $A \cup \{b\} \sim_i A$*
2. *If $u_i(b) \geq u_i(a)$ for all $a \in A$ and $u_i(b) > u_i(a)$ for some $a \in A$, then $A \cup \{b\} \succ_i A$*
3. *If $u_i(b) \leq u_i(a)$ for all $a \in A$ and $u_i(b) < u_i(a)$ for some $a \in A$, then $A \succ_i A \cup \{b\}$*

(ii) For any non-empty set $\emptyset \neq A \subseteq Z$, $A \succ_i \emptyset$;

Note that the Gärdenfors principle rules out cases where players consider only the best (or worst) states they contain (such as the maxi-max or maxi-min extension rules). It does not rule out, however, cases where players consider the best (worst) states first, but in cases of indifference go on to consider the second best (worst), and so forth. Lexicographic maxi-max or lexicographic maxi-min are therefore two examples

[10] This modified version of the Gärdenfors principle is formulated in Pattanaik and Peleg (1984).

[11] In Dutta and Vohra (2017) and Ray and Vohra (2019), for instance, this assumption is implicit in the restriction that the expectation function is absorbing.

[12] We do not formally include the requirement that $\succcurlyeq_i$ is an extension of i's preferences over states because this is already implied by Condition (i).

of extension rules that are compatible with our restrictions. See Pattanaik and Peleg (1984) for their formal definitions.

2.2.3 Discussion

Condition *(OP)* requires that there does not exist an alternative to m, denoted m', which satisfies certain conditions. First, note that m' is identical to m in its prescriptions at all states but one. This is without loss of generality due to the one-shot deviation property (see Subsection 2.4.6). Let a denote the state at which m and m' differ. Compared to m, $m' = [m \setminus R(a)] \cup \{(a, c, T)\}$ removes some (set of) moves $R(a)$ and\or "adds" a (single) move (a, c, T). Note that (a, c, T) is allowed to belong to m (hence the quotation marks around "adds"). Also note that restricting attention to deviations that add only one move is inconsequential because of the dynamic consistency condition.

Condition *(OP).(i)* can be interpreted as a feasibility condition: it is *feasible* for coalition T to remove $(a, b, S) \in R(a)$ from m only if at least one member of T also belongs to S. The assumption underlying this interpretation is that unanimous agreement among all (and only) coalition members is required in order to execute a move. Hence, for T to block the execution of (a, b, S) it must have a "mole" in S, and having one such mole is sufficient.

Condition *(OP).(ii)* can be interpreted as a profitability condition: it is *profitable* for coalition T to deviate from m to m' if the move they support under m' leads to a set of stationary states that is preferred by them over the sets of stationary states that the moves they support under m lead to. Note that due to the dynamic consistency condition, the quantifier "there exists" in Condition *(OP).(ii)* is perfectly interchangeable with "for all".

Lastly, we comment on Conditions *(DC)* and *(OP).(iii)*. The former requires that the expectation under consideration is dynamically consistent. The latter restricts attention to deviations to expectations which are themselves dynamically consistent. The rationale in both cases is that farsighted players are able to foresee cases where they will fail to adhere to their own plans, and therefore avoid making such plans in the first place. Examples 2.3 and 2.4 illustrate the effects of these two conditions on the solution concept. Example 2.3 illustrates why one should require that at any given state $a \in Z$, any player i is indifferent among all the moves she supports at that state, i.e. among all moves in $m_i(a)$. Example 2.4 illustrates why one should further insist on the stronger requirement that there exists a single move in $m_i(a)$ such that i is indifferent between executing this move while supporting no other, and executing any of the moves in while supporting all moves in $m_i(a)$.

Example 2.3. *Consider the game in Figure 2.3. Clearly, when at b, player 2 wants to move to d and player 3 wants to move to e. Likewise, when at c player 2 wants to move to f. Hence, any SVRE m must include those moves. Now, assume that player 1's preferences satisfy* $\{d,e,f\} \succ_1 \{d,e\} \succ_1 \{f\} \succ_1 \{a\}$ *and fix* $m = M$.[13] *Player 3 has no incentive to remove the move from c to a, as this move opens up the possibility that she ends up with 6 or 7, which is better than ending up with 5 for sure. Whether or not m is SVRE therefore depends entirely on player 1's behavior at a.*

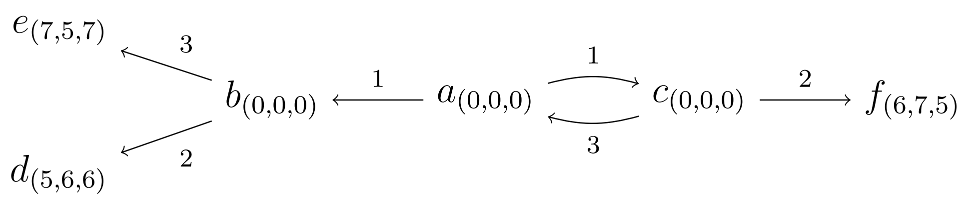


Figure 2.3

Had Conditions (DC) and (OP).(iii) were not in place, m would have been (the unique) SVRE: removing the move to b is not profitable, as $\{d,e\} \succ_1 \{f\}$*; removing the move to c is not profitable, as* $\{d,e\} \succ_1 \{f\}$*; removing both the move to b and the move to c is not profitable, as* $\{d,e\} \succ_1 \{f\} \succ_1 \{a\}$*. However, m is not dynamically consistent: even though player 1 "optimally plans" to mix between moving to b and to c, when the moment of truth arrives she always finds it better not to adhere to her plan and instead move to c for sure. This is because* $Y(c,m) = \{d,e,f\} \succ_1 \{d,e\} = Y(b,m)$*. To appreciate the absurdity of this situation note that the very reason she prefers moving to c is her conviction that, **in the future**, she **will** move to b with some strictly positive probability. But this future never arrives.*

Condition (DC) eliminates this absurd situation from the set of SVREs. Condition (OP).(iii) makes sure that deviations to this absurd situation will not be sufficient grounds to destabilize a potential SVRE. Indeed, if Player 1 is farsighted we would not expect her to deviate to a plan of action that she can see in advance she would not adhere to. With both conditions in place we obtain that $m = M \setminus \{(a,c,\{1\})\}$ *is the unique SVRE: neither removing the move to b, nor removing it and adding the move*

[13]These preferences correspond to the lexicographic maximax extension defined in Pattanaik and Peleg (1984).

to c are profitable deviations; adding the move to c (without removing the move to b) results in a plan of action that is dynamically inconsistent.

Example 2.4. *Consider the game in Figure 2.4. Clearly, when at b, Player 3 wants to move to d and Player 1 wants to move to e. Likewise, when at c Player 2 wants to move to f and Player 4 wants to move to g. Hence, any SVRE m must include those moves. Now, assume that player 1's preferences satisfy $\{d,e,f,g\} \succ_1 \{f,g\} \succ_1 \{d,e\} \succ_1 \{a\}$ and fix $m = M$. Player 3 has no incentive to remove the move from c to a, as this move opens up the possibility that she ends up with 7, which is better than ending up with 0 for sure. Similarly, Player 2 has no incentive to remove the move from b to a. Player 1 has no incentive to remove any (or all) of her moves from a, as she prefers the big set $\{d,e,f,g\}$ over any other attainable set. Thus, had Conditions (DC) and (OP).(iii) were not in place, m would have been (the unique) SVRE.*

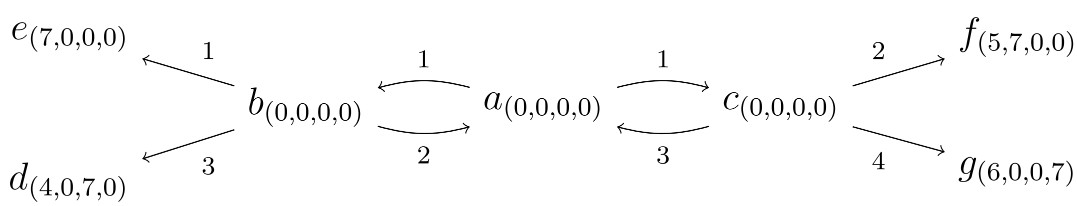


Figure 2.4

However, m is not dynamically consistent: even though Player 1 is indifferent among all her moves at a, there does not exist a single move in $m_1(a)$ such that she is indifferent between executing this move while supporting no other, and executing any of the moves in $m_1(a)$ while supporting all of them. This is because under $m_1(a)$, any move leads to $\{d,e,f,g\}$, while under any $m'_1(a)$ that includes only a single move, this single move leads to either $\{d,e\}$, or $\{f,g\}$, or $\{a\}$. To appreciate the absurdity of this situation note that the very reason Player 1 might opt to choose $m_1(a)$ is the hope that the moves she executes would be immediately undone. Condition (DC) eliminates this absurd situation from the set of SVREs and forces Player 1 to choose her plan of action at a based on a comparison between $\{d,e\}$ and $\{f,g\}$. Assuming $\{f,g\} \succ_1 \{d,e\}$, we obtain that $m = M \setminus \{(a,b,\{1\})\}$ and $m = M \setminus \{(a,b,\{1\}),(c,a,\{3\})\}$ are the only SVREs in this example.

When we restrict all players to support only one move at every state, i.e. $|m_i(a)| = 1$ for all $i \in N$ and $a \in Z$, Conditions *(DC)* and *(OP)* are trivially satisfied. Thus, imposing such a restriction offers the benefit of significantly simplifying the definition of SVRE. Example 2.5 illustrates the negative consequences of such a restriction.

Example 2.5. *Consider an abstract game such that* $M(a) = \{(a, b, \{1\}), (a, a, \{1\})\}$ *and* $M(b) = \{(b, c, \{2\}), (b, d, \{2\}), (b, b, \{2\})\}$*, that is, Player 1 has the mandate to decide whether to stay at* a *or move to* b*, and Player 2 has the mandate to decide whether to stay at* b *or move to* c*, or move to* d*. Suppose any SVRE* m *in this game satisfies the following:* $Y(c, m) \succ_2 \{b\}$*,* $Y(c, m) \sim_2 Y(d, m)$*,* $\{a\} \succ_1 Y(c, m)$*,* $\{a\} \succ_1 Y(d, m)$*,* $Y(c, m) \cup Y(d, m) \succ_1 \{a\}$*, that is, Player 2 prefers moving away from* b *but is indifferent where to, and Player 1 prefers staying at* a *unless Player 2 mixes between the moves she can effect away from* b*.*

Suppose now that we restrict set-valued expectations to prescribe each player to support only one move at every state. Then, state a *is stationary under any SVRE* m*. In the absence of this restriction, however,* a *is non-stationary for some SVRE* m*. Moreover, there is an argument to be made that* a *should be non-stationary for* ***any*** *SVRE* m*. This is because Player 2's indifference between her moves away from* b *renders her unable to make a credible commitment to follow just one.*

2.2.4 Predictions

For a given SVRE m, the moves it includes are the only moves predicted to be executed with strictly positive probability; the set $Y(m)$ is the set of states at which the game is predicted to end (the "predicted set"); and the set $Y(a, m)$ is the set of states at which the game is predicted to end assuming $a \in Z$ is the status quo. A state $a \in Z$ is said to be *supported* by an SVRE m if there exists such a m for which $a \in Y(m)$. A set of states $A \subseteq Z$ is said to be supported by some SVRE m if there exists such a m for which $A = Y(m)$.

2.2.5 Examples Revisited

To illustrate how the SVRE solution concept resolves the issues invoked by the games in Examples 2.1 and 2.2, we now apply it to them.

Example 2.1 (Continued)**.** *Consider again the game in Figure 2.1 and take* $m = \{(a, b, S) \mid a \neq b\} \cup \{(a, a, \{2\}), (c, c, \{1\})\}$*. That is,* m *contains all moves that replace a state with another and the required individual inaction moves to make it a valid* *set-valued expectation. Since* $|m_i(z)| = 1$ *for all* $i \in N$ *and* $z \in Z$*, it is also dynamically consistent. Now, at* a*, removing* $(a, b, \{1\})$ *results in* $Y(a, m') = \{a\} \not\succ_1 \{c\} = Y(a, m) = Y(b, m)$*. So this is not a profitable deviation. At* b*, removing* $(b, a, \{1\})$ *results in* $Y(b, m') = \{c\} \not\succ_1 \{c\} = Y(a, m) = Y(b, m)$*. So this is not a profitable deviation. Still at* b *removing* $(b, c, \{2\})$ *results in* $Y(b, m') = \emptyset \not\succ_2 \{c\} = Y(b, m) = Y(c, m)$*. By Assumption 2.1.(ii) this is not a profitable deviation. We established that* m *is dynamically consistent and there are no profitable deviations away from it. Hence, it is SVRE.*

To see that $\{c\}$ *is the unique predicted set supported by an SVRE* m*, first observe that* $c \in Y(m)$ *for any* m*, as it is terminal w.r.t.* M*. Suppose* m *is such that* $Y(m) = \{a, b, c\}$*.* $m' = [m \setminus \{(b, b, \{1\})\}] \cup \{(b, a, \{1\})\} \in M^d$ *is a profitable deviation for* $\{1\}$*. Suppose* m *is such that* $Y(m) = \{b, c\}$*.* $m' = [m \setminus \{(b, b, \{2\})\}] \cup \{(b, c, \{2\})\} \in M^d$ *is a profitable deviation for* $\{2\}$*. Suppose* m *is such that* $Y(m) = \{a, c\}$*. Since* $b \notin Y(m)$ *it must be that either* $(b, c, \{2\}) \in m$ *or* $(b, a, \{1\}) \in m$*, or both. Suppose* $(b, c, \{2\}) \in m$ *but* $(b, a, \{1\}) \notin m$*.* $m' = [m \setminus \{(b, b, \{1\})\}] \cup \{(b, a, \{1\})\} \in M^d$ *is a profitable deviation for* $\{1\}$*. Suppose* $(b, a, \{1\}) \in m$ *but* $(b, c, \{2\}) \notin m$*.* $m' = [m \setminus \{(b, b, \{2\})\}] \cup \{(b, c, \{2\})\} \in M^d$ *is a profitable deviation for* $\{2\}$*. Suppose* $(b, a, \{1\}), (b, c, \{2\}) \in m$*.* $m' = [m \setminus \{(b, a, \{1\})\}] \cup \{(b, b, \{1\})\} \in M^d$ *is a profitable deviation for* $\{1\}$*. Hence, no set but* $\{c\}$ *is supported by an SVRE* m*.*

Example 2.2 (Continued)**.** *Consider again the game in Figure 2.2. Let us first analyze the subgame that contains only states* b*,* c*, and* d*. If* $(b, c, \{2\}) \notin m$ *it must be that* $c \notin Y(b, m)$*. But then* $m' = [m \setminus \{(b, b, \{2\})\}] \cup \{(b, c, \{2\})\}$ *is a feasible and profitable deviation which does not violate dynamic consistency. Hence, if there exists a SVRE* m *it must contain* $(b, c, \{2\})$*. By the same reasoning, it must also contain* $(b, d, \{3\})$*. Hence, any SVRE* m *satisfies* $Y(b, m) = \{c, d\}$*. This implies that player 1's behavior at state* a *must be dictated by her preference relation between* $\{a\}$ *and* $\{c, d\}$*. In other words, player 1 necessarily takes into account both continuation paths. Note that our assumptions on preferences over sets allow for player 1 to hold any preference relation over these two sets.*

We also note that by Assumption 2.1.(i) we have $\{c\} \succ_2 \{c, d\}$ *and* $\{d\} \succ_3 \{c, d\}$*. Hence, any SVRE* m *must also satisfy* $(b, b, \{2\}), (b, b, \{3\}) \notin m$*. The interpretation is that when at* b*, any player that gets the opportunity to move surely takes advantage of it. In the absence of this assumption the counterintuitive prediction that players choose to pass an opportunity to move away from* b *and by doing so risk ending up at their less* *preferred state could have been produced by an SVRE m (for instance under maxi-max preferences, which imply $\{c\} \sim_2 \{c,d\}$ and $\{d\} \sim_3 \{c,d\}$).*

2.3 Benchmarks

2.3.1 Essentially Single-Valued SVREs

As a benchmark, and in order to relate our solution concept to existing literature, we start by analyzing set-valued expectations which are "essentially single-valued", i.e. that prescribe at most one move away from every state. Formally, $m \in M^e$ is *essentially single-valued* if for all $a \in Z$, $|\{(a,b,S) \in m \mid b \neq a\}| \leq 1$.[14] Proposition 2.1 provides an alternative characterization of essentially single-valued SVREs. More particularly, it shows that essentially single-valued SVREs are equivalent to "standard" expectation function satisfying the conditions proposed in Ray and Vohra (2019). We first reiterate those conditions here. A standard (history-independent) expectation function σ is an object that maps every state to a new state, along with the coalition making the move.[15] Formally, for each $z \in Z$, $\sigma(z) = \{f(z), S(z)\}$, where $f(z)$ is the state that follows z and $S(z) \in E(z, f(z))$ is the coalition implementing the change. (If $f(z) = z$, then $S(z)$ is empty, and this is interpreted as "nothing happens"). Note that σ induces a unique continuation chain from every state. $x^\sigma(z)$ denotes the final\absorbing state in this unique chain (assuming it is finite). The conditions Ray and Vohra (2019) propose to impose on σ are as follows:

- *Absorption.* For any state z, the continuation path σ prescribes terminates at a stationary state, i.e. some z' satisfying $f(z') = z'$.
- *Coalitional Acceptability.* For any state z, $u_i(x^\sigma(z)) \geq u_i(z)$ for all $i \in S(z)$.[16]
- *Absolute Maximality.* For any state z, there does not exist $T \in E(z, z')$ such that $u_i(x^\sigma(z')) > u_i(x^\sigma(z))$ for all $i \in T$.

[14] We call such set-valued expectations *essentially* single-valued, rather than simply "single-valued", because they are not truly single-valued: they may include multiple moves from a state, as long only one of them is to a *different* state (meaning that the rest are inaction moves).

[15] They call σ a "negotiation process" rather than an expectation function.

[16] This requirement is weaker than the standard "external stability" condition, which requires players in S to be strictly better-off in the final state $x^\sigma(z)$ compared to the status quo z.

For any essentially single-valued m let its corresponding "standard" expectation function be denoted σ^m and define it as follows. For every $a \in Z$, $\sigma^m(a) = \{f(a), S(a)\}$, where if there exists $(a, b, S) \in m(a)$ such that $a \neq b$, then $f(a) = b$ and $S(a) = S$, and if there does not exist such (a, b, S) then $f(a) = a$ and $S(a) = \emptyset$. Conversely, for any (history-independent) expectation function σ let m^σ be defined as follows. For every $a \in Z$, for all $i \in N \setminus S(a)$, $m_i^\sigma(a) = \{(a, a, \{i\})\}$ and for all $i \in S(a)$, $m_i^\sigma(a) = \{(a, f(a), S(a))\}$.

Proposition 2.1. *If m is an essentially single-valued SVRE then σ^m is absorbing, coalitionally acceptable and absolutely maximal. Conversely, if σ is absorbing, coalitionally acceptable and absolutely maximal then m^σ is an essentially single-valued SVRE.*

Proof. $\implies$. Throughout the proof, we let (a, b, S) denote the unique move away from a prescribed by m (if exists).

Suppose m is an essentially single-valued SVRE but σ^m is not absorbing. Take $a \in Z$, such that $Y(a, m) = \emptyset$ (such a must exist given that σ^m is not absorbing). Consider $m' = [m \setminus \{a, b, S\}] \cup \{(a, a, S)\}$. It is clearly a feasible deviation for S and does not cause a violation of dynamic consistency. In addition, it is profitable for S because $Y(a, m') = \{a\} \succ_i Y(a, m) = \emptyset$ for all $i \in S$. Hence, m is not an essentially single-valued SVRE. Contradiction. If m is an essentially single-valued SVRE σ^m must be absorbing.

Suppose m is an essentially single-valued SVRE but σ^m is not coalitionally acceptable. Take $a \in Z$, such that $\{a\} \succ_i Y(a, m) \neq \emptyset$ for all $i \in S$ (such a must exist given that σ^m is not coalitioanlly acceptable). Consider $m' = [m \setminus \{a, b, S\}] \cup \{(a, a, S)\}$. It is clearly a feasible deviation for S and does not cause a violation of dynamic consistency. In addition, it is profitable for S because $Y(a, m') = \{a\} \succ_i Y(a, m) = \emptyset$ for all $i \in S$. Hence, m is not an essentially single-valued SVRE. Contradiction. If m is an essentially single-valued SVRE σ^m must be coalitionally acceptable.

Suppose m is an essentially single-valued SVRE but σ^m is not absolutely maximal. Take $a \in Z$, such that there exist $T \in E(a, c)$ and $Y(c, m) \succ_i Y(a, m)$ for all $i \in T$ (such a must exist given that σ^m is not absolutely maximal). Consider $m' = [m \setminus \bigcup_{i \in T} m_i(a)] \cup \{(a, c, T)\}$. It is clearly a feasible deviation for T and does not cause a violation of dynamic consistency. In addition, it is profitable for T because $Y(c, m') = Y(c, m) \succ_i Y(a, m)$ for all $i \in T$. Hence, m is not an essentially single-valued SVRE. Contradiction. If m is an essentially single-valued SVRE σ^m must be absolutely maximal.

$\Longleftarrow$. Suppose σ is absorbing, coalitionally acceptable and absolutely maximal. First, note that m^σ is essentially single-valued. Due to absorption, m^σ satisfies $|Y(a, m^\sigma)| = 1$ for all $a \in Z$, which implies that Condition (DC) holds. Due to coalitional acceptability, at no state $a \in Z$, no coalition can benefit from removing any of its prescribed moves. Due to absolute maximality, at not state, no coalition can benefit from adding any of the moves it can effect. Condition (OP) therefore also holds, meaning that m^σ is an essentially single-valued SVRE. □

Our next result relates essentially single-valued SVREs to farsighted stable sets. Stating it requires the following definitions. A *farsighted improving path* is a finite sequence of states and coalitions $\{z^0, S^1, z^1, ..., S^K, z^K\}$ such that for all $1 \geq k \geq K$: *(i)* $S^k \in E(z^{k-1}, z^k)$; (ii) for all $i \in S^k$, $u_i(z^K) < u_i(z^{k-1})$. A set of states $Y \subseteq Z$ is a *Farsighted Stable Set* if: *(i)* there does not exist a farsighted improving path between any two states within Y; *(ii)* from every state not in Y there exists a farsighted improving path terminating at some state within Y. A set of states $Y \subset Z$ is *single payoff* if $u_i(a) = u_i(b)$ for all $a, b \in Y$ and all $i \in N$. Dutta and Vohra (2017) showed that there exists a rational expectation function supporting a single-payoff set Y as the set of stationary states if and only if Y is a Farsighted Stable Set. As Ray and Vohra (2019) point out (see proof of their proposition 1), when the final payoff is unique, no coalition can find a profitable deviation, so absolute maximality also holds. We can therefore apply Proposition 2.1 to obtain the following result.

Proposition 2.2. *If $Y \subseteq Z$ is a single-payoff farsighted stable set then there exists an essentially single-valued SVRE m that supports it, i.e. one for which $Y(m) = Y$.*

2.3.2 Myopic SVREs

An additional benchmark case worth discussing before turning to the main analysis is that of *myopic* SVREs. The SVRE solution concept reflects foresight in that players take into account *chains* of moves. That is, they evaluate the relative attractiveness of being at a state a by the set of stationary states it *eventually* leads to, i.e. $Y(a, m)$. One could entertain a myopic version of this solution concept where, instead, players evaluate the relative attractiveness of being at a state a simply by the utility that *it* provides. With this in mind, we define *myopic SVRE* by replicating the original SVRE definition but replacing $Y(b, m)$ by b and $Y(c, m')$ by c. The dynamic consistency conditions are omitted because myopic players do not take into account any dynamics.[17]

Definition 2.4. *A set-valued expectation $m \in M^e$ is myopically rational (myopic SVRE) if no coalition T has a feasible and myopically profitable deviation to an alternative set-valued expectation. Formally, for any $a \in Z$, there do not exist $R(a) \subseteq m(a)$ and $(a, c, T) \in M \setminus R(a)$, such that:*

(i) For all $(a, b, S) \in R(a)$, $T \cap S \neq \emptyset$.

(ii) For all $i \in T$, there exists $(a, b, S) \in m_i(a)$ such that $u_i(c) > u_i(b)$;

Proposition 2.3 relates myopic SVREs to the weak and strong cores. The *weak core* of an abstract game, denoted $WC(\Gamma)$, is the set of states such that no coalition can move away from them to a state that all of its members strictly prefer. Formally, $WC(\Gamma) = \{a \in Z \mid \nexists T \in E(a, c) \text{ s.t. } \forall i \in T,\ u_i(c) > u_i(a)\}$. Analogously, the *strong core*, denoted $SC(\Gamma)$, is the set of states such that no coalition can move away from them to a state that all of its members weakly prefer and at least one strictly prefers. Formally, $SC(\Gamma) = \{a \in Z \mid \nexists T \in E(a, c) \text{ s.t. } \forall i \in T,\ u_i(c) \geq u_i(a),\ \exists i \in T,\ u_i(c) > u_i(a)\}$.

The proposition asserts that myopic SVREs necessarily support states that are in the strong core and may support any state that is in the weak core. Indeed, there are no compelling reasons to hold any state in $WC(\Gamma) \setminus SC(\Gamma)$ as myopically stable, nor as myopically unstable. The myopic SVRE concept avoids making an arbitrary choice in either direction and simply allows for both possibilities. To state the proposition, we let S^p denote the set of ms which are myopic SVREs.

Proposition 2.3. *In any abstract game, any $m \in S^p$ satisfies $SC(\Gamma) \subseteq Y(m) \subseteq WC(\Gamma)$. Moreover, there exist $m, m' \in S^p$ such that $Y(m) = SC(\Gamma)$ and $Y(m') = WC(\Gamma)$.*

Proof. We first show that if a state is stationary under a myopic SVRE then it belongs to the weak core, i.e. $Y(m) \subseteq WC(\Gamma)$ for any $m \in S^p$. We then show that if a state belongs to the strong core then it is stationary under any $m \in S^p$, i.e. $SC(\Gamma) \subseteq Y(m)$ for any $m \in S^p$. We then show how to construct myopic SVREs that support as stationary exactly the sets of weak\strong core states.

