Bilateral Bargaining with Externalities[*]

by

Catherine C. de Fontenay *and* Joshua S. Gans
University of Melbourne

First Draft: 12th August, 2003
This Version: 30th October, 2007

This paper provides an analysis of a non-cooperative pairwise bargaining game between agents in a network. We establish that there exists an equilibrium that generates a coalitional bargaining division of the reduced surplus that arises as a result of externalities between agents. That is, we provide a non-cooperative justification for a cooperative division of a non-cooperative surplus. The resulting division is akin to the Myerson-Shapley value with properties that are particularly useful and tractable in applications. We demonstrate this by examining firm-worker negotiations and buyer-seller networks. *Journal of Economic Literature* Classification Number: C78.

Keywords. bargaining, Shapley value, Myerson value, networks, games in partition function form.

---

[*] We thank Roman Inderst, Stephen King, Roger Myerson, Ariel Pakes, Ariel Rubinstein, Michael Schwarz, Jeff Zwiebel, seminar participants at New York University, Rice University, the University of Auckland, the University of California-San Diego, the University of Sydney, the University of Southern California, the University of Toronto, the University of Washington (St Louis), Wharton, Yale University, participants at the Australian Conference of Economists, the Australasian Meetings of the Econometric Society, the International Industrial Organization Conference, and, especially, Anne van den Nouweland and three anonymous referees for helpful comments. Responsibility for all errors lies with the authors. Correspondence to: J.Gans@unimelb.edu.au. The latest version of this paper is available at www.mbs.edu/jgans.

1. Introduction

There are many areas of economics where market outcomes are best described by an on-going sequence of interrelated negotiations. When firms negotiate over employment conditions with individual workers, patent-holders negotiate with several potential licensors, and when competing firms negotiate with their suppliers over procurement contracts, a network of more or less bilateral relationships determines the allocation of resources. To date, however, most theoretical developments in non-cooperative bargaining have either focused on the outcomes of independent bilateral negotiations or on multilateral exchanges with a single key agent.

The goal of this paper is to consider the general problem of the outcomes that might arise when many agents bargain bilaterally with one another and where negotiation outcomes are interrelated and generate external effects. This is an environment where (1) surplus is not maximized because of the existence of those external effects and the lack of a multilateral mechanism to control them; and (2) distribution depends upon which agents can negotiate with each other. While cooperative game theory has developed to take into account (2) by considering payoff functions that depend on the precise position of agents in a graph of network relationships, it almost axiomatically rules out (1). In contrast, non-cooperative game theory embraces (1) but restricts the environment considered – symmetry, two players, small players, etc. – to avoid (2).

Here we consider the general problem of a set of agents who negotiate in pairs. All agents may be linked, or certain links may not be possible for other reasons (e.g., antitrust laws preventing horizontal arrangements among firms). Our environment is such that pairs of agents negotiate over variables that are jointly observable. This might be a joint action – such as whether trade takes place – or an individual action undertaken by one agent but observed by the other (e.g., effort or an investment). We specify a non-cooperative game whereby each pair of agents in a network bargains bilaterally in sequence. Pairwise negotiations utilize an alternating offer approach where offers and acceptances are made in anticipation of deals reached later in the sequence. Moreover, those negotiations take place with full knowledge of the network structure and how terms relate to that structure should it change. Specifically, the network may become 'smaller' should other pairs of agents fail to reach an agreement.

We consider a situation in which the precise agreement terms cannot be directly observed outside a pair. Thus, agents can observe the network of potential agreements but not the details of agreements they are not a party to. This is a reasonable assumption in a number of applied settings. In a labor market setting, this would be akin to firms observing the employment levels in rivals but not wages or hours. In a wholesale market, this is akin to rival suppliers observing competing product lines being sold downstream but not exact quantities or supply terms.

With some restrictions on beliefs, there is a unique equilibrium outcome of the incomplete information game. That outcome involves agents negotiating actions that maximize their joint surplus (as in Nash bargaining) taking all other actions as given. Hence, with externalities, outcomes are what might be termed "bilaterally efficient" rather than socially efficient.