[17]Moreover, the myopic version of dynamic consistency is already implied by the inexistence of feasible and myopically profitable deviations.

Suppose $a \in Z$ is supported as stationary by some myopic SVRE m. Since a is stationary all moves in $m_i(a)$ (for any $i \in N$) are to a itself. Keeping that in mind and choosing $R(a) = \emptyset$, by the definition of myopic SVRE there does not exist $(a, c, T) \in M$ such that for all $i \in T$, $u_i(c) > u_i(a)$. Hence, it belongs to the weak core. This establishes $Y(m) \subseteq WC(\Gamma)$ for any $m \in S^p$.

Next, take $a \in SC(\Gamma)$ and suppose m does not support it as stationary, i.e. it prescribes some move away from a. Denote some such move by (a, b, S). Since a is in the strong core we know that there exists $i \in S$ such that $u_i(a) > u_i(b)$. But this means it is both feasible and profitable for $\{i\}$ to object to (a, b, S) and support $(a, a, \{i\})$ instead. Hence, m cannot be SVRE. Any state in the strong core must be supported by any SVRE m, i.e. $SC(\Gamma) \subseteq Y(m)$ for any $m \in S^p$

Next, we need to show how to construct myopic SVREs that support as stationary exactly the set of weak and strong core states. Algorithm 2.1 shows how to do the former.

Algorithm 2.1 Construct a myopic SVRE that supports all states in the weak core

Set $m(a) = \emptyset$ for all $a \in Z$
for $a \in Z$:
  for $i = [1, 2, ..., N]$:
    Order all moves in $M_i(a)$ in a column, from most to least preferred (in a myopic sense). Whenever indifferences between moves *away* from a arise, order them randomly. Whenever indifferences between an "inaction move" and a move away from a arise, place an individual inaction move first.
  Concatenate all n columns to create a table T.
  Let $L = \max_{i \in N} |M_i(a)|$ denote the longest column of T.
  for $(k, i) = [(1, 1), ...(1, n), (2, 1), ..., (L, n)]$:
    for $(q, j) = [(1, 1), ...(1, n), (2, 1)..., (j, 1), ..., (j, k)]$:
      Whenever the cell $T(k, i)$ contains some move, let (a, b, S) denote the move it contains (if $T(k, i)$ does not contain a move, just continue to the next (k, i)).
      Count the number of times the condition $T(q, j) = (a, b, S)$ & $m_j(a) = \emptyset$ is satisfied.
      if the count reaches $|S|$:
        Update $m(a) = m(a) \cup \{(a, b, S)\}$
return m

We argue that this process concludes in a myopic SVRE that supports all, and only, states that belong to the weak core. First, note that having $(a, a, \{i\}) \in M_i(a)$ for every $i \in N$ and $a \in Z$ guarantees that $m_i(a) \neq \emptyset$ for all players and states, meaning that the process results in a set-valued expectation. Second, the way we constructed m guarantees that $m_i(a)$ contains i's most preferred move (myopically speaking) such that she can enlist all coalition members to execute. Hence, m is a myopic SVRE. Third, since in cases of indifferences we have placed individual inaction moves above others, all moves such that a player is indifferent between executing or not will *not* be included in m. This means that all states in the weak core are supported as stationary under m. Fourth, since every state outside the weak core has a move away from it such that all members of the moving coalition strictly prefer over the status quo, no such state is supported as stationary under m. This establishes that the constructed m is a myopic SVRE satisfying $Y(m) = WC(\Gamma)$.

To construct a myopic SVRE m that supports exactly the set of states that belong to the strong core we follow the same procedure as above, only that in cases of indifferences all inaction moves are placed below others. This guaranteed that when some players are indifferent between executing a move or not, but others strictly prefer to execute it, it *will* be included in m. In turn, this implies that only states in the strong core are supported as stationary. The rest of the argument remains intact.

□

To appreciate the strength of Proposition 2.3, we note that the core of abstract games coincides with a wide variety of well-established solution concepts. More particularly, for an appropriately defined effectivity correspondence, the core can be shown to coincide with Nash equilibrium, strong Nash, pairwise stability, pairwise Nash, stable matching, Condorcet winner, and others. These results originate from the fact that solution concepts applied to abstract games are only meant to ensure optimal behavior, whereas the object defining the set of possible deviations (or "moves", in the language of abstract games), i.e. the effectivity correspondence, is entirely separate and taken as a primitive of the game. The traditional approach, on the other hand, endows solution concepts with a double burden: ensuring optimal behavior *and* defining the set of possible deviations. Under the traditional framework, any change in the set of considered deviations necessitates defining a new solution concept. As a result, we are left with an array of solution concepts that differ from one another in little conceptual content. The abstract games approach may offer a cure to this problem.

2.4 SVRE

We now turn to SVREs in their general form.

2.4.1 Absorption

In Subsection 2.3.1 we saw that an essentially single-valued SVRE is necessarily absorbing. Does this result continue to hold when the set-valued expectation is not restricted to prescribe at most one move away from each state? As Example 2.6 illustrates, the answer is no. In this example, while both players would have been better off by settling on having *some* stationary state (recall that players strictly prefer any non-empty set of stationary states over the empty one), they may lack the ability to coordinate on a deviation that would give rise to one.

Example 2.6. *Consider the game in Figure 2.5 and take $m = M$. There does not exist a coalition that can remove enough moves from m so as to make one of the states stationary. Hence, there are no profitable deviations from this m and it is SVRE. Nonetheless, it is not absorbing.*

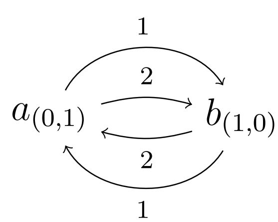


Figure 2.5

Proposition 2.4 provides a sufficient condition on the set of feasible moves to guarantee that any SVRE m is absorbing. To state the proposition, let $N(a) = \{i \in N \mid \exists (a, b, T) \in M \text{ s.t. } a \neq b, i \in T\}$ denote the set of *active players* at a, i.e. those that, when at a, need to choose whether or not to support a move away from it.

Proposition 2.4. *Suppose that from every state $a \in Z$ there exists a sequence of moves in M leading to a state $a \in Z$ that is either terminal with respect to M (i.e. $a \in Y(M)$) or satisfies $(a, a, N(a)) \in M$. Then, if m is SVRE then it is absorbing, i.e. we have $Y(a, m) \neq \emptyset$ for any $a \in Z$.*

Proof. Let $P_0 = \{a \in Z \mid a \in Y(M),\text{ or, } (a,a,N(a)) \in M\}$, and for any $k \geq 1$ recursively define $P_k = \{a \in Z \mid \exists (a,b,S) \in M,\text{ s.t. } b \in P_{k-1}\}$. We prove by induction that for any k, if m is SVRE then we have $Y(a,m) \neq \emptyset$ for any $a \in P_k$.

Base case. Suppose m is SVRE and consider P_0. Clearly, any state $a \in P_0$ such that $a \in Y(M)$ necessarily satisfies $Y(a,m) = \{a\}$. For any other state $a \in P_0$ we necessarily have $Y(a,m) \neq \emptyset$, as otherwise $m' = [m \setminus m(a)] \cup (a,a,N(a))$ would have been a feasible and profitable deviation for $N(a)$ (which does not violate dynamics consistency). The existence of $(a,a,N(a))$ in M follows from the definition of P_0. The feasibility of this deviation is immediate from the fact that $N(a)$ includes all players that are involved in all moves away from a. The profitability of this deviation follows from Assumption 2.1.(ii), according to which $\{b\} \succ_i \emptyset$ for all $i \in N$. m' does not violate dynamic consistency because under m' each player is supports a single move at a. Hence, if m is SVRE we have $Y(a,m) \neq \emptyset$ for all $a \in P_0$.

Induction step. Suppose m is SVRE and assume $Y(a,m) \neq \emptyset$ for all $a \in P_{k-1}$. For any $a \in P_k$ we necessarily have $Y(a,m) \neq \emptyset$, as otherwise $m' = [m \setminus \bigcup_{i \in S} m_i(a)] \cup (a,b,S)$ such that $b \in P_{k-1}$ would have been a feasible and profitable deviation for S (which does not violate dynamics consistency). The existence of (a,b,S) such that $b \in P_{k-1}$ in M follows from the definition of P_k. The feasibility of this deviation is immediate from the fact that the moves that are removed are those members of S are involved in. The profitability of this deviation follows from Assumption 2.1.(ii) as well as the induction assumption. That is, the fact that players always prefer a nonempty set over an empty one and that $Y(b,m) \neq \emptyset$. m' does not cause a violation of dynamic consistency because under it members of S support a single move at a. Hence, if m is SVRE and $Y(a,m) \neq \emptyset$ for all $a \in P_{k-1}$, we have $Y(a,m) \neq \emptyset$ for all $a \in P_k$.

This induction argument establishes that if m is SVRE we have $Y(a,m) \neq \emptyset$ for any $a \in \bigcup_{k=0}^{\infty} P_k$. To complete the proof we need to show $\bigcup_{k=0}^{\infty} P_k = Z$. Indeed, this follows from the assumption that from every state $a \in Z$ there exists a sequence of moves in M leading to a state in P_0. □

2.4.2 External Stability

Just as we asked whether the absorption property carries over to general SVREs, we may ask whether the external stability\coalitional acceptability property carries over. Example 2.1 already answered this question negatively. In this example, while $m = M$ is SVRE, the paths it prescribes from b are not farsightedly improving (as they include

$(b, c, \{2\})$). The SVRE solution concept departs from the majority of existing farsighted solution concepts in that it produces predictions that involve the execution of moves that do not lie on any farsighted improving path. Indeed, we argue that executing such moves is not only compatible with farsighted behavior but sometimes, as in Example 2.1, implied by it. Proposition 2.1 on the other hand showed one instance where external stability *should* be expected to hold (in the sense that it is implied by farsighted and utility-maximizing behavior): when players expect no more than one move away from every state.

2.4.3 Existence

Proposition 2.4 provides a condition under which $Y(m) \neq \emptyset$ for any SVRE m. It does not guarantee, however, that an SVRE m exists. The two following propositions provide sufficient conditions for existence. Proposition 2.5 says that an SVRE m exists in any game such that no strict subset of players has the power to impose any state to be stationary. Moreover, it asserts that under this condition every singleton set containing a weakly Pareto efficient that is reachable from all other states via moves in M can be supported as stationary by some SVRE m. Proposition 2.6 says that an SVRE m exists in any game that is acyclic such that only one coalition can move away from each non-terminal state.

Proposition 2.5. *Suppose that for all $b \in Z$ and $S \subset N$, if $(b, b, S) \in M$, then there exists $(b, c, T) \in M$ such that $b \neq c$ and $S \cap T = \emptyset$. Then, for any weakly Pareto efficient state $a \in Z$ that is reachable from all other states via moves in M, there exists an SVRE m such that $Y(z, m) = \{a\}$ for all $z \in Z$.*

Proof. Let $a \in Z$ be a weakly Pareto efficient state that is reachable from all other states via moves in M and let $m = M \setminus \{(a, z, S) \in M \mid z \neq a\}$, i.e. m contains all feasible moves besides those that replace a by another state. a is terminal w.r.t. this m. Moreover, since it is reachable from all other states, we have $Y(z, m) = \{a\}$ for all $z \in Z$. We want to show that this m is SVRE.

First, since $Y(z, m) = \{a\}$ for all $z \in Z$, m is dynamically consistent. Hence, our task is to show that there does not exist m' satisfying conditions (OP).(i)-(iii).

Consider $b \in Z$, $b \neq a$, and note that $m(b) = M(b)$ and $Y(b, m) = \{a\}$. A deviation at b cannot turn stationary any state other than itself. If it does not turn b to a stationary state we have $Y(b, m) = Y(b, m') = \{a\}$ for any m' that differs from m at b, so this cannot be a profitable deviation. By the assumption in the proposition, turning b to a stationary state is not feasible for any coalition but, perhaps, N. On the other hand, turning b to a stationary state is profitable only if all members of the deviating coalition strictly prefer b over a. Since a is weakly Pareto efficient there does not exist $b \in Z$ such that all $i \in N$ strictly prefer it over a. Hence, no deviation turning b to a stationary state is both feasible and profitable. We conclude that no feasible and profitable deviations at any $b \neq a$ exist.

Next, we show that there are no profitable deviations from a. A deviation at a may only turn a to a non-stationary state, i.e. for any m' that differs from m at a, either $Y(a, m') = \{a\}$, or $Y(a, m') = \emptyset$. Neither option can make any player strictly better off compared to m, so no profitable deviations from $m(a)$ exist. Hence, m is SVRE. □

To state the next proposition, we denote by $\mathcal{N}(a) = \{S \in \mathcal{N} \mid \exists (a, b, S) \in M \text{ s.t. } a \neq b\}$ the set of *active coalitions* at a, i.e. those that, when at a, need to choose whether or not to support a move away from it. In addition, we define a game as *acyclic* if whenever state a is reachable from b via (a sequence of) moves in M, b is not reachable from a.

Proposition 2.6. *Suppose the game is acyclic. If $|\mathcal{N}(a)| \leq 1$ for all $a \in Z$, then there exists a SVRE m.*

Proof. We show that when the game is acyclic SVRE ms can be constructed by backward induction. For any $z \in Z$, let $\Gamma(z)$ denote the subgame starting at z and let $l(\Gamma(z))$ be the length of this game (the length of the path from z to the farthest terminal state). For any b such that $l(\Gamma(b)) = 1$, letting S denote the unique active coalition at b, set $m(b) = \{(b, a, S)\}$ where $(b, a, S) \in M$ and a is not Pareto dominated for S by any other a' such that $(b, a', S) \in M$ (note that (a, a, S) is included in M, and recall that we do not explicitly specify all the "inaction moves" required to make m a set-valued expectation, but they are implicitly assumed). Since (b, a, S) is not strictly Pareto dominated by any other move S can effect, m is SVRE in each subgame.

Now consider all c such that $l(\Gamma(c)) = 2$, and let T denote the unique active coalition at c. Taking into account the previous step, set $m(c) = \{(c, b, T)\}$ where $(c, b, T) \in M$ and the payoff from b, i.e. $Y(b, m)$, is not Pareto dominated for T by any other b' such that $(c, b', T) \in M$. Again, this m is SVRE in each subgame. Continuing this way until all states are exhausted, we obtain a m that is SVRE in the entire game. □

2.4.4 Inexistence

Example 2.7 illustrates a case of inexistence.[18]

Example 2.7. *Consider the game in Figure 2.6. All possible set valued expectations are given by:*

$$m^1 = \{(a,a,\{1\}),(a,a,\{2\}),(b,b,\{1\}),(b,b,\{2\}),(c,c,\{1\}),(c,c,\{2\})\}$$
$$m^2 = \{(a,a,\{1\}),(a,b,\{2\}),(b,b,\{1\}),(b,b,\{2\}),(c,c,\{1\}),(c,c,\{2\})\}$$
$$m^3 = \{(a,b,\{1\}),(a,b,\{2\}),(b,b,\{1\}),(b,b,\{2\}),(c,c,\{1\}),(c,c,\{2\})\}$$
$$m^4 = \{(a,b,\{1\}),(a,a,\{2\}),(b,b,\{1\}),(b,b,\{2\}),(c,c,\{1\}),(c,c,\{2\})\}$$

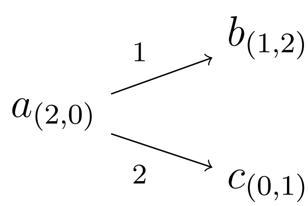


Figure 2.6

From m^1, coalition $\{2\}$ has a feasible and profitable deviation to m^2 (as $Y(a,m^2) = \{c\} \succ_2 Y(a,m^1) = \{a\}$). From m^2, coalition $\{1\}$ has a feasible and profitable deviation to m^3 (as $Y(a,m^3) = \{b,c\} \succ_1 Y(a,m^2) = \{c\}$). From m^3, coalition $\{2\}$ has a feasible and profitable deviation to m^4 (as $Y(a,m^4) = \{b\} \succ_2 Y(a,m^3) = \{b,c\}$). From m^4, coalition $\{1\}$ has a feasible and profitable deviation to m^1 (as $Y(a,m^1) = \{a\} \succ_1 Y(a,m^4) = \{b\}$). Hence, there is an inevitable cycle of profitable deviations and no SVRE exists.

2.4.5 Pareto Efficiency

Proposition 2.5 provided conditions under which every weakly Pareto efficient state can be supported as stationary by some SVRE m. We may ask under what conditions *only* Pareto efficient states are supported by SVREs. The propositions in this subsection offer several answers to this question.

[18] I thank Eran Hanany for pointing out this example.

Proposition 2.7. *Suppose $A \subseteq Z$ satisfies the following conditions:*

(i) All $a \in A$ strictly Pareto dominate all $b \in Z \setminus A$;

(ii) Every $a \in A$ satisfies $a \in Y(M)$ or $(a, a, N(a)) \in M$ and is reachable via moves in M from any $b \in Z \setminus A$.

Then, any SVRE m satisfies $Y(m) \subseteq A$.

Proof. Let $A \subseteq Z$ satisfy all the conditions in the proposition. We need to show that no SVRE m supports any $b \in Z \setminus A$. Assume by contradiction that m is SVRE but there exists $b \in (Z \setminus A) \cap Y(m)$. We first argue that $Y(m)$ must also include some $a \in A$. We then establish a contradiction by showing that this implies $b \notin Y(m)$.

Suppose $Y(m) \cap A = \emptyset$. Note that must mean that no $a \in A$ satisfies $a \in Y(M)$, implying that every $a \in A$ satisfies $(a, a, N(a)) \in M$. Now, take some $a \in A$ and consider $m' = [m \setminus \bigcup_{i \in N(a)} m_i(a)] \cup \{(a, a, N(a))\}$. We argue that this is a feasible and profitable deviation for $N(a)$ which does not cause a violation of dynamic consistency. It is feasible because only moves players in $N(a)$ are involved in are removed. It is profitable because $Y(a, m) \subseteq Z \setminus A$ while $Y(a, m') = \{a\}$, and a strictly Pareto dominates all states in $Z \setminus A$. It does not cause a violation of dynamic consistency because under m' every player is supporting a single move at a. Hence, if m is SVRE and some $b \notin A$ is supported as stationary, some $a \in A$ must also be supported as stationary.

On the other hand, if $a \in A$ is supported as stationary, all states from which a can be reached and are strictly Praeto dominated by it must be non-stationary. But b is such a state. Contradiction. If m is SVRE then $Y(m) \subseteq A$.

□

Corollary 2.1. *Suppose $a \in Z$ satisfies $a \in Y(M)$ or $(a, a, N(a)) \in M$, is reachable from all other states via moves in M, and strictly Pareto dominates every such state. Then, $\{a\}$ is the unique set supported by an SVRE m.*[19]

Proof. Let $a \in Z$ be a state that strictly Pareto dominates all other states. By Proposition 2.7, $\{a\}$ is the unique candidate to be supported by an SVRE m. Hence, we just need to show that there exists an SVRE m that supports it.

Consider $m = M \setminus \{(a, b, S) \in M \mid b \neq a\}$, i.e. m contains all feasible moves besides those that replace a by a different state. a is terminal w.r.t. this m. Since a is reachable

[19] See Theorem 7 in Herings et al. (2009) for a similar result.

via moves in M from all states and m contains all feasible moves from states different than a, we have $Y(m) = Y(z, m) = \{a\}$ for all $z \in Z$. Since no player strictly prefers any set over $\{a\}$, no profitable deviations exist. In addition, since $|Y(z, m)| = 1$ for all $z \in Z$, m is dynamically consistent. Hence, m is SVRE. □

2.4.6 One-Shot Deviation Property

The optimality condition in the definition of SVRE requires immunity to one-shot deviations, i.e. to deviations that differ from the original expectations in their prescriptions *at a single state*. Proposition 2.8 shows that restricting attention to one-shot deviations is without loss of generality, as immunity to one-shot deviations implies immunity to all deviations. To state the proposition, we let M_T denote the set of moves that are feasible for coalition T and $M_T(a)$ be the set of such moves from a. In addition, we define m as *globally optimal* if there does not exist $m' = [m \setminus R] \cup A$, where $R \subseteq m$ and $A \subseteq M_T \setminus R$, such that: *(i)* For all $(a, b, S) \in R$, $T \cap S \neq \emptyset$; *(ii)* For all $(a, c, T) \in A$, for all $i \in T$, there exists $(a, b, S) \in m_i$ such that $Y(c, m') \succ_i Y(b, m)$; *(iii)* $m' \in M^d$. Note that unlike the definition of (plain) optimality, this definition allows m' to differ from m in its prescriptions at multiple states.

Proposition 2.8. *If m is SVRE then it is globally optimal.*

Proof. Assume by contradiction that m does not satisfy global optimality but is SVRE, i.e. satisfies dynamic consistency and (plain) optimality. This means that there exists a dynamically consistent m' that differs from m in more than one state and satisfies requirements (i), (ii) and (iii) of the global optimality condition, but there does not exist a m'' that differs from m is a single state and satisfies requirements (i), (ii) and (iii) of the (plain) optimality condition. Let T denote the coalition that deviates from m to m'.

Case 1. Suppose there exists a state a that is non-stationary under m but is stationary under m'. Then, $\{a\} \succ_i Y(a, m)$ for all $i \in T$. But then $m'' = [m \setminus \bigcup_{i \in T} m_i(a)] \cup \{(a, a, T)\}$ differs from m in a single state and satisfies requirements (i), (ii) and (iii) of the (plain) optimality condition. Contradiction.

Case 2. Suppose there does not exist a state that is non-stationary under m but is stationary under m'. Let $\Pi(a, m) = \{(a^0, a^1, a^2, ..., a^K) \mid a^0 = a, \text{ and, for all } k = 0, 1, ...K-1 \text{ there exists } (a^k, a^{k+1}, S) \in m)\}$ be the set of sequences of states originating from a that are supported by m. We establish the existence of a move $(a, c, T) \in M$

such that $m'(a) \neq m(a)$ and $\Pi(c, m') = \Pi(c, m)$ and then use it to construct a feasible and profitable one-shot deviation. First, since m' is a deviation from m there must exist $a \in Z$ such that $m'(a) \neq m(A)$. Second, note that $Y(m') \neq \emptyset$ (otherwise m' could not have been a profitable deviation) and $Y(m') \subseteq Y(m)$ (since we assume there does not exist a state that is non-stationary under m but is stationary under m'). This implies the existence of $d \in Z$ such that $\Pi(d, m') = \Pi(d, m)$. In particular, $\Pi(d, m') = \Pi(d, m)$ holds for any $d \in Y(m')$. If the hypothesized move (a, c, T) does not exist it must be that any move by the deviating coalition $(a, c, T) \in m'$ satisfies $Y(c, m') = \emptyset$. But this contradicts m' being a profitable deviation for T.

Now, let $(a, c, T) \in M$ be a move satisfying $m'(a) \neq m(a)$ and $\Pi(c, m') = \Pi(c, m)$. We establish a contradiction by showing that $m'' = [m \setminus \bigcup_{i \in T} m_i(a)] \cup \{(a, c, T)\}$ is a feasible and profitable one-shot deviation that does not violate dynamic consistency. It is feasible because m'' differs from m only in moves members of T are involved in. It is profitable because $Y(c, m'') = Y(c, m')$, and m' is a profitable deviation. It does not violate dynamic consistency because m'' is identical to m at all states but a (and m is dynamically consistent) and at a m'' prescribes the support of only one move for all members of T. □

2.5 Applications

2.5.1 Strategic Form Games

In order to analyze strategic form games using the SVRE solution concept we first transform them into abstract games. The general idea is to treat each combination of actions as a state, and set players to be effective in the move from one state to another only if they differ only in the action of that player. Figure 2.7 provides a simple example. Formally, a strategic form game is a triple $(N, \{Z_i\}_{i \in N}, \{u_i\}_{i \in N})$, where N is the set of players and for $i \in N$, Z_i is the nonempty set of strategies of player i and u_i is player i's payoff function, $u_i : Z_N \to \mathbb{R}$, where for $S \subseteq N$, Z_S denotes the Cartesian product of Z_i over $i \in S$, i.e., $Z_S = \Pi_{i \in S} Z_i$. The set of players in the corresponding effectivity function form is N. The set of states is $Z = Z_N$. For every $i \in N$ and $a, b \in Z$ the effectivity correspondence is such that $E(a, b) = \{i\}$ if and only if $a_{-i} = b_{-i}$. Note that this definition does not allow coalitional moves.

		Player 2	
		l	r
Player 1	u	$5, 5$	$0, 6$
	d	$6, 0$	$3, 3$

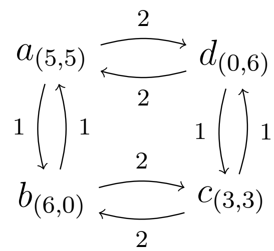


Figure 2.7: Transformation from normal form to effectivity function form

While strategic-form games are usually interpreted as static one-shot games, their abstract version lends itself to a dynamic interpretation where players are free to switch between actions whenever they wish. Under this dynamic interpretation, the set of Nash equilibria emerges as the set of stationary states when players are assumed to behave myopically. More particularly, the set of states supported by myopic SVREs coincides with the set of Nash equilibria. The following proposition shows that when players are farsighted and have the power to coordinate on not moving away from any state, all, and only, Pareto efficient states are supported as stationary.

Proposition 2.9. *Take a generic, finite, strategic form game and suppose* $(a, a, N) \in M$ *for any* $a \in Z$*. Then,* $A \subseteq Z$ *is supported by some SVRE* m *if and only if it is a singleton set containing a Pareto efficient state.*[20]

Proof. The proof of Proposition 2.5 shows that in a finite and generic strategic form game any singleton set containing a Pareto efficient state can be supported by some SVRE m. It remains to show that **only** such sets can be supported. We do so in four steps:

1. Show that for any SVRE m and $a \in Z$, $|Y(a, m)| = 1$.

2. Claim that under the condition in (1), and assuming genericity, if there exists a player that can move from one basin of attraction to another, then m is not SVRE.

3. Show this implies that for any SVRE m, $|Y(m)| = 1$.

4. Show this implies that for any SVRE m, the unique state $Y(m)$ contains is Pareto efficient.

[20] See Theorem 4.2 in Granot and Hanany (2022), as well as Brams and Ismail (2022), for similar results.

Step 1. By Lemma 2.1, if m is SVRE, $Y(a,m)$ is single payoff for all $a \in Z$. Since the game is generic, no two states provide any player the same payoff. Hence, if $Y(a,m)$ is single payoff it must contain a unique state.

Step 2. Suppose there exist $a, b \in Z$ such that $Y(a,m) \neq Y(b,m)$. Note that this implies that there do not exist $(a,b,S),(b,a,S) \in m$ (otherwise one of the sets would have been a strict superset of the other, contradicting it being single payoff). Under genericity, every player has a strict preference between $Y(a,m)$ and $Y(b,m)$. Hence, if some i can move from a to b and vice-versa, that is, if there exist $(a,b,\{i\}),(b,a,\{i\}) \in M$, then m cannot be stationary: one of those moves could profitably be added to m.

Step 3. Suppose m is SVRE. By proposition 2.4 (as well as Step 1 in this proof), $|Y(m)| \neq 0$. Take some $a \in Y(m)$. By Step 2, any state b such that there exists i for which $(a,b,\{i\}),(b,a,\{i\}) \in M$ must have $Y(b,m) = \{a\}$. Similarly, all states c that are i-adjacent to those bs (for some i) must have $Y(c,m) = \{a\}$. Continuing this way, all states are exhausted. Hence, $Y(a,m) = \{a\}$. Put otherwise, $|Y(m)| = 1$.

Step 4. Suppose m is SVRE and $Y(m) = \{a\}$ but a is not Pareto efficient. Letting b denote a state that (strictly) Pareto dominates it (such b must exist if a is not Pareto efficient). $m' = [m \setminus m(b)] \cup \{(b,b,N)\}$, is a feasible and profitable deviation for N. In addition, it is dynamically consistent. Contradiction. Hence, the unique state in $Y(m)$ for any SVRE m must be Pareto efficient. This concludes the proof. □

Lemma 2.1. *Suppose $(a,a,N) \in M$, all players can (individually) move away from all states, and all individual moves are bi-directional. Then, for any SVRE m and $a \in Z$, $Y(a,m)$ is single-payoff.*

Proof. Consider some m. Let m_{-i} denote the plan of action of all coalitions but $\{i\}$. Let $y(a,m_{-i})$ denote the set of *potentially* stationary states reachable from a given m_{-i}. Formally, $y(a,m_{-i}) = Y(a, m_{-i} \cup m_i^c)$, where $m_i^c = \{(z,z,\{i\}) \mid z \in Z\}$. Note that since all individual moves are bi-directional, it is always possible to choose m_i such that all states in $y(a,m_{-i})$ are reachable from one another via moves in $m_{-i} \cup m_i$ (in other words, the subgraph defined by the set of vertices $y(a,m_{-i})$ and the corresponding edges in $m_{-i} \cup M_i$ is strongly connected).

We argue that if $m_{-i} \cup m_i$ is SVRE, then it must be the case that $\emptyset \neq Y(a, m_{-i} \cup m_i) \subseteq \arg\max_{z \in y(a,m_{-i})} u_i(z)$. First note that since $(z,z,N) \in M$ for all $z \in Z$, we know (by Proposition 2.4) that $\arg\max_{z \in y(a,m_{-i})} u_i(z) \neq \emptyset$. If none of the states in $\arg\max_{z \in y(a,m_{-i})} u_i(z)$ belongs to $Y(a, m_{-i} \cup m_i)$ then $\{i\}$ can profitably deviate to a dynamically consistent m' by choosing to remove all moves away from one of them (and add the individual inaction move), thus making it stationary. If $Y(a, m_{-i} \cup m_i)$ contains a state not in $\arg\max_{z \in y(a,m_{-i})} u_i(z)$ (on top of states that do belong to it), then there exists a state from which i could reach one of the stationary states in $\arg\max_{z \in y(a,m_{-i})} u_i(z)$ but doesn't, meaning she has a profitable deviation to a dynamically consistent m'. □

2.5.2 Extensive Form Games

Extensive-form games are transformed into abstract games by letting each node in the extensive form be a state in the abstract form and imputing preferences over non-terminal states such that the active player at each of them would always have an incentive to move down the game tree. Figure 2.8 provides a simple example. Formally, for every extensive form game, let Z be the set of nodes in the game tree and partition Z into $Z_0, Z_1, ...Z_n$, where Z_i, $i \in N$, is the set of nodes that belong to player i and Z_0 is the set of terminal nodes. For every $i \in N$ and $a \in Z \setminus Z_0$ set $u_i(a) < u_i(b)$ for all $b \in Z_0$. For any $i \in N$ and $a \in Z_i$ set $E(a, b) = \{i\}$ if b is adjacent from a.

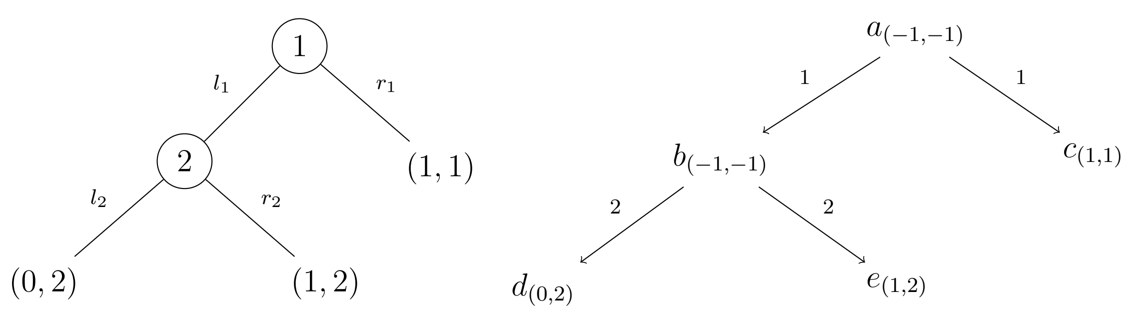


Figure 2.8: Transformation from extensive form to effectivity function form

The following proposition states that in every perfect-information extensive-form game, the predictions of SVREs and subgame perfect equilibria coincide.

Proposition 2.10. *In any finite, perfect-information, extensive-form game, a path is prescribed by an SVRE m if and only if it is prescribed by some subgame prefect equilibrium.*

Proof. A perfect information extensive form game is a special case of an acyclic abstract game where there is exactly one active coalition at every non-terminal state (in particular, one where coalitions are always of size 1) and payoffs from non-terminal states are set to be very low. The proof of Proposition 2.6 therefore shows that if a path is prescribed by a subgame prefect equilibrium then it is prescribed by an SVRE m.

In the other direction, suppose a path is prescribed by some SVRE m but not by any subgame perfect equilibrium. This means that there exist $i \in N$ and $a \in Z$ such that, holding constant the strategies of all players but i, as well as i's strategies at all states but a, i can profitably deviate from her prescribed action at a. But this means that Conditions (OP).(i)-(iii) hold, implying that m is not SVRE. Contradiction. □

2.5.3 Partition Function Form Games

A partition function form game is described by a finite set of players N and a "partition function" that takes a coalition structure (or "partition") $\pi \in \Pi$ and maps each of the coalitions embedded in it $S \in \pi$ to a non-empty set of feasible payoff vectors $V(S, \pi)$. We assume that $V(S, \pi)$ for all coalitions $S \subseteq N$ under all partitions $\pi \in \Pi$ contains a finite amount of payoff vectors. For singleton coalitions, we take the stronger assumption that $V(\{i\}, \pi)$ contains a *unique* payoff vector: for any $\pi \in \Pi$ and $i \in N$, if $\{i\} \in \pi$ then $|V(\{i\}, \pi)| = 1$. This reflects the idea that there is no room for negotiation within a coalition that contains only one player.

In translating these games into abstract games we broadly follow the approach taken in Ray and Vohra (2015) and Ray and Vohra (2019). A "state" is defined by a partition and a payoff vector that is feasible for all coalitions given this partition. Formally, a generic state z is a pair (π, u) (or $(\pi(z), u(z))$, when we need to be explicit), where $u(z)$ is feasible for all $S \in z(\pi)$ given $z(\pi)$.

The effectivity correspondence is required to satisfy two conditions. The first is that a moving coalition S has no direct control over the coalition structure and payoffs of the set of players in the complement of S, denoted S^c. To formalize this idea we define a "default function" $f(z, S, \pi^S) = (\pi^{S^c}, u^{S^c})$, which associates a partition π^{S^c} of S^c and a feasible payoff vector u^{S^c} for the players in S^c with every state z, coalition S and a partition of S, π^S. The feasibility of u^{S^c} in this definition is taken with respect to the union of structure S considers forming, π^S, and the one f dictates for S^c, π^{S^c}, i.e. it is required that for every $T \in \pi^{S^c}$ we have $u^T \in V(T, \pi^S \cup \pi^{S^c})$. The formal condition on the effectivity correspondence is then as follows: if $S \in E(z, z')$ then $(\pi^{S^c}(z'), u^{S^c}(z')) = f(z, S, \pi^S)$.

There exist various natural assumptions that can be imposed on f (for instance, that coalitions disjoint from S remain intact), however, the precise details of this function do not matter for our analysis. To get an intuition for why this is the case, note that f does not intend to capture S's *expectation* of S^c's reaction to its move (as in Bloch and Van den Nouweland (2014), for example), but merely serve as some default immediate implication of S's move. Players in S^c are free to make adjustments in subsequent steps, and since we are analyzing farsighted behavior those adjustments are expected by members of S to begin with.

The second condition on the effectivity correspondence is that if S wants to move from z, it can do so by reorganizing itself in any way it wishes (including breaking up into smaller pieces, captured by the partition π^S of S), so long as the resulting payoff it obtains is feasible given the partition pinned down by π^S and f. Formally: for any state z, coalition S, partition π^S of S, and feasible payoff u^S, there is $z' \in Z$ such that $S \in E(z, z')$, $\pi^S \subseteq \pi(z')$, and $u^S(z') = u^S$. Like in the first condition, the feasibility of u^S is taken with respect to the union of the structure S wants to form, π^S, and the one f dictates for S^c, π^{S^c}, i.e. it is required that for every $T \in \pi^S$ we have $u_T \in V(T, \pi^S \cup \pi^{S^c})$.

Proposition 2.11 characterizes the set of states that are supported as stationary by some SVRE m in terms of myopically beneficial deviations. In particular, it says that a state is supported as stationary by some SVRE m if and only if there does not exist a move away from it by a coalition of size either n or $n-1$ that is strictly myopically beneficial for all its members.

Proposition 2.11. *In any partition function form game satisfying our assumptions, $a \in Z$ is supported as stationary by some SVRE m if and only if there does not exist $(a, b, S) \in M$ satisfying $|S| \geq n-1$ and $u_i(b) > u_i(a)$ for all $i \in S$.*

Proof. Suppose $a \in Z$ satisfies the condition in the proposition and let $m = M \setminus \{(a, z, S) \in M \mid z \neq a\}$, i.e. m contains all feasible moves besides those that replace a by another state. a is terminal w.r.t. this m. Moreover, since it is reachable from all other states, we have $Y(z, m) = \{a\}$ for all $z \in Z$. We want to show that this m is SVRE.

First, since $Y(z, m) = \{a\}$ for all $z \in Z$, m is dynamically consistent. Hence, our task is to show that there does not exist m' satisfying conditions (OP).(i)-(iii).

Consider $b \in Z$, $b \neq a$. A deviation at b cannot turn stationary any state other than itself. If it does not turn b to a stationary state we have $Y(b, m) = Y(b, m') = \{a\}$ for any m' that differs from m at b, so this cannot be a profitable deviation.

Consider deviations that do turn b to a stationary state. If $\pi(b)$ does not contain an isolated player then turning b into a stationary state is not feasible for any coalition but N (since $m(b) = M(b)$). For $S = N$, turning b into a stationary state is not profitable, as we necessarily have $(a, b, N) \in M$ and by the assumption in the proposition, there exists $i \in N$ that does not strictly benefit from it. Hence, there do not exist feasible and profitable deviations from any $b \neq a$ such that $\pi(b)$ does not contain an isolated player.

If $\pi(b)$ *does* contain an isolated player, then, for the same reason as above, turning b into a stationary state is not feasible for any coalition but those of size weakly larger than $n-1$. For any such coalition, i.e. for any S such that $|S| \geq n-1$, turning b into a stationary state is not profitable, as we necessarily have $(a, b, S) \in M$, and, by the assumption in the proposition, there exists $i \in S$ that does not strictly benefit from it. Hence, we established that there do not exist feasible and profitable deviations from any $b \neq a$.

Next, we show that there are no profitable deviations from a. A deviation at a may only turn a to a non-stationary state, i.e. for any m' that differs from m at a, either $Y(a, m') = \{a\}$, or $Y(a, m') = \emptyset$. Neither option can make any player strictly better off compared to m, so no profitable deviations from $m(a)$ exist. Hence, m is SVRE.

We now show that *only* states satisfying the condition in the proposition can be supported as stationary. Assume by contradiction that m is SVRE and $a \in Y(m)$, but a does not satisfy the condition in the proposition, i.e. there exists $(a, b, S) \in M$ such that $|S| \geq n-1$ and $u_i(b) > u_i(a)$ for all $i \in S$. Let M^B_{n-1} denote the set of such moves by coalitions of size $n-1$, M^B_n the set of such moves by coalitions of size n, and B the set of such states. Formally, $M^B_{n-1} \equiv \{(a, b, S) \in M \mid |S| = n-1,\ u_i(b) > u_i(a)\ \forall i \in S\}$, $M^B_n \equiv \{(a, b, S) \in M \mid |S| = n,\ u_i(b) > u_i(a)\ \forall i \in S\}$, and $B \equiv \{b \in Z \mid (a, b, S) \in M^B_{n-1} \cup M^B_n\}$. Our contradiction assumption then reads $B \neq \emptyset$.

First, observe that no $b \in B$ belongs to $Y(m)$: otherwise, there exists a move in M^B that its executing coalition would like to add to m, implying that m is not SVRE. Also, observe that the grand coalition is effective in moving from any state to any state. Taken together, these two observations imply that all $z \in Y(m)$ satisfy $u_i(z) \leq u_i(a)$ for all $i \in N$. But this yields a contradiction: for any $(a, b, S) \in M^B_{n-1} \cup M^B_n$, adding (b, b, S) to m and removing all moves away from b is feasible, profitable, and does not cause a violation of dynamic consistency. To see that removing all moves away from b is feasible for any $(a, b, S) \in M^B_{n-1}$ note that any state such that a coalition $S = N \setminus \{i\}$ can move to has a coalition structure under which i is an isolated player. When i is isolated she cannot, on her own, move to any state with a coalition structure different than the status quo. In addition, due to the assumption that $|V(\{i\},\pi)| = 1$ for any $\pi \in \Pi$ and $\{i\} \in \pi$), she also cannot move to any state with the same coalition structure but a different payoff vector. Hence, $\{i\}$ is not effective in any move away from a state such that $S = N \setminus \{i\}$ can move to, implying that S can remove all moves away from it and enforce its stationarity. This establishes the required contradiction. If m is SVRE, no a violating the condition in the proposition can be supported as stationary.

□

Proposition 2.11 says that to determine farsighted stability it is enough to consider strictly myopically beneficial deviations by coalitions of sizes n and $n-1$. This requirement is clearly weaker than that of the weak core, which seeks immunity to (strictly) myopically beneficial moves by coalitions of *any size*. A corollary of Propositions 2.11 and 2.3 is therefore that all states supported by some *myopic* SVRE (said otherwise, that belong to the weak core) are also supported by some SVRE. The converse need not hold: there may exist states that are supported by an SVRE but not by a myopic SVRE. To state the corollary, we let S^f denote the set of all ms which are SVREs and recall that S^p is used to denote the set of all ms which are myopic SVREs.

Corollary 2.2. *For any partition function form game satisfying our assumptions, letting Γ denote its abstract game translation, we have $\bigcup_{m\in S^p} Y(m) = WC(\Gamma) \subseteq \bigcup_{m\in S^f} Y(m)$.*

The two results above indicate that the (myopic) core of cooperative games is linked to farsighted stability. For similar results see Ray (1989), Konishi and Ray (2003) and Ray and Vohra (2015). The first shows that the core (in its entirety) is immune to farsighted objections provided they are "nested" in the sense that every subsequent move must be effected by a subset of the coalition effecting the previous move. The second shows that the core can be described as the limit of a dynamic process of a coalition formation with a stream of payoffs and a discount factor that is close enough to 1. The third shows that all single-payoff farsighted stable sets are core allocations, and every payoff allocation in the interior of the core (along with appropriate coalition structures) forms a farsighted stable set. Our results may also be contrasted with Dutta and Vartiainen (2020), who analyze partition function form games using their history-dependent "HREFS" solution concept.

2.6 Comparison to Related Solution Concepts

2.6.1 SVRE vs. REEFS (Karos and Robles, 2021)

While we tackle the counterfactual critique by introducing set-valued expectations, Karos and Robles (2021) do so by introducing *extended* expectation functions. An extended expectation function specifies for each state an ordered list of coalitions and their moves. Thus, each coalition knows that if it won't move, the next one on the list will have the floor. This knowledge allows coalitions to build a counterfactual for what would happen had they decided not to move.

Definition 2.5. *An extended expectation function is a map F that assigns to each $z \in Z$ an ordered list $\left(F^1(z), ..., F^{k(z)}(z)\right)$ such that $F^l(z) = \left(f^l(z), S^l(z)\right) \in Z \times 2^N$ with $S^l(z) \in E(z, f^l(z))$ for all $l = 1, ..., k(z)$, $S^l(z) \neq S^{l'}(z)$ for all $l \neq l'$, $f^l(z) \neq z$ for all $l \neq k(z)$, and $S^{k(z)}(z) = \emptyset$.*

In accordance with the current paper (as well as most of the related literature), Karos and Robles (2021) assume that players only care about their utility in final states. In addition, like in the current paper, they assume that players always prefer reaching *some* final state over never reaching a final state at all. The following function captures these ideas. Note that F^1 is a "basic" expectation function, which we denoted by σ in Subsection 2.3.1, and recall that for any basic expectation function σ we use $x^\sigma(z)$ to denote the final state that it leads to from z (assuming such a final state exists).

$$U_i(z, F) = \begin{cases} u_i(x^{F^1}(z)) & \text{if } x^{F^1}(z) \text{ exists} \\ -\infty & \text{otherwise} \end{cases}$$

Using this function, Karos and Robles (2021) define a *rational* extended expectation function (REEFS) as follows.

Definition 2.6. *An extended expectation function F is rational (REEF) if it satisfies the following three conditions:*

(I) For all $z \in Z$ and all coalitions $T \notin \{S^1(z), ..., S^{k(z)}(z)\}$ there is $l \leq k(z)$ such that for each $z' \in Z$ with $T \in E(z, z')$ there is $i \in T$ with $U_i(f^l(z), F) \geq U_i(z', F)$.

(E) For all $z \in Z$ and for all $l = 1, ..., k(z)-1$ it holds that $U_i(f^l(z), F) > U_i(f^{l+1}(z), F)$ for all $i \in S^l(z)$.

(M) *For all* $z \in Z$ *and for all* $l = 1, ..., k(z) - 1$ *it holds that if there is* $z' \neq f^l(z)$ *such that* $S^l(z) \in E(z, z')$*, then there is* $i \in S^l(z)$ *with* $U_i(f^l(z), F) \geq U_i(z', F)$*.*

Note that a REEF is still "single-valued" in the sense that in equilibrium at most one move away from each state is expected to be executed. As a consequence, REEF does not address the overconfidence critique. We illustrate this below by revisiting Example 2.2 once more.

Example 2.2 (Continued)**.** *Consider again the game in Figure 2.1. There are two REEFs in this game: one in which Player 1 is overly confident that at b player 2 will take precedence and move to c, and hence decides not to move away from a; and one in which she is overly confident that at b player 3 will take precedence and move to d, and hence decides to move away from a. Formally, the first is given by* $F(b) = ((c, \{2\}), (d, \{3\}), (b, \emptyset))$*, and* $F(a) = (a, \emptyset)$*; and the second is given by* $F(b) = ((d, \{3\}), (c, \{2\}), (b, \emptyset))$*, and* $F(a) = ((b, \{1\}), (a, \emptyset))$*.*

Under the SVRE concept, Player 1 takes into account both continuation paths simultaneously. Suppose she prefers $\{c, d\}$ *over* $\{a\}$*. Then, while REEF predicts that state a may be stationary (under the first REEF described above), SVRE predicts that it is never stationary.*

While the REEF as a concept does not prefer one order of play over another, every particular REEF is associated with some particular order of play. We note that assumptions about the order of play can be implemented via the definition of the game, and therefore do not exclude the relevance of the SVRE concept. For the game in Example 2.2, for instance, one could implement the assumption that at b Player 2 takes precedence by amending the game as shown in Figure 2.9. It could then be analyzed using the SVRE concept. In the amended game, following a move away from a Player 2 is the only active player, and she needs to decide whether to move to c or leave the floor to Player 3 by moving to b (in principle she can also stay at b', but we have imputed her payoff at this state to be lower than in all other states to make sure this does not happen). At b player 3 is the only active player and she gets to choose whether to stay there, which would result in the same payoffs as in the original b state, or move to d. Hence, this game is equivalent to the original one with the additional assumption that at b Player 2 is given the floor first. Indeed, in the amended game the predictions of REEF and SVRE coincide.

In general, a (partial) coincidence result between REEF and SVRE is achieved whenever all active coalitions at every state can agree on a single preferred move, i.e.

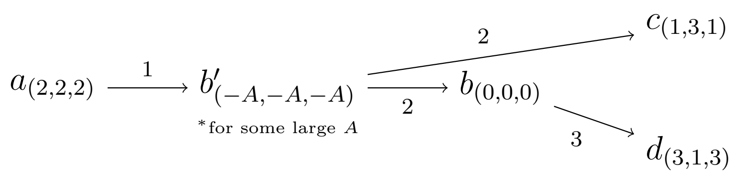


Figure 2.9

when there exists a (history independent) "basic" expectation function satisfying Ray and Vohra (2019)'s conditions. This is stated in the corollary below, which follows immediately from Proposition 2.1 in the current paper and Lemma 3.3. in Karos and Robles (2021).

Corollary 2.3. *Suppose σ satisfies Ray and Vohra (2019)'s conditions and is history independent, then m^σ and $F = \{\sigma, F^2\}$ are equivalent and are, respectively, SVRE and REEF.*

2.6.2 SVRE vs. SPCS (Granot and Hanany, 2022)

Granot and Hanany (2022) model the (history-dependent) evolution of play resulting from coalitions' deviations as an (infinite) extensive form game. They assume a protocol according to which at each state nature draws a coalition that is given the opportunity to make a move. The protocol either has full support, meaning that any coalition that is effective in some move has a positive probability to be chosen ("full support protocol"), or at least partial support, meaning that at least one coalition that can affect a move has a positive probability to be selected ("regular protocol"). A set of states is called "Subgame Perfect Consistent Set" (SPCS) if it contains all states which can be supported as stable by a subgame perfect equilibrium, refined to additionally satisfy both internal and external consistency in the spirit of the Farsighted Stable Set, whereby all stable states that are reachable by some continuation subgame perfect equilibrium are expected by potentially moving coalitions from an initial state. Since this solution concept does not rely on the notion of farsighted improving paths, it is immune to the counterfactual critique. Under a full support protocol, since every coalition that is effective in some move has a positive probability of being selected, it is also immune to the overconfidence critique. Nonetheless, some significant differences between it and the SVRE concept exist. We expose them after providing the formal definition, as appears in Granot and Hanany (2022).

Formally, the protocol is treated as the strategy of a player referred to as "nature". The nature player is denoted c and the extended set of players is denoted $N_c = N \cup \{c\}$. $\mathcal{H}$ is the set of all possibly infinite sequences $h = (h_k)_{k=0}^{K_h}$, including the empty sequence when $K_h = -1$, such that h_0 is some initial state in Z, for all odd numbers $k \geq 1$, h_k is some coalition $S \subseteq N$, and for all even numbers $k \geq 2$, $h_k \in Z$, such that $h_k = h_{k-2}$ or $h_{k-1} \in E(h_{k-2}, h_k)$. For two finite histories $h, h' \in \mathcal{H}$ with even, positive cardinality such that $h'_0 = h_{K_h - 1}$ or $h_{K_h} \in E(h_{K_h-1}, h'0)$, denote by (h, h') the history in $\mathcal{H}$ obtained when h is followed by h'. Denote the set of infinite histories by $\mathcal{H}_\infty$. Define the player function $P : \mathcal{H} \setminus \mathcal{H}_\infty \to N_c$ by $P(\emptyset) = c$, $P(h) = c$ for every finite history h with odd cardinality $|h|$, and $P(h) = h_{K_h} \subseteq N$ for every finite history h with even, positive cardinality.

A history $h \in \mathcal{H}$ is said to converge for player i if there exist $\bar{z}_i(h) \in Z$ and an even, positive number $k_{h,i}$ such that $u_i(h_k) = u_i[\bar{z}_i(h)]$ for every even k such that $k_{h,i} \leq k \leq K_h$. $k^0_{h,i}$ is defined to be the minimal such $k_{h,i}$. A history h is said to converge if it converges for all players, in which case the player index is omitted. Denote by $\bar{\mathcal{H}}$ (resp., $\bar{\mathcal{H}}_i$) the set of all infinite converging histories (resp., for player i). When an infinite history h does not converge for player i, it is said to lead for that player to 'swinging', denoted by w, in which case $\bar{z}_i(h) = w$ and $k^0_{h,i} = \infty$. $\bar{Z} = Z \cup \{w\}$ denotes the extended states set, and the utility function u_i is extended so that $u_i(w) = -\infty$, for all i.