The equilibrium set of transfers also gives rise to a precise structure; namely, a payoff that depends upon the weighted sum of values to particular coalitions of agents. This has a coalitional bargaining structure but with several important differences. First, the presence of externalities means that coalitions do not maximize their surplus, as equilibrium actions are bilaterally efficient rather than socially efficient. Second, coalitions may impose externalities on other coalitions; thus, the partition of the whole space is relevant. Thus, the equilibrium outcome is a Shapley allocation generalized to partition function spaces (as in Myerson 1977b) and further to networks in those partition spaces, but over a surplus that is characterized by bilateral rather than social efficiency.[1] Third, the restricted communication space may give rise to further inefficiencies, if certain agents are missing links between them and cannot negotiate, but instead choose individually optimal actions (see Jackson and Wolinsky, 1996).

In sum, we have a non-cooperative equilibrium that is a generalized Shapley division of a non-cooperative surplus, which is easy to use in applied settings. To our knowledge, no similar simple characterization exists in the literature for a multi-agent bargaining environment with externalities.

The usefulness of this solution to applied research seems clear. The seminal paper in the theory of the firm, Hart and Moore (1990), assumes that agents receive the Shapley value in negotiations; capturing the impact of substitutability without the extreme solutions of other concepts such as the core. However, there is an inherent discomfort to applying Shapley values in non-cooperative settings, because Shapley values assume that groups always agree to maximize their surplus, even in the presence of externalities. As a result, the theory of the firm has limited the types of strategic interactions that can be studied.[2]

---

[1] In the absence of externalities, it reduces to the Myerson value, and if, in addition, the network is complete, it reduces to the Shapley value.

[2] Stole and Zwiebel (1996) examined an environment where a firm bargains bilaterally with a given set of workers. While their treatment is for the most part axiomatic, focusing on a natural notion of stable agreements, they do posit an extensive form game for their environment. In their extensive form game, there is a fixed order in which each worker bargains with the firm over the wage for a unit of labor (i.e. there is no action space). Any given negotiation has the worker and firm taking turns in making offers to the other party that can be accepted or rejected. Rejected offers bring with them an infinitesimal probability of an irreversible breakdown where the worker leaves employment forever. Otherwise, a counter-offer is possible. If the worker and firm agree to a wage (in exchange for a unit of labour), the negotiations move on to the next worker. The twist is that agreements are not binding in the sense that, if there is a breakdown in any bilateral negotiation, this automatically triggers a replaying of the sequence of negotiations between the firm and each remaining worker. This new subgame takes place as if no previous wage agreements had been made (reflecting a key assumption in Stole and Zwiebel's axiomatic treatment that wage agreements are not binding and can be renegotiated by any party at any time).

Stole and Zwiebel (1996, Theorem 2) claim that this extensive form game gives rise to the Shapley value as the *unique* subgame perfect equilibrium outcome (something they also derive in their axiomatic treatment).

This game also allows us to contribute to the modeling of buyer-seller networks. Up until now, the papers addressing this issue have needed to restrict their attention to environments with a restrictive network structure, such as common agency, or to an environment with no competition in downstream markets.[3] Our solution combines the intuitiveness and computability of Shapley values with the consequences of externalities for efficiency. As such, it is capable of general application in these environments.[4]

The paper proceeds as follows. In the next section, we introduce our extensive form game. The equilibrium outcomes of that game are characterised in Sections 3 and 4; first with the equilibrium outcomes as they pertain to actions and then to distribution. Section 5 then considers particular economic applications including wage bargaining with competing employers and buyer seller networks. A final section concludes.