Nature's strategy, named protocol, is a function μ defined over $\{h \in \mathcal{H} \setminus \mathcal{H}_\infty \mid P(h) = c\}$ such that μ_h at history h for which $P(h) = c$ is a probability measure defined over Z when $h = \emptyset$ and over 2^N otherwise, specifying the distribution over initial states and over the coalitions selected to make a choice. μ is said to be "regular" if for every finite history h with odd cardinality $|h|$, $\mu_h(S) > 0$ for some coalition S, if one exists, that has the effectiveness to move at the current state. μ is said to be "full support" if it has full support at every history h for which $P(h) = c$.

A pure strategy of a coalition S is a function $\sigma_S : \{h \in \mathcal{H} \setminus \mathcal{H}_\infty \mid P(h) = S\} \to Z$ such that $\sigma_S(h) \in \{z \in Z \mid z = h_{K_h-1} \text{ or } S \in E(h_{K_h-1}, z)\}$, specifying the action taken by this coalition after any history in the game where this coalition is selected to make a choice (a complete contingent plan).[21] A strategy profile is a vector $\sigma = [\mu, (\sigma_S)_{S \subseteq N}]$ specifying the protocol and the strategy of each coalition S. Denote by σ_{-S} the protocol together with the vector of strategies for all coalitions except for S.

[21] This σ should not be confused with the σ defined earlier in the paper as a single-valued expectation
.

Let Σ_h denote the set of strategy profiles which, following h, generate a probability measure having support consisting only of infinite converging histories. Whenever $\sigma \in \Sigma_h$, this strategy profile generates for player i a distribution $d_{\sigma|h}$ over $\bar{z}(h') \in Z$. Let $\Psi_{\sigma|h}$ denote the support of this distribution, interpreted as the set of potential (non-swinging) final states according to σ following h. More generally, a strategy profile σ (which might support non-converging histories following h), generates for player i a distribution $d_{i,\sigma|h}$ over $\bar{z}_i(h') \in \bar{Z}$. Let $\Psi_{\sigma|h}$ denote the support of this distribution, interpreted as the set of potential final states according to σ following h. For each player i let $\Psi_{i,\sigma|h}$ denote the support of this distribution.

In a subgame that follows a finite history h with $P(h) = S$, coalition S has a strict preference relation $\succ_{S,h}$ over strategy profiles, where weak preference $\succcurlyeq_{S,h}$ and indifference $\sim_{S,h}$ are defined from the strict preference $\succ_{S,h}$ in the usual way. When the coalition is a singleton player i, Granot and Hanany (2022) assume the following: For each player i, the preference relation $\succ_{i,h}$ is continuous, and $\sigma \succ_{S,h} \sigma'$ is implied if $d_{i,\sigma|h}$ strictly first-order stochastically dominates $d_{i,\sigma'|h}$, where $\bar{Z}$ is ordered according to u_i. For general coalitions they assume the following: For each coalition S, the preference relation $\succ_{S,h}$ satisfies that $\sigma \succ_{S,h} \sigma'$ is implied if $\sigma \succcurlyeq_{i,h} \sigma'$ for all members $i \in S$, with strict preference for at least one member of S.

Given coalition preferences, Granot and Hanany (2022) define a strategy profile σ as a subgame perfect equilibrium if $\sigma \succsim_{S,h} (\hat{\sigma}_S, \sigma_{-S})$ for every coalition S, every finite history h with $P(h) = S$, and every strategy $\hat{\sigma}_S$ for this coalition. A state z^2 is said to be reachable from a state z^1 if for a subgame in which z^1 is the initial state, there exists a subgame perfect equilibrium strategy profile σ such that z^2 is a final state. $R(z^1)$ denotes the set of states z^2 reachable from z^1 when σ is constrained to prescribe full support protocols. $R^*(z^1)$ denotes the set of states z^2 reachable from z^1 when σ is constrained to prescribe regular protocols.

Lastly, Granot and Hanany (2022) define equilibrium dynamics of play with respect to a set of states $X \in Z$ as follows.

Definition 2.7. *Given a strategy profile σ and a set of states $X \in Z$, say that σ is X-farsighted, denoted σ^X, if σ is a subgame perfect equilibrium with full support protocol μ that satisfies the following two requirements:*

(a) for any history $h = (z, S^1, z, S^2, ..., z, S^t)$ such that $z \in R(z)$ and $\left\{S^l\right\}_{l=1}^{t} \supseteq 2^N \backslash \emptyset$, $\sigma_{S^t}(h) = z$; and

(b) for any history $h = (z^1, S^1, z^1, S^2, \ldots, z^1, S^t, z^2, S^{t+1})$ *such that* $z^1 \neq z^2$ *and* $\left\{S^l\right\}_{l=1}^{t} \not\supseteq 2^N \setminus \emptyset$*, the set of final states is* $\Psi_{\sigma|h} = X \cap R(z^2)$

Say that σ *is* X*-farsighted*, denoted* σ^{*X}*, if the protocol* μ *is only required to be regular, and with* R *replaced with* R^**.*

The set of states supported as stationary by σ^X is called a Subgame Perfect Consistent Set (SPCS). Formally, $X \in Z$ is a Subgame Perfect Consistent Set (SPCS, resp. SPCS*) if there exists σ^X (resp. σ^{*X}) such that $z \in X$ if, and only if, $\sigma_S^X(h) = z$ (resp. $\sigma_S^{*X}(h) = z$) for any coalition S and any finite history $h = (z, S^1, z, S^2, ...z, S)$.

Like the SVRE concept, the SPCS does not take an assumption about players' optimism\conservatism. As desirable property as that may be, combining it with the fact that $R(z^1)$ is defined to include the final states under *all* possible subgame perfect equilibria (restricted to a full support protocol) leads to problems of existence that the SVRE concept does not necessarily face. Example 2.8 illustrates this point.

Example 2.8. *Consider the game in Figure 2.10. To find a SPCS (if exists) we first need to compute the reachability function* R*. Any subgame perfect equilibrium* σ *with full support protocol must prescribe that at* b*, whenever* $\{2\}$ *or* $\{3\}$ *are selected, they move to* d *or* e *respectively. Hence, when at* a*,* $\{1\}$ *clearly does not choose to stay at* a*, and this implies that when at* o*,* $\{3\}$ *always moves to* a*. The only thing left to determine is whether at* a *coalition* $\{1\}$ *moves to* b *or to* c*. This decision is dictated by the first-order stochastic dominance relation between a probability distribution over obtaining a payoff of 2 or 0, and a degenerate distribution yielding a payoff of 1 for sure. However, neither distribution first-order stochastically dominates the other. Hence, neither move contradicts optimality. This means that there exists a subgame perfect equilibrium under which* $\{1\}$ *moves to* b*, and another under which* $\{1\}$ *moves to* c*. Let us denote the first by* σ^1 *and the second by* σ^2*. Since* R *is computed based on both* σ^1 *and* σ^2*, we have* $R(z) = \{z\}$ *for* $z = c, d, e$*,* $R(b) = \{d, e\}$*,* $R(a) = \{c, d, e\}$*, and* $R(o) = \{c, d, e\}$*.*

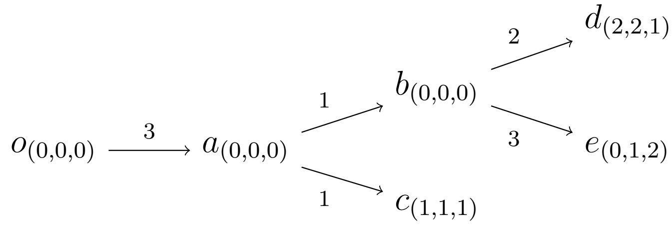


Figure 2.10

Now consider the history $h = (o, \{3\}, a)$. For σ^1 we have $\Psi_{\sigma^1|h} = \{d, e\}$. Requirement (b), asks that $\Psi_{\sigma^1|h} = X \cap R(a)$, i.e. $\{d, e\} = X \cap \{c, d, e\}$, which pins down $X = \{d, e\}$. But if that is the case, following the history $h' = \{a, \{1\}, c\}$ we have $\Psi_{\sigma^1|h'} = \{c\} \neq X \cap R(c) = \{d, e\} \cap \{c\} = \emptyset$, so Requirement (b) fails. Hence σ^1 is not X-farsighted for any $X \subset Z$.

For the same h as above, consider now σ^2. Requirement (b), asks that $\Psi_{\sigma^2|h} = X \cap R(a)$, i.e. $\{c\} = X \cap \{c, d, e\}$, which pins down $X = \{c\}$. But if that is the case, following the history $h'' = \{a, \{1\}, b\}$ we have $\Psi_{\sigma^2|h''} = \{d, e\} \neq X \cap R(b) = \{c\} \cap \{d, e\} = \emptyset$, so Requirement (b) fails. Hence σ^2 is also not X-farsighted for any $X \subset Z$. Since there are no other subgame perfect equilibria in this game, an X-farsighted equilibrium fails to exist.

In contrast, there exists an SVRE m for any possible ranking of Player 1 between $\{c\}$ and $\{d, e\}$. In particular, if $\{c\} \succ_1 \{d, e\}$, then $m = M \setminus \{(a, b, \{1\})\}$ is SVRE; if $\{d, e\} \succ_1 \{c\}$, then $m = M \setminus \{(a, c, \{1\})\}$ is SVRE; and if $\{d, e\} \sim_1 \{c\}$, then $m = M$ is SVRE.

An additional point of divergence between SPCS and SVRE relates to the dynamics that are assumed to take place *within* states. The SPCS concept assumes that within each state nature sequentially selects coalitions to be allowed to execute moves. In particular, when S is selected, S has the ability to make a move and no subset of S does. This implies that in equilibrium S may choose to execute some move even though there exists $T \subset S$ that has a move all its members prefer over it. The SVRE concept, in contrast, does not allow for such a situation. Example 2.9 illustrates this point.

Example 2.9. *Consider the game in Figure 2.11. Suppose o is the status quo, coalition $\{1, 2\}$ is selected and μ is a full support protocol. $\{1, 2\}$'s decision to stay at o or move to a is dictated by the first-order stochastic domination relation between a probability distribution yielding 2 for sure and a probability distribution yielding either 1 or 3. Since neither first-order stochastically dominates the other, no decision contradicts optimality. Hence, the SPCS concept predicts that a may be reached from o.*

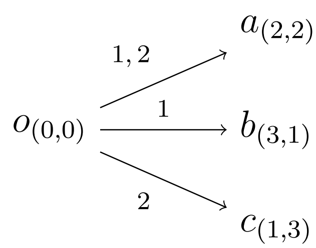


Figure 2.11

Under the SVRE concept, in contrast, neither Player 1, nor 2, will ever plan to support the move $(o, a, \{1, 2\})$*. This is because whenever they are able to give their support for some move, they prefer to give it to* $(o, b, \{1\})$ *or* $(o, c, \{2\})$*, respectively.*

A general coincidence result between SPCS and SVRE is obtained when restricting attention to strategic form games. The following corollary formally states this result, which follows immediately from Proposition 2.9 in the current paper and Theorem 4.2 in Granot and Hanany (2022).

Corollary 2.4. *Take a generic, finite, strategic form game, suppose* $(a, a, N) \in M$ *for any* $a \in Z$ *and let* $A \subseteq Z$*. The following statements are equivalent:*

- *A is a singleton set containing a Pareto efficient state.*
- *A is supported by some SVRE m.*
- *A is SPCS.*
- *A is SPCS*.*

2.7 Concluding Remarks

Abstract games provide a unified framework that allows describing a wide range of games from both the cooperative and non-cooperative branches of game theory. Recent papers have proposed to capture optimal farsighted behavior in this framework using expectation functions, which specify a unique continuation path for each state. In our view, existing solutions in this literature make compromises either on their interpretation of "optimal behavior" or on their faithfulness to the free-form protocol of play characterizing abstract games (and cooperative game theory in general). Dutta and Vohra (2017) compromise "optimal behavior" in that the existence of a coalition that has the power to block the move prescribed by the expectation and execute another one instead is not necessarily sufficient to render the expectation "unstable" (in particular, this happens when this coalition shares a non-empty intersection, but is not contained in, the coalition prescribed to move). On the other hand, Karos and Robles (2021) compromise the free-form protocol of play in that every equilibrium their solution admits is associated with a specific order of play. While Ray and Vohra (2019) do not compromise on any of the aspects above, their solution loses existence as soon as a disagreement among coalitions about the preferred move arises.

We go beyond existing "rational expectations" solutions by allowing expectations to be set-valued. This allows insisting on a strong notion of optimality *and at the same time* on a free-form protocol of play. On the flip side, since set-valued expectations render continuation paths non-unique, an extension to preferences over sets of states is required. The assumptions we make on the extension rule are, however, minimally invasive, and all of our results hold under any mode of behavior compatible with them, be it fully optimistic, fully pessimistic, or anything in-between. Note that the set-valued expectations we posit are history-independent. Allowing for history dependence is left for future work.

One of our results states that the proposed solution concept reduces to subgame perfect Nash equilibrium when extensive form games (with perfect information) are considered. This result points at a different way to interpret the proposed solution concept: while we chose to present it as an extension of existing solutions for abstract games, it may also be cast as a generalization of subgame perfection for a type of extensive form game that allows for coalitional moves and avoids specifying an order of play. Indeed, much like Nash equilibrium, our solution requires the prescribed behavior to be a fixed point in the sense that whenever all players expect it, no one wants to deviate from it, and much like subgame perfection it requires behavior to be optimal both on and off equilibrium path.

Lastly, we comment on the inexistence example (Example 2.7). The example hinges on the fact that, when considering a *deviation* from m to some m', players take m' as final and do not consider further deviations this could trigger to m'', m''' and so on. This stands in stark contrast to the idea that when considering a *move* from some state to another players are capable of taking into account the entire chain of moves it may trigger: while players perfectly foresee chains of moves, they are entirely myopic in terms of deviations. Pushing foresight into the realm of deviations may be a fruitful avenue for future research.

Chapter 3

Farsighted Reasoning, Coordination and Cooperation: a Network Formation Experiment

Abstract

We conduct a lab experiment to test the predictions of various myopic and farsighted solution concepts in dynamic multi-player games, as well as the way the dynamics of play are affected by the composition of players in terms of their cognitive abilities and the disclosure of this information. The experiment consists of an initial part measuring participants' cognitive ability, and a subsequent part where participants strategically interact within groups. Results lend empirical support to farsighted solution concepts allowing for preemptive moves. In addition, we find that high-ability individuals tend to insist harder on achieving Pareto-efficient outcomes. This effect is further amplified by the disclosure of information on other group members' cognitive abilities.

Keywords: Cognitive ability, Coordination, Cooperation, Foresight, Myopia, Rational expectations.

JEL codes: C73, C92, D74, D84, D85.

3.1 Introduction

Game-theoretic solution concepts differ from one another in the set of assumptions they reflect. Some reflect myopic behavior in the sense that players are assumed to take into account only the immediate consequences of their actions (e.g. the Core). Others reflect farsighted behavior in the sense that players are assumed to take into account the long-term consequences of their actions (e.g. the Farsighted Stable Set). Yet others assume players take into account not only the long-term consequences of their actions, but also of *lack of actions* on their part (e.g. the Set-Valued Rational Expectations solution proposed in Dekel (2023)). This paper tests the predictive power of each of these categories of solution concepts using a dynamic multi-player game played in laboratory settings. In addition, it examines how observed play is affected by: *(i)* groups' compositions in terms of the cognitive ability of their constituent members; and, *(ii)* publicly revealing information on other group members' cognitive abilities.

The paper brings together two separate literatures. The first deals with empirically testing the predictions of myopic and farsighted solution concepts (e.g. Kirchsteiger et al. (2016), Teteryatnikova and Tremewan (2020), Carrillo and Gaduh (2021)). The second deals with the effects of cognitive ability on strategic behavior (e.g. Jones (2008), Burks et al. (2009), Gill and Prowse (2016), Alaoui and Penta (2016), Proto et al. (2019), Lambrecht et al. (2021), Proto et al. (2022)). Experimental studies belonging to the "solution concepts literature" involve having participants play games that are designed so that different solution concepts (each reflecting a different type of myopic\farsighted behavior), produce different predictions regarding the dynamics and\or final outcomes of the game. Experimental studies "cognitive ability literature" have participants complete some task that provides a measure of their cognitive ability as a first step, and then proceed to pair participants (either of similar or of different cognitive ability) to play a game against one another. While the first literature is well-motivated by theory, it ignores the potential effects that individual heterogeneity in cognitive abilities may have. In view of this, the current paper can be interpreted either as a test of myopic and farsighted solution concepts *that takes into account heterogeneity in cognitive abilities*, or as a *theoretically-motivated* study of the effects of cognitive ability on strategic behavior.

Drawing from the cognitive ability literature, our experimental design is composed of two parts, where the first measures participants' cognitive abilities and the second matches participants to play a game. The cognitive task used in the first part is a novel questionnaire based on the "Hit 15" game, which is meant to measure participants' ability to reason farsightedly in a non-strategic environment and allows assigning each a "farsighted reasoning score". In the second part, participants are matched in groups of four to play a dynamic network formation game. The game starts at the empty network and proceeds by sequentially offering randomly selected group members to form\delete links with other group members. Following any such offer, all group members are presented with the new state of the network (and associated payoffs) and are asked whether they want to stop the formation process at the current network. If they all say YES, the game ends and the current network is declared "final". Otherwise, another group member is randomly offered to form\delete links with other group members. Only final networks affect participants' earnings.

The payoffs associated with the various network structures are designed to distinguish between three categories of solution concepts: *(i)* myopic; *(ii)* farsighted without preemptive actions; *(iii)* farsighted with preemptive actions. The latter two categories differ in what players are assumed to believe would happen had they decided not to take an action. In category *(ii)* players are assumed to believe no other player would either. This implies they never have an incentive to take a preemptive action, i.e. to act only in order to prevent others from doing so. Hence, no preemptive actions are predicted to take place. In category *(iii)* players are assumed to believe that that others react optimally to inaction. Thus, preemptive actions might be predicted to take place.

Each experimental session is randomly assigned a matching treatment and an information treatment. In the "random" matching treatment groups are composed at random, while in the "homogeneous" matching treatment groups are composed to minimize the variance of their members' farsighted reasoning scores (as measured by the questionnaire administered in the first part of the experiment). In the "private" information treatment farsighted reasoning scores remain private information, while in the "public" information treatment they are publicly disclosed to all group members. These treatments allow examining how are the observed dynamics of play and final outcomes affected by groups' compositions in terms of the farsighted reasoning scores of their constituent members, as well as the disclosure of information on others' farsighted reasoning scores.

Our design differs from other experiments within the literature testing myopic and farsighted solution concepts in two important ways. First, unlike other papers in this literature, we precede the game with a cognitive task aimed at measuring participants' cognitive ability. This allows taking cognitive ability into account when matching participants into groups and controlling for cognitive ability in all subsequent analyses. Since myopic and farsighted behaviors are likely to be affected by cognitive ability, we view this as an essential part of the experiment. Second, we distinguish between farsighted solution concepts that do not admit preemptive actions and farsighted solution concepts that do. Solution concepts in the latter category are very recent (have only appeared in Karos and Robles (2021), Granot and Hanany (2022) and Dekel (2023)) and, to the best of our knowledge, have never been explicitly tested before.

With respect to experiments on the effects of cognitive ability on strategic behavior, our design differs in four important ways. First, these experiments most commonly use infinitely repeated 2×2 games, which model situations where every choice of strategies (i.e. a stage in the multistage game) realizes a payoff for each player and these are aggregated across stages. In contrast, in our game payoffs are realized only once a consensus is reached. Our design therefore addresses the effects of cognitive abilities on negotiations-type interactions – a question left unadressed thus far. Second, while most previous experiments in this literature use a Raven test to measure cognitive ability, we use an original questionnaire based on the Hit-15 game. This is motivated by the belief that the central aspect of cognitive ability that is relevant to the experimental game we deploy is the ability to compute several steps ahead, and the Hit-15 game requires exactly that. Third, while previous experiments use 2-player games, ours is played in groups of four. This enriches the set of possible group compositions. Lastly, while most previous experiments study either the effect of groups' compositions *or* the effect of information disclosure on strategic behavior, we do both.

The generated experimental data supports four main results. First, we find that in 87% of the rounds played in the lab groups converge to the predictions made by the "farsighted with preemptive moves" solution concepts. This suggests that participants not only take into account others' reactions to their actions, but also to lack of action on their part. Second, with respect to the level of farsighted reasoning scores, we find that the farsighted reasoning scores computed based on out Hit-15 questionnaire are strongly negatively correlated with taking myopically rational actions. This suggests that the devised Hit-15 questionnaire is a useful tool for measuring the ability to reason farsightedly. Third, we find that participants with higher farsighted reasoning scores tend to insist harder on achieving the predictions made by "farsighted with preemptive moves" solution concepts, characterized by begin Pareto efficient. This suggests that high-score individuals are less prone to coordination failures. There is limited evidence that their insistence on achieving Pareto efficiency is further amplified by the provision of information on other's scores. Fourth, with respect to the dispersion of farsighted reasoning scores, we find that groups with low dispersion consisting of high-score individuals tend to converge to a final outcome faster than groups with high dispersion. This further illustrates high-score individuals' strategic competence.

The paper proceeds as follows. Section 3.2 presents the experimental design. Section 3.3 presents a theoretical analysis of the experimental game and generates predictions. Section 3.4 shows the experimental evidence. Section 3.5 concludes. The supplementary appendices contain further experimental design details, proofs, and the instructions read to participants.

3.2 Experimental Design

Our experimental design consists of two parts. In the first part, participants perform a cognitive task based on the Hit 15 game. We designed it with the aim of measuring participants' farsighted reasoning ability in a non-strategic environment, i.e. in a way that is independent of their beliefs about the farsighted reasoning abilities of others. In the second part, participants are matched in groups of four to play a dynamic network formation game. At each stage of the game, one randomly selected group member is allowed to form or delete one of their links and the game ends once all group members unanimously agree on it. Each experimental session is randomly assigned a matching algorithm and an information treatment. The matching algorithm either matches participants at random ("random algorithm") or groups together participants with similar farsighted reasoning abilities ("homogeneous algorithm"). Participants were not given any information about the matching algorithm being used. Information about the farsighted reasoning abilities of other group members is either made public among group members ("with information treatment") or not ("no information treatment").

3.2.1 Part I

In Part I of the experiment participants complete a questionnaire based on the Hit 15 game (a.k.a. "Race game"). The game consists of two players taking turns adding tokens to a basket. At the beginning of the game, the basket contains no tokens. At each turn, a player adds 1, 2 or 3 tokens (not adding any tokens is not allowed). The goal is to be the one who places the 15th token in the basket. After the game is explained, participants are asked to imagine that they are playing against a *"smart and experienced opponent who never misses an opportunity to win"* and answer the following 10 questions: *"There are currently x tokens in the basket and it is your turn. How many tokens should you add?"*, where $x = 13, 12, 6, 9, 5, 8, 2, 4, 1, 0$, in that order. They have 18 minutes to complete the questionnaire, are not allowed to go back and forth between questions, and are provided pens and paper for computations.

The instructions stress that every question has a correct answer. Participants receive 10 experimental currency units (= 1 Euro) for every correct answer they submit. The amount of earned experimental currency units is privately revealed to each participant at the end of the questionnaire. While we interpret those earnings as reflecting "farsighted reasoning scores", this terminology (or any other loaded terminology) is never used in front of participants.

The Hit 15 game has a dominant strategy: "add tokens to reach 4\7\11\15 whenever possible"[1] As a consequence, it provides a measure of one's ability to reason farsightedly that is relatively clean from beliefs about the reasoning abilities of others. The only caveat to this claim arises when a player believes her opponent will not take advantage of an opportunity to win when one is present (i.e. the belief that the opponent does not follow the dominant strategy described above). In this case, a player is indifferent between following the dominant strategy in its entirety or following it only at the later stages of the game. Our decision to ask questions about hypothetical play and have participants assume they are playing against an opponent "who never misses an opportunity to win" is meant to neutralize this caveat and allow us to interpret "wrong answers" as resulting from failure to reason farsightedly rather than as best-replies to some beliefs. Another reason to ask hypothetical questions is to avoid imitation behavior, i.e. allow us to interpret correct responses as resulting from genuine discovery of the dominant strategy rather than from observing others using it and imitating them. Note that the order in which questions are asked is important, as it is likely to affect the difficulty of discovering the dominant strategy. The order we chose was tested in a pilot to ensure it induces a well-balanced distribution of correct responses.

[1] Note that it is possible to reach one of 4\7\11\15 whenever the current number of token in the basket is not one of those, and that none of the question in the questionnaire ask participants to assume that it is one of those.

3.2.2 Part II

In Part II of the experiment, participants are matched into groups of four according to one of two algorithms. The first algorithm matches participants at random ("random algorithm"). The second algorithm ranks participants according to their farsighted ability scores and matches together the top four participants, next top four, etc' ("homogeneous algorithm"). Thus, the two algorithms result in a set of groups that differ from one another both in terms of the average scores (average level of foresight) and in terms of the variance of the scores (homogeneity in the level of foresight). Participants are not given any information about the matching algorithm being used. At the beginning of the game, every group member is assigned an identifying label (A, B, C, or D). Groups and identifying labels remain fixed throughout the game.

Each group plays four rounds of the following network formation game. At the beginning of each round ($t = 0$) no links exist. At every subsequent stage ($t > 0$), a player is randomly selected and is allowed to (unilaterally) update the linking status of a single link of her choice (i.e. form it in case it does not currently exist, or delete it in case it does), provided she is a side to it. Selected players may always choose not to effect any changes. Once the selected player has made her choice, the resulting "current network" is displayed to all group members via a graphical interface and each of them is asked whether they want to end the formation process at the current network. Players reply with either a YES or a NO. In case all group members say YES, the formation process concludes and the round ends. Otherwise, a new stage commences, meaning that another participant is randomly selected to make a change.[2] To ensure that an end is reached, a random stopping rule is implemented after stage 25: at every $t \geq 26$ the game ends with probability 0.2, regardless of whether agreement on ending the game is achieved.

Every network position in every network configuration is assigned a payoff, which are summarized in Table 3.1. What makes this succinct representation possible is the fact that payoffs depend only on positions and configurations, and in particular *not* on players' identities, i.e. the fact that the payoff function satisfies anonymity. Allocated payoffs are not meant to represent any real-life situation or network formation principles (e.g. costs of maintaining links, utility from indirect links, attention allocation etc').[3]

[2] The same player cannot be selected in two consecutive rounds, but no other restriction is imposed.