2. Bargaining Game

There are N agents and a graph, L (the network), of connections between them. Each linked pair, $ij \in L$, has associated with it a joint action, $x_{ij} \in X_{ij}$,[5] where X_{ij} is a compact interval of the reals. We normalize x_{ij} so that if a pair is not linked, $ij \notin L$, then $x_{ij} = 0$.[6]. Each

---

However, we demonstrate below that the informational structure between different bilateral negotiations must be more precisely specified (Stole and Zwiebel implicitly assume that the precise wage that is paid to a worker is not observed by other workers), and certain specific ‘out of equilibrium’ beliefs specified, for their result to hold. As will be apparent below, our extensive form bargaining game is a natural extension of theirs to more general economic environments.

[3] For example, Cremer and Riordan, 1987; Kranton and Minehart, 2001; Inderst and Wey, 2003; and Prat and Rustichini, 2003.

[4] There is also a literature on inefficiencies that arise in non-cooperative games with externalities (see, for example, Jehiel and Moldovanu, 1995). The structure of our non-cooperative game is of a form that eliminates these and we focus, in particular, on equilibria without such inefficiencies. As such, that literature can be seen as complementary to the model here.

[5] The action here is listed here as a scalar but could easily be considered to be a vector.

[6] For example, if x_{ij} is an action that is taken only by *i*, *i* chooses the optimal level for their own payoff, which we normalize to zero. The extension to action spaces in which the optimal level depends on the actions of others is trivial, as will be seen in Theorems 1 and 2: *i* will choose its best response, holding as given all other actions.

agent, *i*, has a payoff function, $u_i(\mathbf{x}) - \sum_{j\in N} t_{ij}$, where the first term is a utility function and t_{ij} is a transfer (positive or negative) from *i* to *j*. The utility function is a strictly concave and continuously differentiable function of the vector of all joint actions involving *i*; but we impose no structure on the utility to *i* of actions not involving *i* (that is, externalities) except to assume that $\sum_{i\in N} u_i(\mathbf{x})$ is globally concave in $\mathbf{x}$.

Fix an exogenous ordering of linked pairs.[7] When its turn in the order comes, each pair, *ij*, negotiates over (x_{ij}, t_{ij}). The pairwise bargaining game is described below. Importantly, it is assumed that, if there is an agreement in that game, only *i* and *j* can observe the agreed (x_{ij}, t_{ij}), however, it is assumed that breakdowns between pairs is common knowledge. As a breakdown will sever a pair's link, a new network state will arise (e.g., if *ij*'s negotiations break down, the new network is $K \equiv L - ij$). Formally, it is this network state that is common knowledge.

We follow Stole and Zwiebel (1996) and assume that agreements are *non-binding* with respect to a change in network state. Thus, in the event of a breakdown, any agreement between a pair still linked on the new network state can be *unilaterally re-opened*. In the model, we presume that the negotiation game is simply repeated for the new network state, because one party will always find it attractive to renegotiate. Critically, however, it is the anticipation of equilibrium outcomes in renegotiation subgames that plays a critical role in negotiated outcomes in the initial network state.

This modeling choice effectively assumes some contractual incompleteness with respect to a change in the network state.[8] An alternative approach would be to assume, following Inderst

---

[7] In the equilibrium we focus on, the precise ordering will not matter.

[8] Many contracts contain clauses that allow for renegotiation when a "material change in circumstances" arises.

and Wey (2003),[9] that initial negotiations are not just over $\left\{x_{ij}(L), t_{ij}(L)\right\}$ that would arise in the initial network state but also over each $\left\{x_{ij}(K), t_{ij}(K)\right\}$ for all possible network states, $K \subseteq L$, where $ij \in K$. That is, agreements would be *network contingent* and *binding*. It turns out that the equilibrium of interest that we analyze below arises in both the non-binding and binding cases. For expositional ease, we focus on the non-binding case and demonstrate the extension to the binding case in the appendix.