[3] In that sense, our game is not a network formation game in its essence and is merely represented as such.

Only payoffs from the networks at which rounds end ("final networks") are relevant for participants' earnings. In particular, earnings are equal to the sum of payoffs of two randomly drawn final networks (with a conversion rate of 10 payoff units = 1 Euro).

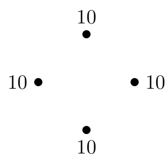
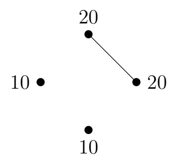
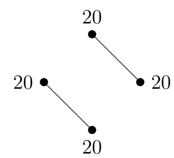
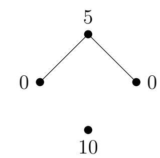
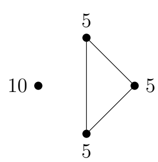
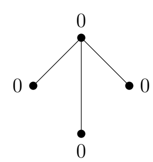
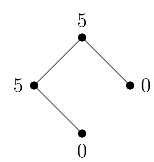
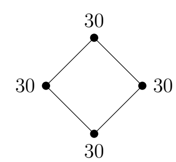
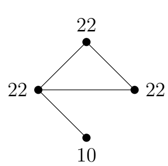
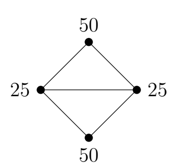
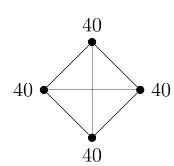
Class 1	**Class 2**	**Class 3**	**Class 4**
# players with 0 links: 4 # players with 1 links: 0 # players with 2 links: 0 # players with 3 links: 0 Example and payoffs: 	# players with 0 links: 2 # players with 1 links: 2 # players with 2 links: 0 # players with 3 links: 0 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 4 # players with 2 links: 0 # players with 3 links: 0 Example and payoffs: 	# players with 0 links: 1 # players with 1 links: 2 # players with 2 links: 1 # players with 3 links: 0 Example and payoffs: 
Class 5	**Class 6**	**Class 7**	**Class 8**
# players with 0 links: 1 # players with 1 links: 0 # players with 2 links: 3 # players with 3 links: 0 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 3 # players with 2 links: 0 # players with 3 links: 1 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 2 # players with 2 links: 2 # players with 3 links: 0 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 0 # players with 2 links: 4 # players with 3 links: 0 Example and payoffs: 
Class 9	**Class 10**	**Class 11**	
# players with 0 links: 0 # players with 1 links: 1 # players with 2 links: 2 # players with 3 links: 1 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 0 # players with 2 links: 2 # players with 3 links: 2 Example and payoffs: 	# players with 0 links: 0 # players with 1 links: 0 # players with 2 links: 0 # players with 3 links: 4 Example and payoffs: 	

Table 3.1: Payoffs from network classes

The instructions for Part II of the experiment are given only after all participants complete Part I. Table 3.1 is included in the instructions so that participants know the payoff structure before play begins. In addition, they are encouraged to familiarize themselves with it before starting to play and consult it during play. After instructions are read, participants are required to solve a quiz about payoffs from various networks.

All participants must answer all questions correctly for play to begin. A hint is provided whenever a participant submits a wrong answer. To further help participants navigate the payoffs associated with each network, a "simulation tool" is available for them to use at any time during play. It consists of a graphical representation of a network over which participants can freely form\delete any link. The payoffs of all players from the simulated network are always displayed. A button to copy the "current network" into the simulation tool is included, thus allowing participants to quickly explore various continuation paths from the current state of the actual network. Screenshots of the interface are included in Appendix 3.6.

We implement two information conditions. In the "no information" treatment, no information about other group members' earnings in Part I is revealed. In the "with information" treatment, the earnings of each group member in Part I are presented in the graphical interface next to their identifying letter. Participants in the "with information" treatment are aware that the numbers next to the identifying letters represent earnings in Part 1, but are not directed to interpret those numbers in any particular way. After the game is completed, participants are asked to complete a short questionnaire about their understanding of the game, the strategies they chose and demographic information.

The network formation game described above closely follows the one used in Kirchsteiger et al. (2016), which deploys it in order to test the performance of myopic and farsighted solution concepts. Our game differs from theirs in three main aspects: first, the payoff structure; second, unlike them, we allow for unilateral link formation; third, while they reshuffle group members' identifying letters between rounds, we do not. The modification regarding the payoff structure is meant to put in place the incentive structure we are interested in studying (see paragraph below). The modification regarding unilateral link formation is meant to reduce the complexity of the game. The modification regarding the identifying letters is meant to amplify participants' opportunity to learn about other individuals in their group (rather than just about aggregate behavior).

The payoff structure was designed to induce tension along three dimensions of decision-making and strategic interactions: foresight vs. myopia, coordination vs. miscoordination, and cooperation vs. competition. Class 11 is clearly a very attractive candidate for a final outcome as it is egalitarian and socially efficient. Class 10, however, may be seen as a threat to its stability, as any player can, by deleting one of her links, move from Class 11 to Class 10 and obtain 50 instead of 40 (assuming the game will end immediately after their move). In other words, conditional on arriving at Class 10, players may either choose to compete on obtaining 50 or cooperate to obtain an equal payoff of 40. Class 8 represents a way out from this conundrum – players who do not want to engage in competing over obtaining 50 or do not believe others would cooperate to obtain 40 may aim to make Class 8 a final outcome, yielding a payoff of 30 for everybody. This would be an unfortunate example of miscoordination, as all players are better off in Class 11 compared to Class 8 (i.e. Class 11 strictly Pareto dominates Class 8). Lastly, Class 3 may turn out as a final outcome in case players strictly follow myopic reasoning, i.e. execute moves if and only if they maximize their immediate payoff, ignoring the fact that further moves might take place. Indeed, the unique myopically rational move from the starting point (Class 1) is to Class 2, the unique myopically rational move from Class 2 is to Class 3, and there are no myopically rational moves away from Class 3. Clearly, ending up at Class 3 would represent an even more severe case of miscoordination, as classes 8, 10 and 11 all strictly Pareto dominate it. The next section formally introduces solution concepts that predict classes 3, 8, 10, or 11 to be final states.

3.3 Theoretical Analysis and Hypotheses

In this section, we define and apply several solution concepts to the experimental game described above, so as to generate competing hypotheses on the dynamics and final outcomes that would be observed. We divide solution concepts into three broad camps: *(i)* "myopic", which assume players only take into account the immediate consequences of their actions; *(ii)* "farsighted without preemptive actions", which assume players take into account the entire chain of reactions that their own actions might trigger, but presuppose that if they will not act, no one else would either; *(iii)* "farsighted with preemptive actions", which assume players take into account the entire chain of reactions that their own actions *or inaction* might trigger. Myopic solution concepts include, for instance, the Core. Solution concepts in the second category include, for instance, the Farsighted Stable Set, the Largest Consistent Set (Harsanyi, 1974; Chwe, 1994) and the Rational Expectation Farsighted Stable Set (Dutta and Vohra, 2017)). Most farsighted solution concepts belong to this category. Solution concepts in the third category, which have only started to appear recently, include the Equilibrium Stable Set (Karos and Robles, 2021), the Subgame Perfect Consistent Set (Granot and

Hanany, 2022) and the Set-Valued Rational Expectations (SVRE) (Dekel, 2023).

In order to describe these solution concepts in a unified framework, we formulate the experimental game as an "abstract game" (a.k.a. "game in effectivity function form") and use Dekel (2023)'s definition of a set-valued expectation. Since all solution concepts within a given category produce the same predictions for the game at hand, we only describe one in each. As a representative of the myopic solution concepts category we use the "Myopic SVRE" concept defined in Dekel (2023). As a representative of the second category we use the "Rational Expectations Farsighted Stable Set" defined in Dutta and Vohra (2017), which we refer to here as "Dutta-Vohra SVRE". As a representative of the third category we use "SVRE", defined in Dekel (2023).

3.3.1 Abstract Game Representation

An abstract game is defined by $\Gamma = (N, Z, E, \{u_i\}_{i \in N})$, where:

- N is a set of n players.
- Z is a finite set of outcomes\states. Elements in this set are denoted by a, b, c... etc'.
- E is a correspondence from $Z \times Z$ to $\mathcal{N}$ (the set of all subsets of N) describing, for every ordered pair of states, which coalitions can replace the first by the second. If a coalition $S \in \mathcal{N}$ belongs to $E(a, b)$ we say that S "is effective in the move" (or "can move") from state a to state b.
- u_i is a function from Z to $\mathbb{R}$ describing player i's utility from each state.[4]

To formulate our network formation game as an abstract game, we let Z be the set of all possible network configurations. For any two networks $z, z' \in Z$, $z \neq z'$ we let $S \in E(z, z')$ if and only if z and z' differ from one another by exactly one link, $S = \{i\}$ and i is a side in the link that distinguished z from z'. For any network z, we have $S \in E(z, z)$ if and only if either $|S| = 1$ or $|S| = 4$. The former is due to the fact that players can always choose to leave all their links unchanged. The latter is due to the fact that at the end of every stage, players are asked whether they want to end the game, which means that the group as a whole can coordinate on staying at any $z \in Z$.

[4]An abstract game can be represented as a directed graph, where nodes represent states, and labeled directed edges represent the effectivity correspondence, i.e. a directed edge from z to z' labeled S signifies that S is effective in the move from z to z'. It is assumed that the directed graph induced by Z and E is weakly connected (otherwise we treat each graph component as a separate game).

3.3.2 Solution Concepts

The setup described in Dekel (2023) allows discussing different types of myopic and farsighted behaviors in a single unified framework, and is therefore suitable for the purposes of the current paper. We start by reiterating it here. (z, z', S) is used to denote a move from state z to state z' by coalition S. $M = \{(z, z', S) \mid S \in E(z, z')\}$ denotes the set of all possible moves. Given M and any subset $m \subseteq M$, $M(a) = \{(a, b, S) \in M\}$ (respectively, $m(a) = \{(a, b, S) \in m\}$) denotes the set of moves from a that are in M (respectively, m), and $M_i(a) = \{(a, b, S) \in M \mid i \in S\}$ (respectively, $m_i(a) = \{(a, b, S) \in m \mid i \in S\}$) denote the set of moves from a that are in M (respectively, m) and i is involved in. m_i denotes the set of moves in m that i is involved in (from any state). $Y(m) = \{a \in Z \mid \nexists (a, b, S) \in m \text{ s.t. } b \neq a\}$ denotes the set of stationary states under m and $Y(a, m)$ denotes the set of stationary states that are reachable from a via moves in m.[5]. A set-valued expectation is defined as a subset of M that describes all players' (intended) behavior at every state. Formally, $m \subseteq M$ is a set-valued expectation if for all $i \in N$ and $a \in Z$, $m_i(a) \neq \emptyset$. Let M^e denote the set of all set-valued expectations.

Myopic SVRE

A set-valued expectation is *myopically* rational if it prescribes only myopically rational moves. The formal definition is given below. As pointed out by Dekel (2023), for an appropriately defined effectivity correspondence, the set of states supported as stationary by all myopic SVREs coincides with the set of Nash equilibria. Note, however, that myopic SVREs contain more information than just a prediction about the final states that could be reached – they also tell us which subset of the set of potential final states could be reached given the current state ($Y(a, m)$, where a is the current state) and which moves are expected to follow from it ($m(a)$).

Definition 3.1. *A set-valued expectation $m \in M^e$ is myopically rational (myopic SVRE) if no coalition T has a feasible and myopically profitable deviation to an alternative set-valued expectation. Formally, for any $a \in Z$, there do not exist $R(a) \subseteq m(a)$ and $(a, c, T) \in M \setminus R(a)$, such that:*

---

[5] Formally,

$$Y(a, m) = \{a^K \in Z \mid \exists (a^0, ..., a^K) : a^0 = a,\ a^K \in Y(m), \text{and}, \forall k \in \{1, ..., K\},\ \exists (a^{k-1}, a^k, S^k) \in m\}$$

(i) For all $(a, b, S) \in R(a)$, $T \cap S \neq \emptyset$.

(ii) For all $i \in T$, *there exists* $(a, b, S) \in m_i(a)$ *such that* $u_i(c) > u_i(b)$;

Dutta-Vohra SVRE

A Dutta-Vohra SVRE is defined as follows.

Definition 3.2. *A set-valued expectation* $m \in M^e$ *is a Dutta-Vohra SVRE if:*

- *It is "essentially single-valued", i.e. it prescribes at most one move away from any state. Formally, for all* $a \in Z$, $|\{(a, b, S) \in M \mid b \neq a\}| \leq 1$;
- *It is absorbing, i.e. the continuation path it prescribes away from any state terminates at a stationary state. Formally, for all* $a \in Z$, $Y(a, m) \neq \emptyset$;
- *It is internally stable, i.e. no coalition is effective in making a profitable move away from any stationary state. Formally, for any* $a \in Y(m)$ *there does not exist* $(a, b, S) \in M$ *such that* $u_i(B) > u_i(a)$ *for all* $i \in S$*, where* B *denotes the unique state in* $Y(b, m)$;
- *It is externally stable, i.e. for any non-stationary state, the path* m *prescribes away from it terminates at a state all members of the moving coalition prefer over the status quo. Formally, for all* $a \notin Y(m)$, $u_i(B) > u_i(a)$ *for all* $i \in S$*, where* B *denotes the unique stationary state in* $Y(b, m)$, S *denotes the unique coalition prescribed to move away from* a*, and* b *denotes the unique state it is prescribed to move to;*
- *It is maximal, i.e. for any non-stationary state, the move* m *prescribed away from it is optimal for the coalition executing it. Formally, for all* $a \notin Y(m)$*, there does not exist* $(a, c, S) \in M$ *such that* $u_i(C) > u_i(B)$*, where* C *is the unique state in* $Y(c, m)$, S *is the unique coalition prescribed to move away from* a, b *is the unique state it is prescribed to move to, and* B *is the unique state in* $Y(b, m)$.

The requirement that m is essentially single-valued restricts m to prescribe at most one move away from every state. The requirement that m is absorbing reflects an implicit assumption that players always prefer to agree on some final state over cycling among several.[6] The internal and external stability requirements originate from the

[6] This assumption is justified ex-post by the fact that almost all rounds played in the lab end with a unanimous agreement on the final state (only 5% of rounds terminate by the random stopping rule).

standard definition of the Farsighted Stable Set, which can be further traced back to vNM's definition of (myopic) stable sets. Note that according to the external stability condition, players compare the final state their move would lead to ($Y(b, m)$) *against the status quo* (a). This implies that they do not take into account the possibility that others would move away from the status quo, which in turn means they would never be interested in executing a move on a preemptive pretext, i.e. only to prevent others from making a different move. Lastly, The maximality requirement ensures that coalitions make moves that are optimal for them (note that the external stability condition only requires moves to be improving).

SVRE and Symmetric SVRE

The notion of SVRE defined in Dekel (2023) relaxes the requirement that m is essentially single-valued: it allows m to prescribe multiple moves away from every state. As a consequence, $Y(a, m)$ (for any a) contains, in general, multiple states. Comparing the relative attractiveness of different moves therefore requires comparing the *sets* of stationary states they may lead to. Hence, an extension rule from preferences over states to preferences over sets of states should be put in place. Dekel (2023) takes the following assumption.

Assumption 3.1. *The following holds for all $i \in N$:*

(i) For any non-empty set $\emptyset \neq A \subseteq Z$ and for all $b \in Z \setminus A$:

1. *If $u_i(b) = u_i(a)$ for all $a \in A$, then $A \cup \{b\} \sim_i A$*
2. *If $u_i(b) \geq u_i(a)$ for all $a \in A$ and $u_i(b) > u_i(a)$ for some $a \in A$, then $A \cup \{b\} \succ_i A$*
3. *If $u_i(b) \leq u_i(a)$ for all $a \in A$ and $u_i(b) < u_i(a)$ for some $a \in A$, then $A \succ_i A \cup \{b\}$*

(ii) For any non-empty set $\emptyset \neq A \subseteq Z$, $A \succ_i \emptyset$;

The first part reflects the idea that players always prefer adding (removing) the possibility of ending up at a state that is better (worse) than those currently "on the table". Note that it implies, among other things, the most basic requirement one could ask an extension rule to fulfill, i.e. that for any two states $a, b \in Z$, $\{a\} \succ_i \{b\}$ if and only if $u_i(a) \geq u_i(b)$. The second part of the assumption in some way parallels the absorption requirement included in the definition of Dutta-Vohra SVREs, as it, too, reflects the idea that players always prefer to agree on some final state over cycling between states forever. It is therefore justified on the same grounds.

Given preferences over sets that satisfy these assumptions, an SVRE is defined as follows, where $m_i(a)$ is said to be *dynamically consistent* if for all $a \in Z$, for all $(a, b, S) \in m_i(a)$ there exists $(a, c, T) \in m_i(a)$ such that, letting $m' = [m \setminus m_i(a)] \cup \{(a, c, T)\}$ we have $Y(b, m) \sim_i Y(c, m')$.

Definition 3.3. *A set-valued expectation $m \in M^e$ is SVRE if:*

(DC) It is dynamically consistent. i.e. $m_i(a)$ is dynamically consistent for all $a \in Z$ and $i \in N$;

(OP) It is optimal, i.e., no coalition T has a feasible and profitable deviation to an alternative set-valued expectation that is dynamically consistent. Formally, for any $a \in Z$, there does not exist $m' = [m \setminus R(a)] \cup \{(a, c, T)\}$, where $R(a) \subseteq m(a)$ and $(a, c, T) \in M \setminus R(a)$, such that:

(i) For all $(a, b, S) \in R(a)$, $T \cap S \neq \emptyset$;

(ii) For all $i \in T$, there exists $(a, b, S) \in m_i(a)$ such that $Y(c, m') \succ_i Y(b, m)$;

(iii) $m'_i(a)$ is dynamically consistent for all $i \in T$

Definition 3.4. *A set-valued expectation $m \in M^e$ is a symmetric SVRE if it is SVRE and symmetric, i.e. for any state $a \in Z$, if i and j have the same degree in a then $m_i(a) = m_j(a)$.*

3.3.3 Predictions on Final States

Proposition 3.1 provides the predictions on the final states that the solution concepts defined above produce when applied to the experimental game. Myopic SVREs predict the game would end at class 3. Dutta-Vohra SVREs predict it will end at class 8. SVREs predict it will end at either class 10 or class 11. Symmetric SVREs predict it will end at class 11.

Proposition 3.1. *Assume players have no other-regarding preferences and let g^k denote the set of networks that belong to class k.*[7] *Then, in our network formation game:*

[7] To avoid complicating the notation, whenever g^k is a singleton set we use this notation to denote both the singleton set and the unique network it contains.

(i) $\bigcup_{m\in m^m} Y(g^1, m) = g^3$, *where* m^m *is the set of myopic SVREs.*

(ii) $\bigcup_{m\in m^{f_1}} Y(g^1, m) = g^8$, *where* m^{f_1} *is the set of Dutta-Vohra SVREs.*

(iii) $\bigcup_{m\in m^{f_2}} Y(g^1, m) = g^{10} \cup g^{11}$, *where* m^{f_2} *is the set of SVREs.*

(iv) $\bigcup_{m\in m^{f_2'}} Y(g^1, m) = g^{11}$, *where* $m^{f_2'}$ *is the set of symmetric SVREs.*

Proof. See Appendix 3.7. □

The intuition behind this result is as follows. For myopic SVREs, the only paths emanating from the empty network that are composed only of myopically optimal moves terminate at networks in class 3. Hence, if the game is played by myopic players, they are expected to reach class 3 and never move away from it. For Dutta-Vohra SVREs, first note that the external stability condition rules out any m prescribing a move away from class 10 by the players in the position with a degree of 2 (yielding a payoff of 50), as those moves can only set in motion paths that terminate at states that provide those players with a weakly lower payoff (as 50 is the highest possible payoff one could obtain in this game). Since class 11 can only be reached from class 10, this means that no Dutta-Vohra SVRE allows reaching class 11 from any network in any other class. Given that, the network in class 8 must be stationary, as no move away from it can set in motion a path that would terminate at a strictly higher payoff for the moving player.

When it comes to (plain) SVREs, no external stability requirement is imposed. As a result, prescriptions to move from class 10 to class 11 are not ruled out. Moreover, SVREs allow for multiple moves to be prescribed away from every state. Consider then an m prescribing all players to move away from all states besides the network in class 11, implying this network is uniquely stationary and will be reached from any other network. No matter what any individual player does, obtaining a payoff of 50 is out of reach, as other players are expected to move away from any network in class 10. Hence, in such an m all individual players behave optimally. Moreover, the four players as a whole cannot agree on not moving away from any of the non-stationary networks, as none of them Pareto dominates the stationary one. Hence, this m is SVRE. By the exact same argument, an m that prescribes all players to move away from all states besides some network in class 10 is also SVRE. Any m implying that some network in classes 1-9 is stationary is not an SVRE: if it does not imply the stationarity of any network in classes 10 or 11, the four players as a whole can beneficially make one of them stationary (by coordinating on not to move away from it); if it implies the stationarity of some network in classes 1-9 *in addition* to some networks in classes 10 or 11, at least one player can beneficially make it non-stationary (by simply moving away from it).

Lastly, note that any SVRE *either* supports some network in class 10 as stationary *or* the network in class 11, i.e. those two predictions arise from *separate* SVREs. To establish this claim, observe first that no SVRE m prescribes multiple networks in class 10 to be stationary simultaneously. Had that been the case, there must exist a player obtaining 25 in one stationary network and 50 in another, and that player would have had an incentive to make the former non-stationary (by simply moving away from it). This in turn contradicts the assumption that m is SVRE. Now suppose m is SVRE and supports some network in class 10 as stationary *on top* of the one in class 11. A player obtaining 50 in the stationary network in class 10 has an incentive to make the one in class 11 non-stationary, which, again, contradicts the assumption that m is SVRE.

3.3.4 Predictions on Paths of Play

The solution concepts described above provide predictions not only on the final state that would be achieved but also on the paths of play that would take place. A move $(a, b, \{i\})$ belongs to the set of paths predicted by some Myopic SVRE m only if it is myopically rational, i.e. $u_i(b) \geq u_i(a)$. It belongs to the set of paths predicted by some Dutta-Vohra SVRE m only if $a \notin g^8 \cup g^{11}$ and $u_i(a) < 30$. Hence, all moves can be rationalized by the Dutta-Vohra SVRE concept besides those made by players who gain at least 30 in the current network. The set of paths predicted by some SVRE (respectively, symmetric SVRE) is the set of all paths that terminate at some state within $g^{10} \cup g^{11}$ (respectively, g^{11}). Hence, all moves can be rationalized by the SVRE (or symmetric SVRE) concept (even those that replace states in $g^{10} \cup g^{11}$ by another).[8] The reason for the permissiveness of the predictions associated with the farsighted solution concepts is that players are assumed to care only about the payoff they receive at the final state and in particular do not incur any "moving costs". Hence, the length of the path of play is payoff-irrelevant and the only thing that matters is the final state that is eventually achieved.

[8] Note that we are not using the term "path" in its graph-theoretic meaning, as it also refers to walks that include cycles. The precise graph-theoretic term to be used here is "walk".

3.3.5 Hypotheses

All solution concepts above assume a homogeneous population of players in terms of their myopia\foresight ("homogeneity assumptions"). Farsighted solution concepts assume, in addition, that the fact that all players are farsighted is common knowledge ("common knowledge assumption"). Thus, strictly speaking, the group-level predictions produced by Myopic SVREs should only apply to homogeneous groups, and the group-level predictions produced by Dutta-Vohra SVREs and (plain) SVREs should only apply to homogeneous groups subject to the "public" information condition. This leads to the following hypotheses.

Hypothesis 3.1. *Groups composed of participants with low farsighted reasoning scores tend to end the game at networks in class 3.*

Hypothesis 3.2. *Groups composed of participants with high farsighted reasoning scores who are subject to the "public" information condition tend to end the game at networks in class 8.*

Hypothesis 3.3. *Groups composed of participants with high farsighted reasoning scores who are subject to the "public" information condition tend to end the game at networks in classes 10 or 11.*

Hypothesis 3.4. *Groups composed of participants with high farsighted reasoning scores who are subject to the "public" information condition tend to end the game at networks in class 11.*

These hypotheses can be related to the concepts of coordination and cooperation. Hypotheses 3.1 and 3.2 imply coordination failures in the sense that players fail to coordinate on a Pareto efficient network (i.e. one that belongs to classes 10 or 11). Hypothesis 3.3 implies that players are successful in solving this coordination problem. Conditional on successful coordination, players may either choose to compete over obtaining a payoff of 50 (while others obtain 25) or cooperate to obtain a payoff of 40 for all. Hypothesis 3.4 implies that players choose cooperation. Measuring the effect of mixing high and low-score participants and of concealing information on others' scores (i.e. of violating the homogeneity and\or common knowledge assumptions) is possible thanks to the random variation in groups' composition and information condition.

Concerning paths of plat, while the predictions produced by Myopic SVREs are rather restrictive (only myopically rational moves), the predictions produced by the farsighted solution concepts are very permissive. As noted in Subsection , this results from the (unrealistic) assumption that players only care about the final networks that emerge, which in turn implies indifference between short and long paths that terminate as the same final network. Had players preferred shorter paths (e.g. if executing a move entailed a small "moving cost") the farsighted solutions would have predicted players to move efficiently toward the target network and not move away from it once achieved. We base the hypotheses below on this alternative prediction.

Hypothesis 3.5. *Participants with low farsighted reasoning scores tend to make moving decisions that are myopically rational more often than participants with high farsighted reasoning scores.*

Hypothesis 3.6. *Groups composed of participants with high farsighted reasoning scores converge to a final network faster than mixed groups composed of both low-score and high-score participants.*

Lastly, we formulate a hypothesis concerning the Hit-15 questionnaire. The crucial aspect of the questions it contains is that they vary in the number of steps ahead the responder is required to compute in order to answer correctly. For example, questions assuming the current number of tokens in the basket is 12, 13 or 14 do not require the respondent to consider any steps ahead, as she can guarantee a win in one step (by adding 3, 2 or 1 tokens, respectively), while questions assuming the current number of tokens in the basket is 8, 9 or 10 require the respondent to consider one step ahead, as answering correctly requires taking into account one move by the opponent. Similarly, questions assuming the current number of tokens is 4, 5 or 6 (respectively 0, 1 or 2) require thinking two (respectively, three) steps ahead. Since the ability to compute k steps ahead implies the ability to compute $k-1$ steps ahead, one should expect the average rate of correct responses to decrease as the number of steps ahead required to be computed increases.

Hypothesis 3.7. *In the Hit-15 questionnaire, the average rate of correct responses is negatively related to the number of steps ahead required to be computed.*

3.4 Implementation and Results

3.4.1 Implementation

The recruitment was conducted from the Parisian Experimental Economics Laboratory (LEEP) participants pool using ORSEE (Greiner, 2015). All sessions took place in the LEEP Experimental Lab at the Maison de Sciences Economique of the Paris 1 Panthéon-Sorbonne University. A total of 120 participants took part in the experimental sessions. They earned on average around 18.6 Euros, including 5 Euros for participating. The experiment was programmed using oTree (Chen et al., 2016).

We conducted three sessions for the "public" information treatment and three sessions for the "private" information treatment. Among each of those, two sessions used the homogeneous matching algorithm and the remaining one used the random matching algorithm. Table 3.2 reports the dates of the sessions and the number of participants per session. We chose to conduct more "random" than "homogeneous" sessions in order to obtain a well-balanced distribution of group compositions, bearing in mind that the "random" condition has the potential to generate (by chance) groups composed of members with similar farsighted reasoning scores.

Table 3.2

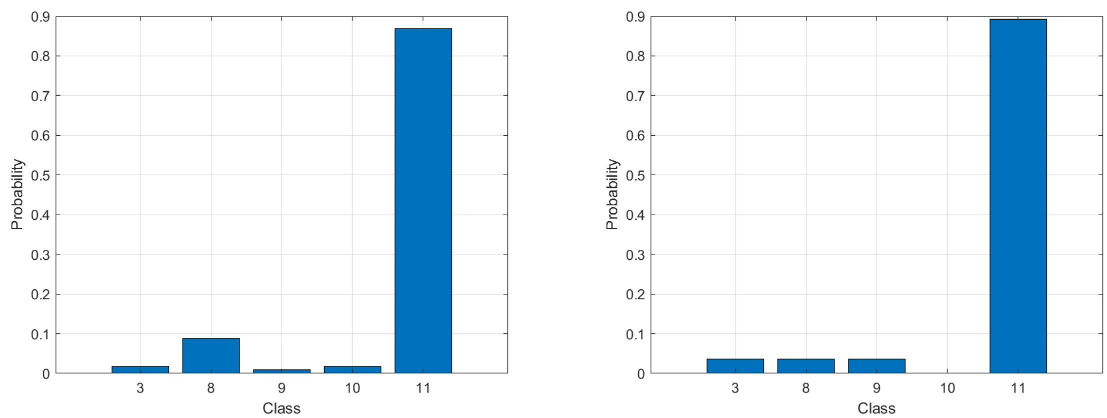
Session	Date	Information treatment	Matching algorithm	# Participants
1	08/02/2023	Public	Homogeneous	20
2	08/02/2023	Private	Homogeneous	24
3	09/02/2023	Public	Random	20
4	09/02/2023	Private	Random	20
5	20/02/2023	Private	Random	16
6	20/02/2023	Public	Random	20

3.4.2 Final Networks

Figure 3.1 presents the distribution of final networks over all rounds that terminated endogeneously (i.e. not by the random stopping rule).[9] It shows that the vast majority of rounds (87%) terminate at class 11. To show that this is not driven by participants' learning from previous rounds, Figure 3.1b presents the distribution of final networks

[9] Only 6 out of 120 rounds were terminated by the random stopping rule.

only in the first round played by every group. It exhibits the same pattern. These statistics clearly go against Hypotheses 3.3 and 3.2 and in favor of Hypotheses 3.3 and 3.4. They therefore lend empirical support to symmetric SVREs. Due to the low variability in final networks, we avoid analyzing how they are affected by groups' compositions and information disclosure. This analysis is pursued in Subsection 3.4.6, however, by considering individual decisions to support networks as final rather than the aggregate outcome they induce. What makes it possible is the increased level of variability that these individual-level decisions present.



(a) All rounds. **(b)** First round only.

Figure 3.1: Distribution of final networks. Rounds that were terminated by the random stopping rule are excluded.

Result 3.1. *The networks predicted by the solution concepts we consider account for 99% of final networks (115 out of 114 rounds). Myopic SVREs account for 1.7% (2 out of 114 rounds). Dutta-Vohra SVREs account for 8.6% (10 out of 114 rounds). SVREs account for 88.6% (101 out of 114 rounds). Symmetric SVREs account for*

Hypothesis 3.4 explains participants' attraction to class 11 by the fact that it is supported by *symmetric* expectations and makes no appeal to other-regarding preferences. Alternatively, one could try to explain this result by assuming participants care about others' payoff and are, in particular, inequality averse. Assuming participants prefer obtaining a payoff of 40 when the remaining players obtain 40, 40 and 40, over obtaining a payoff of 50 when the remaining players obtain 50, 25 and 25, all farsighted solution concepts coincide in predicting the network in class 11 to be uniquely stationary.

While our experimental design does not allow determining participants' preference relation over these two alternatives, their reply to the "ending question" when occupying the position with degree 2 in class 10 (yielding a payoff of 50) offers a clue. The average rate of a YES ending decision when participants are in this position is 0.74. Strictly speaking, we have no way of determining their beliefs about where the game would end if they answered NO. However, given the high rate of rounds terminating in class 11, it is reasonable to assume that this is the counterfactual they have in mind. Under this assumption, a YES ending decision at the position with degree 2 in class 10 can be interpreted as reflecting a preference for the unequal (50,50,25,25) payoff vector (where the concerned individual obtains 50) over the equal (40,40,40,40) payoff vector, i.e. in 74% of the cases participants state they prefer the former over the latter. Hence, inequality aversion does not seem to be a convincing explanation for the observed convergence to class 11.

3.4.3 Farsighted Reasoning Task

A participant's farsighted reasoning score is defined as the number of correct responses she submits in the Hit-15 questionnaire multiplied by 10. Figure 3.2 presents the scores' distribution. It is bell-shaped, and shoes that the minimum score is 20 and the maximal score is 100.

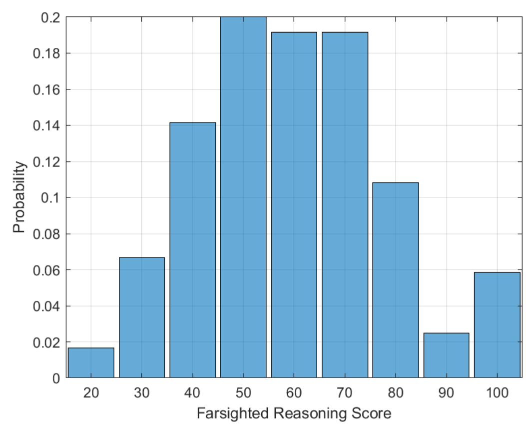


Figure 3.2: Distirbution of farsighted reasoning scores.

According to Hypothesis 3.7, we should expect to observe a negative relation between the required level of reasoning a question entails, ranging from 0 to 3, and the probability that participants answer it correctly. Table 3.3 reports results from regressing the former on the latter. The negative coefficient of "Required level of reasoning" supports the stated hypothesis.

Table 3.3

	Dependent variable:	
	Correct response=1	
	(1)	(2)
Required level of reasoning	−0.1833***	−0.1833***
	(0.0117)	(0.0113)
Constant	0.9083***	0.8117***
	(0.0238)	(0.1371)
Individual FE	No	Yes
Observations	1,200	1,200

Note: A linear probability model. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Result 3.2. *In the Hit-15 questionnaire, the level of reasoning a question requires is negatively correlated with the probability that participants answer it correctly.*

3.4.4 Moving Decisions

At every stage of the network formation game one randomly chosen group member is offered an opportunity to change the status of one of her links. In particular, the selected group member may: (1) form a link with a player she is not currently linked with; (2) delete an existing link with another player; (3) not make any changes. We refer to this type of decision as "moving decisions" as they have the potential to move the formation process from one network configuration to another. According to Hypothesis 3.5, we should expect participants with low farsighted reasoning scores to make moving decisions that are myopically rational more often than participants with high farsighted reasoning scores.

Figure 3.3 displays the average rate of myopically rational moves against scores, which suggests an overall negative relationship. As a more rigorous test, we estimate a linear probability regression model of the probability that a participant executes a myopically rational move on her score.[10] Table 3.4 reports the results.

Column (1) presents the simplest possible specification, which includes only the scores and a constant. Since the decision problems participants face are not randomly

[10] We choose a linear probability model instead of a non-linear model (e.g., logit) because its coefficients are easier to interpret.

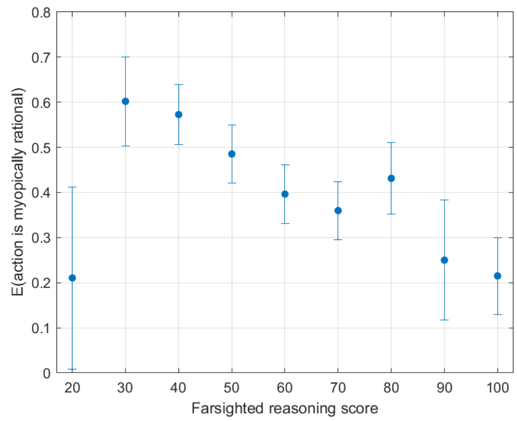


Figure 3.3: Average rate of myopically rational moving decisions by farsighted reasoning scores.

allocated but rather a result of the endogenous network formation process, one might be concerned that the decision problems that low-score individuals tend to face are somehow different from those high-score individuals tend to face.[11] The other specifications in the table control for this by introducing position fixed-effects. A *position* is defined by a combination of the class the current network belongs to and the decision maker's degree in it.[12] Hence, every position is associated with a unique decision problem and conditional on position the decision problem faced by participants is identical. The magnitude of the coefficient associated with the score drops, but remains highly significant.

Column (3) introduces into the regression a dummy variable taking the value 1 if the stage is above 25. Recall that while up to stage 25 a round ends only when all group members agree on it, after stage 25 it ends anyway with probability 0.2. One might hypothesize that the introduction of a random stopping rule would induce participants to behave more myopically, as any network could be final. In contrast to Carrillo and Gaduh (2021), we find no evidence for such an effect: participants seem to behave in the same level of myopia with and without the random stopping rule.

Column (4) adds two controls: the participants' gender and whether or not she reported she plays chess regularly. Results remain stable. The interpretation of the

[11]For example, it is evident from the data that conditional on reaching Class 11, all participants (regardless of their score) choose the myopically irrational move "leave all links unchanged" with a very high probability. But it might be the case that groups composed of high-score individuals tend to arrive at Class 11 more often. If so, the coefficient reported in column (1) would be inflated.

[12]There are 20 positions in total.

Table 3.4

	Dependent variable:			
	Myopically rational move=1			
	(1)	(2)	(3)	(4)
Score	-0.0045^{***}	-0.0013^{***}	-0.0013^{***}	-0.0012^{**}
	(0.0009)	(0.0005)	(0.0005)	(0.0005)
Stage>25			0.0199	0.0218
			(0.0396)	(0.0394)
Female=1				0.0184
				(0.0190)
Plays chess				0.0117
				(0.0229)
Constant	0.7076^{***}	0.9206^{***}	0.9202^{***}	0.9042^{***}
	(0.0640)	(0.0455)	(0.0457)	(0.0520)
Position FE	No	Yes	Yes	Yes
Observations	1,302	1,302	1,302	1,302

Note: The model is estimated using a linear probability model. Standard errors (in parentheses) are clustered at the group level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

coefficient from the specifications with the position fixed-effects is that answering correctly one additional question in the farsighted reasoning questionnaire (i.e. an increase of 10 points in the score) is associated with a decrease of approximately 1 percentage point in the probability of executing a myopically rational move.

Thus, we conclude the following result, which supports Hypothesis 3.5.

Result 3.3. *The farsighted reasoning ability measured by the proposed Hit-15 questionnaire is negatively correlated with making myopically rational moves.*

3.4.5 Speed of Convergence

For every round that ends endogenously, we define "# stages to convergence" as the number of stages until the final network is reached. According to Hypothesis 3.6, we should expect groups composed of only high-score participants to converge to a final network faster than groups composed of both low and high-score participants. One way of testing this hypothesis is to compare "# stages to convergence" among groups with low variance and high average in farsighted reasoning scores to "# stages to

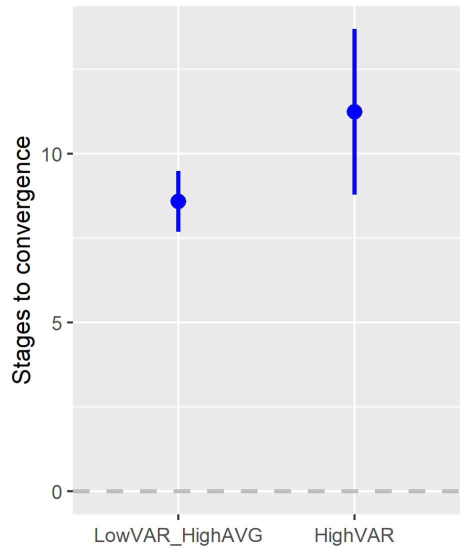


Figure 3.4: Means and 95% confidence intervals (computed based on standard errors clustered at the group level) for: (1) groups with $Variance(\textit{farsighted reasoning scores})$ below the 0.25 quantile and $Average(\textit{farsighted reasoning scores})$ above the median (there are 19 such groups in the data); (2) groups with $Variance(\textit{farsighted reasoning scores})$ above the 0.75 quantile (there are 30 such groups in the data).

convergence" among groups with high variance in farsighted reasoning scores. Figure 3.4 presents averages and 95% confidence intervals (clustered at the group level) assuming that "low variance in farsighted reasoning scores" means below the 0.25 quantile its the distribution, "high variance" means above the 0.75 quantile, and "high average farsighted reasoning scores" means above the median of its distribution. The p-value of the associated F-test for difference in means is 0.046, implying the null is rejected at the 5% significance level.

Since the choice of variance quantile in the exercise above is somewhat random, we repeat it for a range of quantiles q between 0.5 and 1. For each q, "low variance" is interpreted as "below the $1-q$ quantile" and "high variance" is interpreted as "above the q quantile". The p-values and differences in means associated with each q are reported in Figure 3.5. As q increases, the p-values get smaller and are strictly below the 5% significance threshold for any $q > 0.73$. Moreover, given that the average "# stages to convergence" across all (endogenously terminated) rounds is 10.2, the effect sizes are large.

We conclude the following result, which supports Hypothesis 3.6.

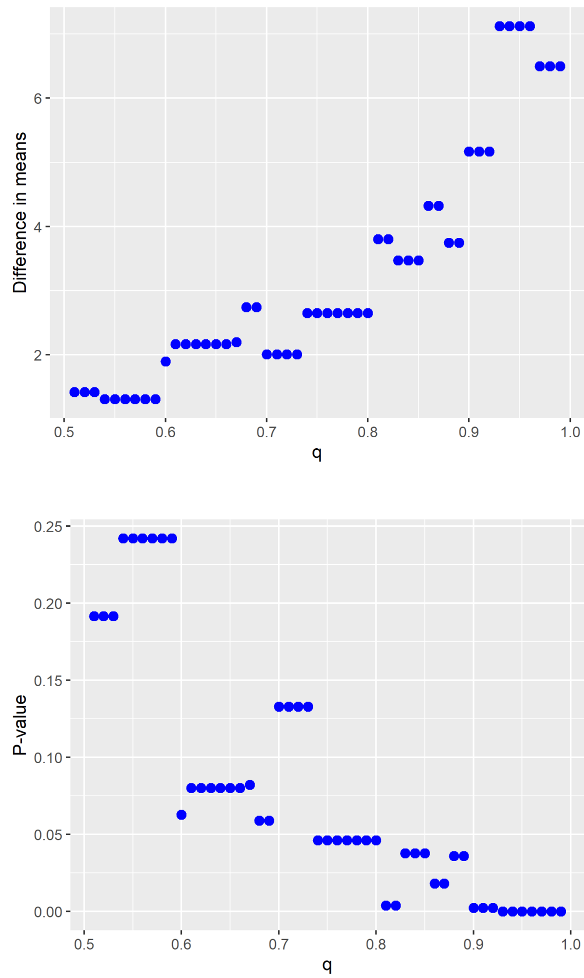


Figure 3.5: Effects of group composition and p-values for a range of variance quantiles.

Result 3.4. *Groups composed of members with high farsighted reasoning scores tend to achieve convergence about 2-7 stages faster than groups composed of both low-score and high-score participants. The disparity grows for groups at the far ends of the variance distribution.*

3.4.6 Ending Decisions

After every moving decision, all group members submit a response to the question "would you like to end the formation process at the current network?". A group member can reply with either a YES or a NO. We refer to this type of decision as "ending decisions" because they have the potential to terminate the current round. We interpret

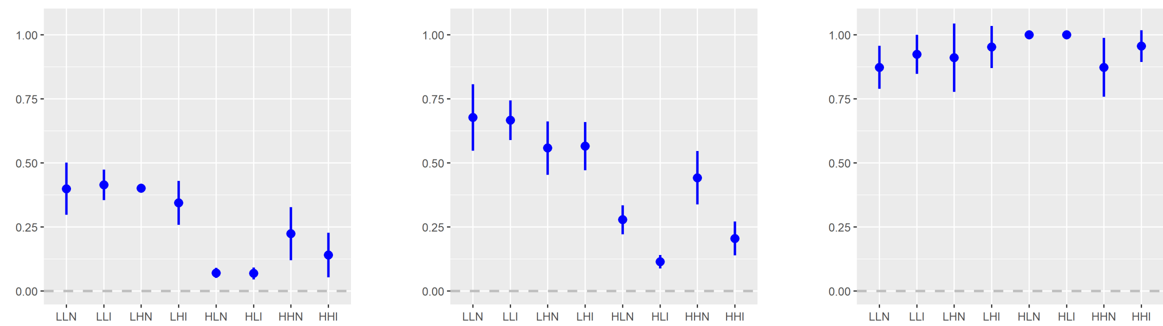
YES ending decisions as indicating willingness to make the current network final. We interpret an ending decision as reflecting *miscoordination* if it takes the value YES at networks that are Pareto dominated (i.e. all networks that belong to classes 1-9). This terminology alludes to the fact such decisions may lead to coordination failures (i.e. having a Pareto-dominated network as a final network). We interpret an ending decision as reflecting *cooperation* if it takes the value YES at the complete network (class 11). This is because refusing to make the complete network final is likely due to a desire to *compete* on obtaining 50.

This subsection examines the effect of scores, group composition and the provision of information on others' scores on rates of miscoordination and cooperation. In order to keep the analysis tractable, we define i as a high-score individual if her score is strictly above the median score (60) and as a low-score individual otherwise. Likewise, the average score among i's group members is considered high if it is above the median (60) and low otherwise. Together with the dummy variable indicating whether information on group members' scores is provided or not, this creates eight subsamples: LLN, LLI, LHN, LHI, HLN, HLI, HHN, and HHI, where the first letter signifies whether the considered individual is of low (L) or high (H) score, the second whether the average among her other group members is low (L) of high (H), and the third whether information is provided (I) or not (N).

Figure 3.6 presents rates of YES ending decisions for each of these subsamples. In panel 3.6a these rates are computed based on positions in classes 1-9. Low-score individuals tend to support ending the game in those classes significantly more than high-score individuals (difference in means: 0.25, p-value of associated F-test: 8.753e-08, number of observations: 3244). This suggests that high-score individuals are less prone to coordination failures.

Result 3.5. *High-score individuals are less prone to miscoordination.*

Some of the data used in panel 3.6a come from very unattractive positions (e.g. those of class 4), for which the variability in ending decisions is very low. To get a clearer view of the effect of information and others' scores on rates of coordination, panel 3.6b restricts attention to the most attractive position among classes 1-9, i.e. the one associated with class 8. Here, it is clear that among high-score individuals, the provision of information significantly reduces the probability of a YES ending decision (difference in means: 0.25; p-value of associated F-test: 0.0024, number of observations: 748). This may be taken to illustrate high-score individuals' responsiveness to information. Note,



(a) Classes 1-9 **(b)** Class 8 **(c)** Class 11

Figure 3.6: Rates of YES ending decisions. Vertical lines represent 95% confidence intervals. Standard errors are clustered at the group level.

however, that it seems as if it is the mere provision of information that has an effect on high-score individuals, an not the content of this information. A possible interpretation is that information provision pushes high-score individuals to think more strategically.

Result 3.6. *The provision of information on others' scores negatively impacts high-score individuals' tendency to miscoordinate. There is no evidence of such an effect for low-score individuals.*

Lastly, panel 3.6c presents the rates of YES ending decisions at class 11. All subsamples exhibit high rates of approval for this network. Ignoring subsamples HLN and HLI (where there is zero variation in ending decision due to the small sample size), no significant differences arise. This suggests that conditional on arrival to class 11, neither one's score, nor those of others, nor whether information is provided have any impact on rates of cooperation.

Result 3.7. *There is no evidence for a relationship between either one's own farsighted reasoning score or those of others on individuals' tendency to cooperate.*

3.5 Concluding Remarks

This paper reports the results of a lab experiment designed to test the predictions of various myopic and farsighted solution concepts in dynamic multi-player games, as well as the way these dynamics are affected by the composition of players in terms of their cognitive abilities and the disclosure of this information. We horse-race three categories of solution concepts: myopic; farsighted without preemptive moves; and farsighted with preemptive moves. To the best of our knowledge, this experiment is the first to explicitly consider, and test, the third category. The final outcomes groups converge to largely conform with the predictions of this third category, lending it strong empirical support. Solution concepts belonging to this category include the Set-Valued Rational Expectations Dekel (2023), the Equilibrium Stable Set Karos and Robles (2021) and the Subgame Perfect Consistent Set Granot and Hanany (2022).

Drawing from the literature relating cognitive ability to strategic behavior, our experimental design is comprised of an initial part aimed at measuring participants' cognitive abilities, and a second part where participants interact in a strategic environment. The cognitive task we deploy is an original questionnaire based on the Hit-15 game. We show that the performance in this task is predictive of behavior in the second part of the experiment, suggesting that it may serve as a useful tool in studying the relation of cognitive ability to strategic behavior.

While we do not find an effect of cognitive abilities, or their disclosure, on the final outcomes groups converge to, we do find an effect on individual behavior and speed of convergence. In terms of individual behavior, we find that participants with high cognitive ability tend to insist on achieving the outcomes predicted by the third category of solution concepts more vigorously than others. Since these outcomes are characterized by being Pareto efficient, this suggests that high-ability individuals are less prone to coordination failures. In addition, we find some (albeit limited) evidence that high-ability individuals' insistence on achieving Praeto efficiency is further amplified by the provision of information on others' abilities. In terms of the speed at which groups converge to a final outcome, we find that groups composed of only high-ability individuals tend to achieve convergence faster than groups composed of both low and high-ability individuals. This further illustrates the strategic competence of high-ability individuals.

3.6 Appendix: Interface

Figures 3.7 and 3.8 presents the interface of the network formation game in the "no information" treatment. Figure 3.7 shows the beginning of a stage, where one randomly selected group member group (in this case, the one labeled "C") can form\delete one of her links (or do nothing). The actual game takes place on the right part of the screen. Bold lines represent formed links and dashed lines represent absent ones. The payoffs associated with the current network appear next to the nodes. The simulation playground is on the left part of the screen. Participants are free to form\delete any link in that part of the screen at any time, and the payoffs associated with the displayed network always appear next to the nodes. A button to copy the structure of the current "actual network" appears on the top right of that part of the screen. This allows participants to quickly explore various paths starting from the current network.

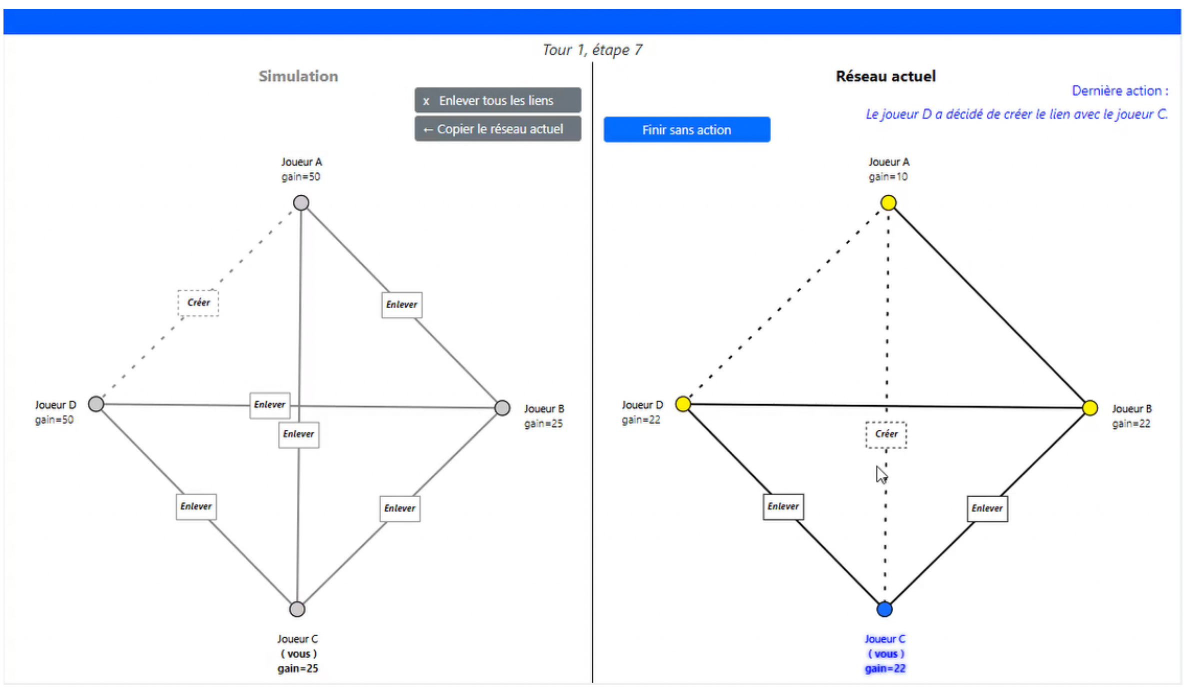


Figure 3.7

Figure 3.8 shows the end of a stage, all group members are asked whether they want to stop the formation process at the current network (which appears on the right part of the screen). Participants can respond either YES or NO.

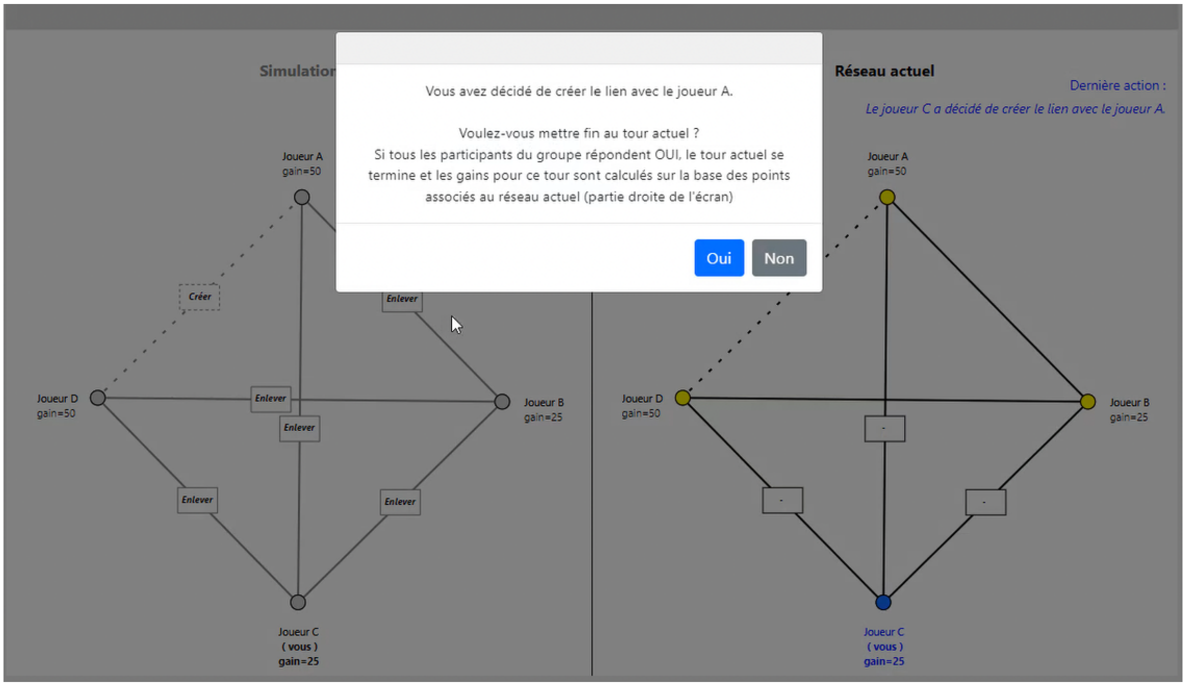


Figure 3.8

3.7 Appendix: Proofs

3.7.1 Proposition 3.1

Proof. Proof of (i). Since all moves away from any network $g \in g^3$ make the moving player strictly worse off, any such g belongs to the strong core. By Proposition 5 in Dekel (2023), a myopic SVRE exists and all states in the strong core are supported by any myopic SVRE. Hence, there exists a myopic SVRE for the game at hand, and any SVRE supports any $g \in g^3$. From $g \in g^1$ the only myopically rational move for any player is to form a link with one other player, which leads to some $g \in g^2$. From any $g \in g^2$, the only myopically rational move for players with degree 1 is not to change their linking status (which would result in staying in the same network), and the only myopically rational move for players with degree 0 is to form a link with the other isolated player, which would lead in some $g \in g^3$. Since any $g \in g^3$ is stationary under any myopic SVRE, no further moves are predicted to take place. Hence under a myopic SVRE m all the network in g^3, and only them, can be reached from g^1, i.e. $\bigcup_{m \in m^m} Y(g^1, m) = g^3$, where m^m is the set of myopic SVREs.

Proof of (ii). We first show that there exists a Dutta-Vohra SVRE m such that

$Y(g^1, m) = g^8$ and then show that no other Dutta-Vohra SVRE m exists. Consider a set-valued expectation m that prescribes exactly one player to move away from every network $g \notin g^8 \cup g^{11}$, exactly zero players to move away from every network $g \in g^8 \cup g^{11}$, only players with degree 3 to move away from any $g \in g^{10}$, and satisfies $Y(g, m) = g^8$ for every $g \notin g^{11}$. We argue that this m is Dutta-Vohra SVRE. It is essentially single-valued because it prescribes at most one move away from every state. It is absorbing because $Y(g, m) = g^8$ for every $g \notin g^{11}$ and $Y(g, m) = g^{11}$ for the unique $g \in g^{11}$. It is internally stable because any move away from $g \in g^{11}$ would terminate at $g \in g^8$, but the payoff from the latter is lower than the payoff for the former for all players, and any move from $g \in g^8$ would terminate at $g \in g^8$, which clearly does not strictly improve the payoff of any player. It is externally stable because the path prescribed from any non-stationary state terminates at a state that strictly improves the payoff of the moving player (note that moves away from $g \in g^{10}$ are prescribed only for players of degree 3). It is maximal because all feasible moves of all players prescribed to move at any non-stationary state terminate at the same stationary state (again, note that moves away from $g \in g^{10}$ are prescribed only for players of degree 3, so moves by players of degree 2 are not relevant for the maximality condition). Hence, there exists a Dutta-Vohra SVRE m such that $Y(g^1, m) = g^8$.

We now want to show that no other Dutta-Vohra SVRE m exists. Assume by contradiction the existence of a Dutta-Vohra SVRE m such that $Y(g^1, m) \neq g^8$. If $Y(g^1, m) =$ then m is not absorbing, which contradicts this assumption. If $Y(g^1, m) \cup g^{11} \neq \emptyset$ then m is not externally stable, because the player that makes the move from class 10 to class 11 (which must be of degree 2 in class 2) does not prefer the network on class 11 over the position she occupies in the class 10 network. Again, this contradicts our assumption. If $Y(g^1, m) \cup g^{11} = \emptyset$ then $g^8 \subseteq Y(g^1, m)$ for any Dutta-Vohra SVRE m, because otherwise the move away from the network in g^8 violates external stability. If $Y(g^1, m) \cup g^{11} = \emptyset$ and $g^8 \subseteq Y(g^1, m)$ then for every network adjacent to $g \in g^8$, m must prescribe some player to move to g, otherwise either internal stability (if this adjacent network is prescribed to be stationary) or maximality (if it is not) are violated. The same holds for all networks that are 2, 3 or 4 steps away from g^8, besides the one in g^{11}. Thus any Dutta-Vohra SVRE m satisfies $Y(g^1, m) = g^8$, and we have $\bigcup_{m \in m^{f_1}} Y(g^1, m) = g^8$, where m^{f_1} is the set of Dutta-Vohra SVREs.

Proof of (iii). By Proposition 5 in Dekel (2023), the game at hand prosesses a SVRE m satisfying $Y(z, m) = \{g\}$ for any weakly Pareto efficient network g. Since $g^{10} \cup g^{11}$ is the set of Pareto efficient networks, we know that all of these are supported as stationary.[13]

We now need to show that *only* those states can be supported as stationary. Suppose m is SVRE but does not support any network in $g^{10} \cup g^{11}$. Then, letting $g \in g^{11}$, $m' = [m \setminus m(g)] \cup \{(g, g, N)\}$ is a feasible and profitable deviation that does not violate dynamic consistency. Contradiction. Suppose m is SVRE but supports as stationary some network *in addition* so a network in $g^{10} \cup g^{11}$, i.e. $Y(m) \setminus [g^{10} \cup g^{11}] \neq \emptyset$. Let player i be one that gains either 40 or 50 at some of the stationary states (such a player must exist since $Y(m) \cap [g^{10} \cup g^{11}] \neq \emptyset$), and denote by g_i this players' least preferred stationary network. Note that there exists some g' such that $(g_i, g', \{i\}) \in M_i(g_i)$. For some such g' the deviation $m' = m \cup \{(g_i, g', \{i\})\}$ is feasible, profitable, and does not violate dynamic consistency. The profitability condition follows from Assumption 3.1: player i weakly prefers all other stationary states and strictly prefers some (in particular, the one where she gains 50 or 40). Contradiction. Thus, all, and only, states in $g^{10} \cup g^{11}$ may be stationary under some SVRE m, i.e. $\bigcup_{m \in m^{f_2}} Y(g^1, m) = g^{10} \cup g^{11}$, where m^{f_2} is the set of SVREs.

proof of (iv). Given *(iii)*, we just need to show that any SVRE m supporting as stationary some $g \in g^{10}$ is not symmetric and that some SVRE m supporting as stationary $g \in g^{11}$ is. Starting with the former, assume by contradiction that m is a symmetric SVRE but there exists $Y(m) \cap g^{10} \neq \emptyset$. This means that some players are prescribed not to move away from some $g \in g^{10}$ when occupying the position yielding a payoff of 25, and some are prescribed not to move away from that network when occupying the position yielding a payoff of 50. The symmetry condition requires this to be the case for *all* players, i.e. that whenever *any* player is in either of these two positions they do not move away from the current network. In turn, this means that all networks in g^{10} are stationary. Let i be a player obtaining a payoff of 25 in the stationary network $g \in g^{10}$. Note that there exists some g' such that $(g, g', \{i\}) \in M_i(g_i)$. For some such g' the deviation $m' = m \cup \{(g, g', \{i\})\}$ is feasible, profitable, and does not violate dynamic consistency. But this contradicts the assumption that m is SVRE. Hence, no $g \in g^{10}$ can be supported by a symmetric SVRE.

The proof of Proposition 5 in Dekel (2023) shows that an m that includes all feasible moves at all states and no moves away from some Pareto efficient state g is SVRE. Let this g be the unique network in g^{11}. Since all positions in g are symmetric, this m is symmetric. Hence, there exists an SVRE m supporting $g \in g^{11}$. This concludes the

[13] Dekel (2023) shows that this is obtained when m contains all feasible moves besides those that replace g by another state.

proof. □

3.8 Appendix: Instructions

3.8.1 General Instructions (for both information treatments)

Instructions Générales

Merci pour votre participation à cette expérience. Veuillez éteindre votre téléphone et le ranger. Il est interdit de communiquer avec les autres participants, sous peine d'exclusion de la session.

Si vous avez des questions de compréhension pendant la lecture des instructions, veuillez lever la main pour les poser. En cas de questions une fois la lecture des instructions terminée, veuillez lever la main et nous viendrons vous répondre individuellement.

Toutes les décision que vous prendrez aujourd’hui sont anonymes.

L'expérience est composée de deux parties. Vous recevrez les instructions pour la seconde partie une fois que vous et tous les autres participants aurez terminé la première partie.

Vous recevrez 5 euros pour votre participation. En outre, vous pouvez accumuler des gains dans les deux parties de l'expérience. Les gains de la première partie dépendent uniquement de vos décisions. Les gains de la seconde partie dépendent de vos décisions ainsi que de celles des autres participants. Tout au long de l'expérience, les gains seront affichés en termes de "points". A la fin de l'expérience, nous convertirons les points que vous avez gagnés en euros selon le taux :

10 points = 1 Euro

Vous serez payés individuellement à la fin de l'expérience.

1/1

3.8.2 Part I Instructions (for both information treatments)

Instructions Partie 1

Dans le cadre de cette partie de l'expérience, il vous est demandé de répondre à 10 questions concernant le jeu décrit ci-dessous. Vous gagnerez 10 points pour chaque bonne réponse et 0 point pour chaque mauvaise réponse (il y a toujours au moins une bonne réponse par question). Par conséquent, le nombre maximal de points que vous pouvez gagner dans cette partie de l'expérience est 10 × 10 = 100.

Description du jeu : A tour de rôle, deux joueurs mettent des jetons dans un "panier de jetons". A chaque tour, un joueur peut ajouter 1, 2 ou 3 jetons (ne pas ajouter de jetons n'est pas autorisé). **Le but de chaque joueur est d'être celui qui place le 15ème jeton dans le panier.**

Toutes les questions sont de la forme suivante :

‘ Il y a actuellement __ jetons dans le panier, et c’est votre tour.
Combien de jetons devriez-vous ajouter ? ’

Dans toutes ces questions, on vous demande d'imaginer que vous jouez contre un adversaire intelligent et expérimenté qui ne manque jamais une occasion de gagner. Nous insistons sur le fait que chaque question comporte au moins une réponse correcte.

Vous avez 18 minutes pour répondre à toutes les questions. Ce délai est amplement suffisant. Vous êtes encouragés à utiliser ce temps, ainsi que la feuille de brouillon fournie, pour trouver les bonnes réponses. Nous ne passerons à la partie 2 de l'expérience que lorsque tous les participants auront répondu à toutes les questions.

Veuillez appuyer sur ‘Suivant’ pour commencer la partie 1 de l'expérience.

1/1

3.8.3 Part II Instructions for “No Information” Treatment

Instructions Partie 2

Dans le cadre de cette partie de l'expérience, vous allez jouer à un jeu avec d'autres participants de cette session. Vous jouerez à ce jeu pendant quatre tours.

Groupes

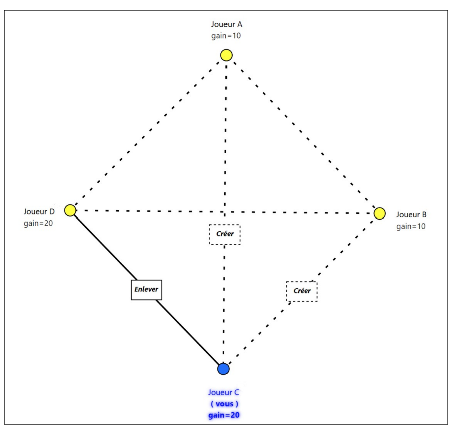
L'ordinateur vous assignera à un groupe avec trois autres participants. La composition des groupes ne change pas au cours de l'expérience. Par conséquent, vous serez dans le même groupe avec les mêmes personnes tout au long des quatre tours.

Le jeu est anonyme. Vous ne connaîtrez pas l'identité des autres personnes de votre groupe, ni pendant ni après l'expérience. De même, les autres participants de votre groupe ne connaîtront pas non plus votre identité.

Chaque participant du groupe se verra attribuer une lettre, A, B, C ou D. Sur votre écran d'ordinateur, à côté de l'icône qui vous représente, il sera marqué "VOUS" en plus de votre lettre d'identification (A, B, C ou D). Sur les écrans des autres membres de votre groupe, vous serez identifié uniquement par votre lettre (A, B, C ou D). Les lettres d'identification restent fixes tout au long des quatre tours.

Parcours de jeu

La tâche consiste à créer et à enlever des liens avec les autres membres du groupe. L'absence de lien est représentée par une ligne pointillée. L'existence d'un lien est représentée par une ligne continue.



Ceci est une capture d'écran de l'écran du joueur C.
Il apparaît donc en bas de l'écran et "vous" apparaît à côté de sa lettre d'identification.

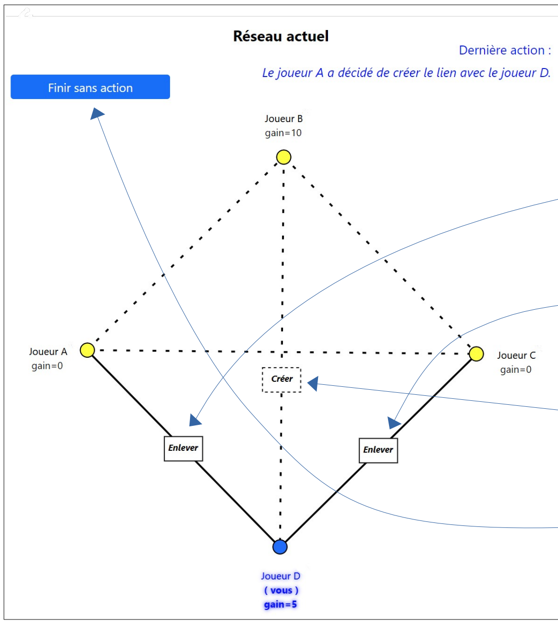
Il existe un lien entre C et D (ligne continue).
Les autres liens sont absents (lignes pointillées).

1/5

Chaque tour est divisé en plusieurs étapes. Lors de la première étape de chaque tour, il n'y aura aucun lien entre les membres de votre groupe.

L'ensemble des liens existants dans votre groupe à un moment donné se nomme "le réseau actuel".

À chaque étape, un membre du groupe (vous ou quelqu'un d'autre) sera sélectionné au hasard pour avoir la possibilité de changer le statut d'un lien de son choix, c'est-à-dire que le participant sélectionné peut soit ne rien faire, soit créer un nouveau lien, soit enlever un lien existant. Le participant sélectionné apparaîtra en bleu. Les autres membres du groupe seront invités à attendre que le participant sélectionné prenne sa décision.



Ici, le joueur D est celui qui a été sélectionné *pour jouer. Il apparaît donc en bleu. Il peut choisir une des actions suivantes:*

1. *Enlever son lien avec le joueur A (en cliquant sur le bouton "Enlever" correspondant)*
2. *Enlever son lien avec le joueur C (en cliquant sur le bouton "Enlever" correspondant)*
3. *Créer un lien avec le joueur B (en cliquant sur le bouton "Créer" correspondant)*
4. *Ne rien faire (en cliquant sur le bouton "Finir sans action")*

Une fois que le participant sélectionné a pris sa décision, tous les membres du groupe verront le nouveau réseau actuel à l'écran. Puis, il sera demandé à tous les membres du groupe s'ils veulent mettre fin au tour actuel. Vous pouvez répondre OUI ou NON.

- **Si tous les participants du groupe répondent OUI, le tour actuel se termine et les gains pour ce tour sont calculés sur la base des points associés au réseau actuel** (plus de détails ci-dessous).
- **Si au moins un membre de votre groupe répond NON, le groupe passe à l'étape suivante**, c'est-à-dire qu'un autre membre du groupe est sélectionné au hasard pour mettre à

2/5

jour le statut d'un de ses liens. Un participant qui a choisi "ne rien faire" dans une étape donnée ne sera pas sélectionné à nouveau à l'étape suivante.

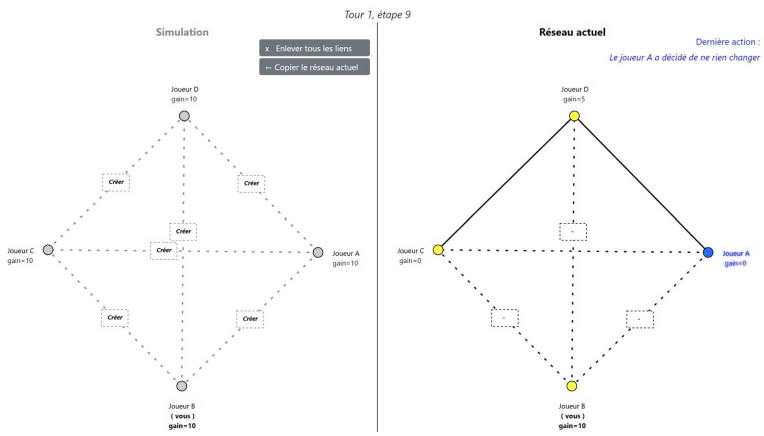
Si un consensus pour mettre fin au tour n'est pas atteint avant l'étape 25, le tour peut se terminer brusquement à n'importe quelle étape suivante avec une probabilité de 0,2. C'est-à-dire qu'aux étapes 26,27,28,... le tour se termine soit si tous les membres du groupe répondent OUI lorsqu'on leur demande s'ils veulent terminer le tour (comme d'habitude), **soit, le cas échéant, il y a une chance de 1/5 que le tour se termine néanmoins.**

Une fois qu'un tour se termine, le suivant commence, et les mêmes règles s'appliquent. Les actions prises lors d'un tour n'affectent pas les tours suivants.

<u>Gains</u>

Un nombre de points est associé à chaque participant de chaque réseau. Seuls les points qui vous sont associés dans le réseau final de chaque tour comptent pour vos gains. Ainsi, **les points qui vous sont associés ou qui sont associés à d'autres membres de votre groupe dans les réseaux de n'importe quel stade, sauf le dernier, n'ont aucune incidence sur vos gains**.

L'écran est divisé en une partie 'réseau actuel' (à droite) et une partie 'simulation' (à gauche). Les points associés au réseau actuel apparaissent toujours sur le côté droit de l'écran.



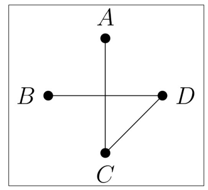
La partie simulation de l'écran est un outil conçu pour vous aider à prendre des décisions tout au long du jeu. Elle vous permet de simuler tous les réseaux possibles en cliquant sur les boutons 'créer'\'enlever' qui apparaissent sur chaque lien potentiel. Les gains associés au réseau simulé apparaissent toujours dans la partie gauche de l'écran. Notez que le bouton 'Copier le réseau actuel' en haut à droite de la région de simulation vous permet de copier immédiatement le réseau actuel dans la région de simulation et d'explorer les conséquences de diverses actions. Nous vous encourageons à utiliser cet outil avant de prendre des décisions concernant le réseau actuel.

3/5

Une autre façon de connaître les points associés à chaque réseau possible est d'utiliser la feuille de points ci-jointe.

Comment lire la feuille de points - Un exemple

Considérons le réseau suivant. D'après la feuille de points, quels sont les points obtenus par chaque participant ?



On observe que dans ce réseau il y a deux joueurs qui ont deux liens (C et D) et deux joueurs qui en ont un seul (A et B). Cela correspond à la Classe 7 dans la feuille de points. La section "exemple et points" sous la Classe 7 indique que les deux joueurs avec deux liens obtiennent 5 et les deux joueurs avec un lien obtiennent 0. D'où :

A obtient 0
B obtient 0
C obtient 5
D obtient 5

Vous êtes encouragés à étudier la feuille de points avant le début du jeu, ainsi qu'à la consulter pendant le jeu.

Calcul des gains

Deux tours sur quatre seront choisis au hasard pour calculer vos gains pour cette partie de l’expérience.

Remarques finales

C'est la fin des instructions. Il est important de s’assurer de leur bonne compréhension. Si vous avez des questions, veuillez lever la main. Pour vérifier que vous avez bien compris les instructions, nous vous demandons de répondre à quelques questions. Lorsque tout le monde aura répondu correctement à ces questions, le jeu commencera.

Veuillez appuyer sur ‘Suivant’ pour commencer la partie 2 de l'expérience.

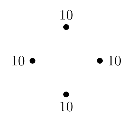
4/5

Feuille de Points

Classe 1

joueurs avec 0 liens: 4
joueurs avec 1 liens: 0
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

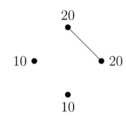
Exemple et points:



Classe 2

joueurs avec 0 liens: 2
joueurs avec 1 liens: 2
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

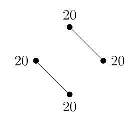
Exemple et points:



Classe 3

joueurs avec 0 liens: 0
joueurs avec 1 liens: 4
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

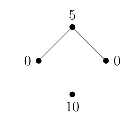
Exemple et points:



Classe 4

joueurs avec 0 liens: 1
joueurs avec 1 liens: 2
joueurs avec 2 liens: 1
joueurs avec 3 liens: 0

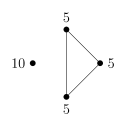
Exemple et points:



Classe 5

joueurs avec 0 liens: 1
joueurs avec 1 liens: 0
joueurs avec 2 liens: 3
joueurs avec 3 liens: 0

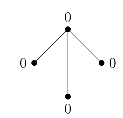
Exemple et points:



Classe 6

joueurs avec 0 liens: 0
joueurs avec 1 liens: 3
joueurs avec 2 liens: 0
joueurs avec 3 liens: 1

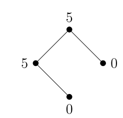
Exemple et points:



Classe 7

joueurs avec 0 liens: 0
joueurs avec 1 liens: 2
joueurs avec 2 liens: 2
joueurs avec 3 liens: 0

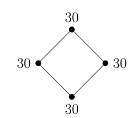
Exemple et points:



Classe 8

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 4
joueurs avec 3 liens: 0

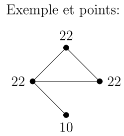
Exemple et points:



Classe 9

joueurs avec 0 liens: 0
joueurs avec 1 liens: 1
joueurs avec 2 liens: 2
joueurs avec 3 liens: 1

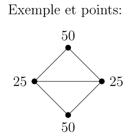
Exemple et points:



Classe 10

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 2
joueurs avec 3 liens: 2

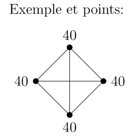
Exemple et points:



Classe 11

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 0
joueurs avec 3 liens: 4

Exemple et points:



5/5

3.8.4 Part II Instructions for "With Information" Treatment

Instructions Partie 2

Dans le cadre de cette partie de l'expérience, vous allez jouer à un jeu avec d'autres participants de cette session. Vous jouerez à ce jeu pendant quatre tours.

Groupes

L'ordinateur vous assignera à un groupe avec trois autres participants. La composition des groupes ne change pas au cours de l'expérience. Par conséquent, vous serez dans le même groupe avec les mêmes personnes tout au long des quatre tours.

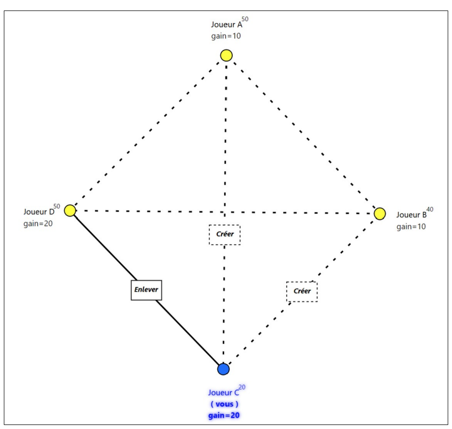
Le jeu est anonyme. Vous ne connaîtrez pas l'identité des autres personnes de votre groupe, ni pendant ni après l'expérience. De même, les autres participants de votre groupe ne connaîtront pas non plus votre identité.

Chaque participant du groupe se verra attribuer une lettre, A, B, C ou D. Sur votre écran d'ordinateur, à côté de l'icône qui vous représente, il sera marqué "VOUS" en plus de votre lettre d'identification (A, B, C ou D). Sur les écrans des autres membres de votre groupe, vous serez identifié uniquement par votre lettre (A, B, C ou D). Les lettres d'identification restent fixes tout au long des quatre tours.

Les gains de chaque membre de votre groupe **dans le cadre de la partie 1** (la tâche que vous venez de réaliser) apparaîtront au-dessus de leur lettre d'identification. Par exemple, B^{40} indique que le participant *B* a gagné 40 points **dans la partie 1 de cette expérience**. Pour référence, nous vous rappelons que le nombre maximal de points pouvant être gagnés dans cette tâche est 100, tandis que le minimum est 0.

Parcours de jeu

La tâche consiste à créer et à enlever des liens avec les autres membres du groupe. L'absence de lien est représentée par une ligne pointillée. L'existence d'un lien est représentée par une ligne continue.



Ceci est une capture d'écran de l'écran du joueur C.
Il apparaît donc en bas de l'écran et "vous" apparaît à côté de sa lettre d'identification.

Les gains des joueurs dans la partie 1 sont indiqués à côté de leurs lettres d'identification. Par exemple, B a gagné 40 points dans la partie 1.

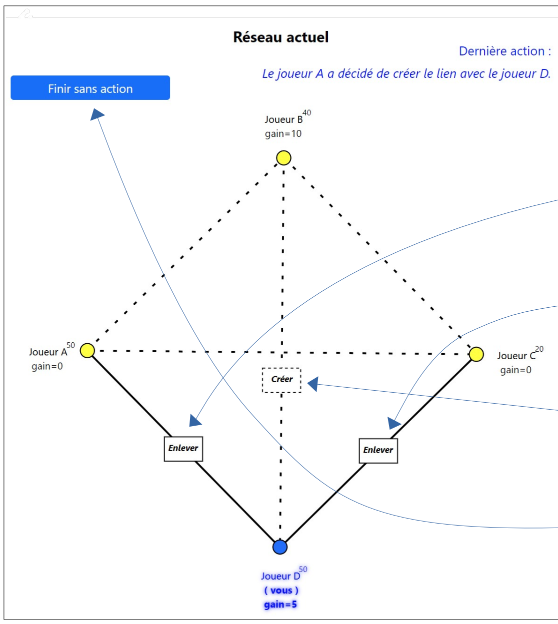
Il existe un lien entre C et D (ligne continue).
Les autres liens sont absents (lignes pointillées).

1/5

Chaque tour est divisé en plusieurs étapes. Lors de la première étape de chaque tour, il n'y aura aucun lien entre les membres de votre groupe.

L'ensemble des liens existants dans votre groupe à un moment donné se nomme "le réseau actuel".

À chaque étape, un membre du groupe (vous ou quelqu'un d'autre) sera sélectionné au hasard pour avoir la possibilité de changer le statut d'un lien de son choix, c'est-à-dire que le participant sélectionné peut soit ne rien faire, soit créer un nouveau lien, soit enlever un lien existant. Le participant sélectionné apparaîtra en bleu. Les autres membres du groupe seront invités à attendre que le participant sélectionné prenne sa décision.



Ici, le joueur D est celui qui a été sélectionné *pour jouer. Il apparaît donc en bleu. Il peut choisir une des actions suivantes:*

1. *Enlever son lien avec le joueur A (en cliquant sur le bouton "Enlever" correspondant)*
2. *Enlever son lien avec le joueur C (en cliquant sur le bouton "Enlever" correspondant)*
3. *Créer un lien avec le joueur B (en cliquant sur le bouton "Créer" correspondant)*
4. *Ne rien faire (en cliquant sur le bouton "Finir sans action")*

Une fois que le participant sélectionné a pris sa décision, tous les membres du groupe verront le nouveau réseau actuel à l'écran. Puis, il sera demandé à tous les membres du groupe s'ils veulent mettre fin au tour actuel. Vous pouvez répondre OUI ou NON.

- **Si tous les participants du groupe répondent OUI, le tour actuel se termine et les gains pour ce tour sont calculés sur la base des points associés au réseau actuel** (plus de détails ci-dessous).
- **Si au moins un membre de votre groupe répond NON, le groupe passe à l'étape suivante**, c'est-à-dire qu'un autre membre du groupe est sélectionné au hasard pour mettre à

2/5

jour le statut d'un de ses liens. Un participant qui a choisi "ne rien faire" dans une étape donnée ne sera pas sélectionné à nouveau à l'étape suivante.

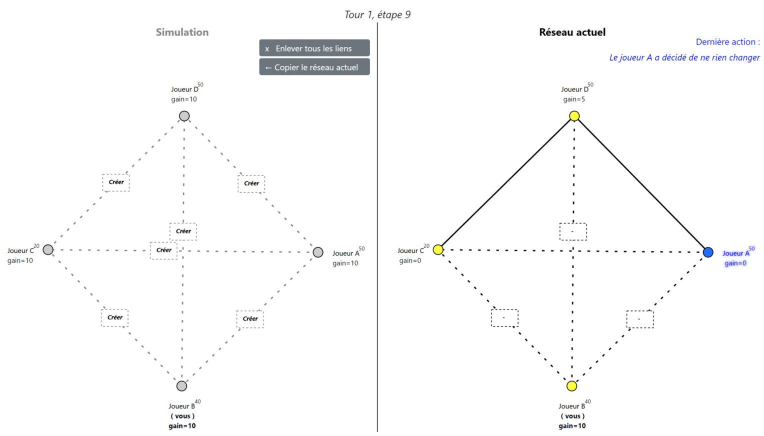
Si un consensus pour mettre fin au tour n'est pas atteint avant l'étape 25, le tour peut se terminer brusquement à n'importe quelle étape suivante avec une probabilité de 0,2. C'est-à-dire qu'aux étapes 26,27,28,... le tour se termine soit si tous les membres du groupe répondent OUI lorsqu'on leur demande s'ils veulent terminer le tour (comme d'habitude), **soit, le cas échéant, il y a une chance de 1/5 que le tour se termine néanmoins.**

Une fois qu'un tour se termine, le suivant commence, et les mêmes règles s'appliquent. Les actions prises lors d'un tour n'affectent pas les tours suivants.

<u>Gains</u>

Un nombre de points est associé à chaque participant de chaque réseau. Seuls les points qui vous sont associés dans le réseau final de chaque tour comptent pour vos gains. Ainsi, **les points qui vous sont associés ou qui sont associés à d'autres membres de votre groupe dans les réseaux de n'importe quel stade, sauf le dernier, n'ont aucune incidence sur vos gains**.

L'écran est divisé en une partie 'réseau actuel' (à droite) et une partie 'simulation' (à gauche). Les points associés au réseau actuel apparaissent toujours sur le côté droit de l'écran.



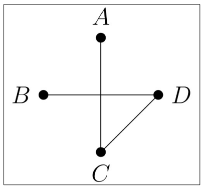
La partie simulation de l'écran est un outil conçu pour vous aider à prendre des décisions tout au long du jeu. Elle vous permet de simuler tous les réseaux possibles en cliquant sur les boutons 'créer'\'enlever' qui apparaissent sur chaque lien potentiel. Les gains associés au réseau simulé apparaissent toujours dans la partie gauche de l'écran. Notez que le bouton 'Copier le réseau actuel' en haut à droite de la région de simulation vous permet de copier immédiatement le réseau actuel dans la région de simulation et d'explorer les conséquences de diverses actions. Nous vous encourageons à utiliser cet outil avant de prendre des décisions concernant le réseau actuel.

3/5

Une autre façon de connaître les points associés à chaque réseau possible est d'utiliser la feuille de points ci-jointe.

Comment lire la feuille de points - Un exemple

Considérons le réseau suivant. D'après la feuille de points, quels sont les points obtenus par chaque participant ?



On observe que dans ce réseau il y a deux joueurs qui ont deux liens (C et D) et deux joueurs qui en ont un seul (A et B). Cela correspond à la Classe 7 dans la feuille de points. La section "exemple et points" sous la Classe 7 indique que les deux joueurs avec deux liens obtiennent 5 et les deux joueurs avec un lien obtiennent 0. D'où :

A obtient 0
B obtient 0
C obtient 5
D obtient 5

Vous êtes encouragés à étudier la feuille de points avant le début du jeu, ainsi qu'à la consulter pendant le jeu.

<u>Calcul des gains</u>

Deux tours sur quatre seront choisis au hasard pour calculer vos gains pour cette partie de l'expérience.

<u>Remarques finales</u>

C'est la fin des instructions. Il est important de s'assurer de leur bonne compréhension. Si vous avez des questions, veuillez lever la main. Pour vérifier que vous avez bien compris les instructions, nous vous demandons de répondre à quelques questions. Lorsque tout le monde aura répondu correctement à ces questions, le jeu commencera.

Veuillez appuyer sur 'Suivant' pour commencer la partie 2 de l'expérience.

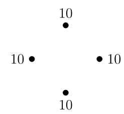
4/5

Feuille de Points

Classe 1

joueurs avec 0 liens: 4
joueurs avec 1 liens: 0
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

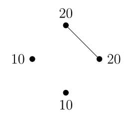
Exemple et points:



Classe 2

joueurs avec 0 liens: 2
joueurs avec 1 liens: 2
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

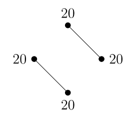
Exemple et points:



Classe 3

joueurs avec 0 liens: 0
joueurs avec 1 liens: 4
joueurs avec 2 liens: 0
joueurs avec 3 liens: 0

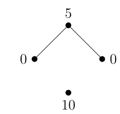
Exemple et points:



Classe 4

joueurs avec 0 liens: 1
joueurs avec 1 liens: 2
joueurs avec 2 liens: 1
joueurs avec 3 liens: 0

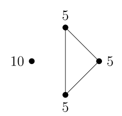
Exemple et points:



Classe 5

joueurs avec 0 liens: 1
joueurs avec 1 liens: 0
joueurs avec 2 liens: 3
joueurs avec 3 liens: 0

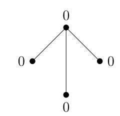
Exemple et points:



Classe 6

joueurs avec 0 liens: 0
joueurs avec 1 liens: 3
joueurs avec 2 liens: 0
joueurs avec 3 liens: 1

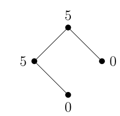
Exemple et points:



Classe 7

joueurs avec 0 liens: 0
joueurs avec 1 liens: 2
joueurs avec 2 liens: 2
joueurs avec 3 liens: 0

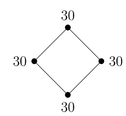
Exemple et points:



Classe 8

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 4
joueurs avec 3 liens: 0

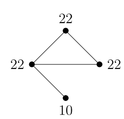
Exemple et points:



Classe 9

joueurs avec 0 liens: 0
joueurs avec 1 liens: 1
joueurs avec 2 liens: 2
joueurs avec 3 liens: 1

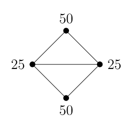
Exemple et points:



Classe 10

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 2
joueurs avec 3 liens: 2

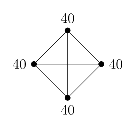
Exemple et points:



Classe 11

joueurs avec 0 liens: 0
joueurs avec 1 liens: 0
joueurs avec 2 liens: 0
joueurs avec 3 liens: 4

Exemple et points:



5/5

Bibliography

Advani, Arun and Bansi Malde, “Empirical Methods for Networks Data: Social Effects, Network Formation and Measurement Error,” *IFS Working Paper*, 2014, p. 91.

Aguirregabiria, Victor and Pedro Mira, “Sequential Estimation of Dynamic Discrete Games,” *Econometrica*, 2007, *75* (1), 1 – 53.

Alaoui, Larbi and Antonio Penta, “Endogenous depth of reasoning,” *The Review of Economic Studies*, 2016, *83* (4), 1297–1333.

Ambrus, Attila and Matt Elliott, “Investments in social ties, risk sharing and inequality,” *Review of Economics Studies (forthcoming)*, 2020.

_ , Markus Mobius, and Adam Szeidl, “Consumption risk-sharing in social networks,” *American Economic Review*, 2014, *104* (1), 149–82.

Badev, Anton, “Nash Equilibria on (Un)Stable Networks,” *Econometrica (forthcoming)*, 2020.

Bala, V and S Goyal, “A noncooperative model of network formation,” *Econometrica*, 2000, *68*, 1181–1230.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, “The diffusion of microfinance,” *Science*, 2013, *341* (6144), 1236498.

Barbera, Salvador, Walter Bossert, and Prasanta K Pattanaik, “Ranking sets of objects,” in “Handbook of utility theory,” Springer, 2004, pp. 893–977.

Benjamini, Yoav and Yosef Hochberg, “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing,” *Journal of the Royal Statistical Society*, 1995, *57* (1), 289–300.

Bjorn, Paul A and Quang H Vuong, “Simultaneous equations models for dummy endogenous variables: a game theoretic formulation with an application to labor force participation,” Technical Report, WP, California Institute of Technology 1984.

Bloch, Francis and Anne Van den Nouweland, "Expectation formation rules and the core of partition function games," *Games and Economic Behavior*, 2014, *88*, 339–353.

— **and Anne van den Nouweland**, "Farsighted stability with heterogeneous expectations," *Games and Economic Behavior*, 2020, *121*, 32–54.

— **, Garance Genicot, and Debraj Ray**, "Informal insurance in social networks," *Journal of Economic Theory*, 2008, *143* (1), 36–58.

Blumenstock, Joshua, "Don't forget people in the use of big data for development," *Nature*, 2018, *561*, 170–172.

Boucher, Vincent and Ismael Mourifie, "My friend far, far away: a random field approach to exponential random graph models," *The Econometrics Journal*, 2017, *20* (3), S14–S46.

Bramoullé, Y., H. Djebbari, and B. Fortin, "Peer Effects in Networks: a Survey," *Annual Review of Economics*, 2020, *12*, 603–629.

Bramoullé, Yann and Garance Genicot, "Diffusion and Targeting Centrality," *Unpublished*, 2023.

— **and Rachel Kranton**, "Risk-sharing networks," *Journal of Economic Behavior & Organization*, 2007, *64* (3-4), 275–294.

Brams, Steven J and Mehmet S Ismail, "Every normal-form game has a Pareto-optimal nonmyopic equilibrium," *Theory and Decision*, 2022, *92* (2), 349–362.

Brandes, Ulrik and Daniel Fleischer, "Centrality measures based on current flow," in "Annual symposium on theoretical aspects of computer science" Springer 2005, pp. 533–544.

Breschi, Stefano and Francesco Lissoni, "" Cross-firm" Inventors and Social Networks: Localized Knowledge Spillovers Revisited," *Annales d'Economie et de Statistique*, 2005, pp. 189–209.

Bresnahan, Timothy F and Peter C Reiss, "Empirical models of discrete games," *Journal of Econometrics*, 1991, *48* (1-2), 57–81.

Burks, Stephen V, Jeffrey P Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive skills affect economic preferences, strategic behavior, and job attachment," *Proceedings of the National Academy of Sciences*, 2009, *106* (19), 7745–7750.

Candelaria, Luis and Takuya Ura, "Identification and Inference of Network Formation Games with Misclassified Links," *working paper arXiv:1804.10118*, 2018.

Carrillo, Juan D and Arya Gaduh, "Dynamics and stability of social and economic networks: Experimental evidence," *Journal of Economic Behavior & Organization*, 2021, *188*, 1144–1176.

Chandrasekhar, Arun, "Econometrics of network formation," *The Oxford Handbook of the Economics of Networks*, 2016, pp. 303–357.

Chandrasekhar, Arun G and Matthew O Jackson, "A network formation model based on subgraphs," *Available at SSRN 2660381*, 2016.

Chandrasekhar, Arun G. and Randall Lewis, "Econometrics of sampled networks," *Unpublished*, 2012.

Charness, Gary and Matthew O Jackson, "Group play in games and the role of consent in network formation," *Journal of Economic Theory*, 2007, *136* (1), 417–445.

Chen, Daniel L, Martin Schonger, and Chris Wickens, "oTree—An open-source platform for laboratory, online, and field experiments," *Journal of Behavioral and Experimental Finance*, 2016, *9*, 88–97.

Chwe, Michael Suk-Young, "Farsighted coalitional stability," *Journal of Economic theory*, 1994, *63* (2), 299–325.

Coate, Stephen and Martin Ravallion, "Reciprocity without commitment: Characterization and performance of informal insurance arrangements," *Journal of development Economics*, 1993, *40* (1), 1–24.

Comola, Margherita and Marcel Fafchamps, "Testing Unilateral and Bilateral Link Formation," *Economic Journal*, September 2014, *124* (579), 954–976.

_ and _ , "The missing transfers: Estimating misreporting in dyadic data," *Economic Development and Cultural Change*, 2017, *65* (3), 549–582.

Dekel, Amit, "Set-valued Rational Expectation and Farsighted Stability," *Unpublished paper, Paris School of Economics.*, 2023.

Dutta, Bhaskar and Hannu Vartiainen, "Coalition formation and history dependence," *Theoretical Economics*, 2020, *15* (1), 159–197.

_ **and Rajiv Vohra**, "Rational expectations and farsighted stability," *Theoretical Economics*, 2017, *12* (3), 1191–1227.

Fafchamps, Marcel and Susan Lund, "Risk-sharing networks in rural Philippines," *Journal of development Economics*, 2003, *71* (2), 261–287.

Genicot, Garance and Debraj Ray, "Group formation in risk-sharing arrangements," *The Review of Economic Studies*, 2003, *70* (1), 87–113.

Gill, David and Victoria Prowse, "Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis," *Journal of Political Economy*, 2016, *124* (6), 1619–1676.

Gilleskie, Donna and Yun Zhang, "Friendship Formation and Smoking Initiation Among Teens," 2009. unpublished.

Goeree, Jacob K, Margaret A McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv, "The 1/d law of giving," *American Economic Journal: Microeconomics*, 2010, *2* (1), 183–203.

Goyal, Sanjeev, Marco Van der Leij, and José Luis Moraga-González, "Economics: An emerging small world," *Journal of Political Economy*, 2006, *114* (2), 403–412.

Graham, Bryan, "An econometric model of network formation with degree heterogeneity," *Econometrica*, 2017, *85(4)*, 1033 – 1063.

_ **and Andrin Pelican**, "Testing for Externalities in Network Formation Using Simulation," in "The Econometric Analysis of Network Data (B. Graham and A. de Paula, Eds.)" 2019, pp. 63 – 82.

Graham, Bryan S, "Methods of identification in social networks," *Annual Review of Economics*, 2015, *7* (1), 465–485.

Granot, Daniel and Eran Hanany, "Subgame perfect farsighted stability," *Unpublished paper, Faculty of Engineering, Tel Aviv University.[172]*, 2022.

Greiner, Ben, "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 2015, *1* (1), 114–125.

Harsanyi, John C, "An equilibrium-point interpretation of stable sets and a proposed alternative definition," *Management science*, 1974, *20* (11), 1472–1495.

Herings, P Jean-Jacques, Ana Mauleon, and Vincent Vannetelbosch, "Farsightedly stable networks," *Games and Economic Behavior*, 2009, *67* (2), 526–541.

Hoshino, Tadao, "Two-Step Estimation of Incomplete Information Social Interaction Models With Sample Selection," *Journal of Business & Economic Statistics*, 2019, *37* (4), 598–612.

Hsieh, Chih-Sheng and Lung Fei Lee, "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 2016, *31* (2), 301–319.

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

_ , Tomas Rodriguez-Barraquer, and Xu Tan, "Social capital and social quilts: Network patterns of favor exchange," *American Economic Review*, 2012, *102* (5), 1857–97.

Jones, Garett, "Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003," *Journal of Economic Behavior & Organization*, 2008, *68* (3-4), 489–497.

Jordan, James S, "Pillage and property," *Journal of economic theory*, 2006, *131* (1), 26–44.

Karos, Dominik and Laura Robles, "Full farsighted rationality," *Games and Economic Behavior*, 2021, *130*, 409–424.

Kimball, Miles S, "Farmers' cooperatives as behavior toward risk," *The American Economic Review*, 1988, *78* (1), 224–232.

Kimya, Mert, "Equilibrium coalitional behavior," *Theoretical Economics*, 2020, *15* (2), 669–714.

Kirchsteiger, Georg, Marco Mantovani, Ana Mauleon, and Vincent Vannetelbosch, "Limited farsightedness in network formation," *Journal of Economic Behavior & Organization*, 2016, *128*, 97–120.

Kocherlakota, Narayana R, "Implications of efficient risk sharing without commitment," *The Review of Economic Studies*, 1996, *63* (4), 595–609.

König, Michael D, "The formation of networks with local spillovers and limited observability," *Theoretical Economics*, 2016, *11* (3), 813–863.

König, Michael, Xiaodong Liu, and Yves Zenou, "R&D networks: Theory, empirics and policy implications," *Review of Economics and Statistics*, 2019, *101* (3), 476–491.

Konishi, Hideo and Debraj Ray, "Coalition formation as a dynamic process," *Journal of Economic theory*, 2003, *110* (1), 1–41.

Krishnan, Pramila and Emanuela Sciubba, "Links and architecture in village networks," *The Economic Journal*, 2009, *119* (537), 917–949.

Lambrecht, Marco, Eugenio Proto, Aldo Rustichini, and Andis Sofianos, "Intelligence disclosure and cooperation in repeated interactions," *CEPR Discussion Paper No. DP16656*, 2021.

Leung, Michael P., "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, 2015, *188* (1), 182–195.

Li, Qi and Jeffrey Scott Racine, *Nonparametric econometrics: theory and practice*, Princeton University Press, 2007.

Mele, Angelo, "A structural model of dense network formation," *Econometrica*, 2017, *85* (3), 825–850.

Miyauchi, Yuhei, "Structural estimation of pairwise stable networks with nonnegative externality," *Journal of Econometrics*, 2016, *195* (2), 224–235.

Myerson, R, "Game Theory: Analysis of Conflict Harvard Univ," *Press, Cambridge*, 1991.

Newey, Whitney K and Daniel McFadden, "Large sample estimation and hypothesis testing," *Handbook of econometrics*, 1994, *4*, 2111–2245.

Pattanaik, Prasanta K and Bezalel Peleg, "An axiomatic characterization of the lexicographic maximin extension of an ordering over a set to the power set," *Social Choice and Welfare*, 1984, *1* (2), 113–122.

Paula, Aureo De, "Econometrics of network models," in "Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress" Cambridge University Press Cambridge 2017, pp. 268–323.

_ **and X. Tang**, "Inference of signs of interactions effects in simultaneous games with incomplete information," *Econometrica*, 2012, *80* (1), 143–172.

_ **, Seth Richards-Shubik, and Elie T Tamer**, "Identifying Preferences in Networks With Bounded Degree," *Econometrica*, 2018, *86*, 263 – 288.

Poirier, Dale J, "Partial observability in bivariate probit models," *Journal of Econometrics*, 1980, *12* (2), 209–217.

Proto, Eugenio, Aldo Rustichini, and Andis Sofianos, "Intelligence, personality, and gains from cooperation in repeated interactions," *Journal of Political Economy*, 2019, *127* (3), 1351–1390.

_ **, _ , and _** , "Intelligence, Errors, and Cooperation in Repeated Interactions," *The Review of Economic Studies*, 2022, *89* (5), 2723–2767.

Rabinovich, Zinovi, Victor Naroditskiy, Enrico H Gerding, and Nicholas R Jennings, "Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types," *Artificial Intelligence*, 2013, *195*, 106–139.

Ray, Debraj, "Credible coalitions and the core," *International Journal of Game Theory*, 1989, *18*, 185–187.

_ **and Rajiv Vohra**, "The farsighted stable set," *Econometrica*, 2015, *83* (3), 977–1011.

_ **and** _ , "Maximality in the farsighted stable set," *Econometrica*, 2019, *87* (5), 1763–1779.

Ridder, Geert and Shuyang Sheng, "Estimation of large network formation games," *working paper arXiv:2001.03838*, 2020.

Rubinstein, Ariel, *Economic fables*, Open book publishers, 2012.

Sheng, Shuyang, "A structural econometric analysis of network formation games through subnetworks," *Econometrica*, 2020, *88* (5), 1829–1858.

Singh, Jasjit, "Collaborative networks as determinants of knowledge diffusion patterns," *Management science*, 2005, *51* (5), 756–770.

Stephenson, Karen and Marvin Zelen, "Rethinking centrality: Methods and examples," *Social networks*, 1989, *11* (1), 1–37.

Tamer, Elie, "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 2003, *70* (1), 147–165.

Teteryatnikova, Mariya and James Tremewan, "Myopic and farsighted stability in network formation games: an experimental study," *Economic Theory*, 2020, *69* (4), 987–1021.

Thirkettle, Matthew, "Identification and Estimation of Network Statistics with Missing Link Data," *working paper*, 2019.

Townsend, Robert M, "Risk and Insurance in Village India," *Econometrica*, May 1994, *62* (3), 539–591.

Udry, Christopher, "Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria," *The Review of Economic Studies*, 1994, *61* (3), 495–526.

Vandenbossche, Joost and Thomas Demuynck, "Network formation with heterogeneous agents and absolute friction," *Computational Economics*, 2013, *42* (1), 23–45.

Vosoughi, Soroush, Deb Roy, and Sinan Aral, "The spread of true and false news online," *Science*, 2018, *359* (6380), 1146–1151.

Vuong, Quang H, "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: Journal of the Econometric Society*, 1989, pp. 307–333.

Weerdt, Joachim De and Marcel Fafchamps, "Social identity and the formation of health insurance networks," *Journal of Development Studies*, 2011, *47* (8), 1152–1177.

_ **and Stefan Dercon**, "Risk-sharing networks and insurance against illness," *Journal of development Economics*, 2006, *81* (2), 337–356.