Bargaining for each pair proceeds according to the Binmore, Rubinstein and Wolinsky (1986) protocol. First, *i* or *j* are randomly selected to be the proposer and makes an offer, based on the current network state *K*, $\left\{x_{ij}(K), t_{ij}(K)\right\}$ which the receiver can accept or reject. Acceptance closes the negotiations and the next pairwise negotiation in the order begins. We assume that prior to any offer being made, there is an exogenous probability, $1-\sigma$, that negotiations between a pair ceases and no agreement can be made, otherwise the negotiation game between *i* and *j* begins again.[10] Thus, rejection may also trigger a breakdown in negotiations in which case this becomes common knowledge and, as past agreements are non-binding, a new order and round of negotiations between all pairs in $K-ij$ begins. We will focus on results where σ is arbitrarily close to 1.

---

[9] Inderst and Wey (2003) model multilateral negotiations as occurring simultaneously; any agent involved in more than one negotiation delegates one agent to bargain on their behalf in each negotiation. This alternative specification may be appropriate for situations where negotiations take place between firms. Agents could not observe the outcomes of negotiations they were not a party to. This would avoid the need to specify beliefs precisely in any equilibrium. As our model applies more generally than just between firms, we chose not to rely on a similar specification here. Note, also, that Inderst and Wey's treatment of individual negotiations is axiomatic rather than a full extensive-form, as they merely posit that agents split the surplus from negotiations in each different contingency. In an extensive form game, one would also have to model how and why pairs choose to negotiate over contingencies that are very unlikely to arise.

[10] Note that usually in such games, rejection triggers a breakdown possibility. Here, for technical rather than substantive reasons, we adopt a symmetric convention that exogenous breakdowns that sever future agreement possibilities between the pair can occur prior to an offer being made. This has the same impact as the alternative assumption but for the possibility of a breakdown prior to *any* offer being made.

The game concludes when all pairs on a given network state have reached an agreement or there are no linked pairs left. In this case, all agents received their agreed payments and choose their contracted actions (if any) with unlinked pairs choosing actions and transfers of 0.

Coalitional value and efficiency

As anticipated in the introduction, the equilibrium we focus on from this bargaining game gives rise to payoffs that reflect those found in coalitional game theory. For that reason, it is useful to provide additional notation to reflect coalitional value. For a given network, *K*, the resulting equilibrium set of actions, $\hat{\mathbf{x}}(K)$, leads to agent payoffs which sum to a 'coalitional value,' $\sum_{i\in N} u_i(\hat{\mathbf{x}}(K))$. When a subset of agents, $S \subseteq N$, are linked only to each other we will also consider the sub-coalition value, $\sum_{i\in S} u_i(\hat{\mathbf{x}}(K))$.

An important concept in this paper is bilateral efficiency, defined as follows:

Definition (Bilateral Efficiency). *For a given graph, K, a vector of actions,* $\hat{\mathbf{x}}(K)=\left(\hat{x}_{12}(K),\hat{x}_{13}(K),...,\hat{x}_{1n}(K),\hat{x}_{23}(K),...\hat{x}_{n-1,n}(K)\right)$ *satisfies bilateral efficiency if and only if:*

$$\hat{x}_{ij}(K)=\begin{cases}\arg\max_{x_{ij}} u_i\left(x_{ij},\hat{\mathbf{x}}(K)\setminus\hat{x}_{ij}(K)\right)+u_j\left(x_{ij},\hat{\mathbf{x}}(K)\setminus\hat{x}_{ij}(K)\right) & \text{if} \quad ij\in K \\ 0 & \phantom{\text{if}} \quad ij\notin K\end{cases}.$$

Under our concavity and continuity assumptions, $\hat{\mathbf{x}}(K)$ exists and is unique for every *K*. Consistent with this definition, we define $\hat{v}(S,K)\equiv\sum_{i\in S} u_i(\hat{\mathbf{x}}(K))$ as the coalitional value to a set *S* of players (linked only to each other) when actions are bilaterally efficient.[11] Note that the values $\hat{v}(.)$ are unique given our concavity assumptions on $u_i(.)$.

It is useful to note that, in some situations, bilateral efficiency will coincide with the efficient outcome normally presumed in coalitional game theory. Specifically, it is easy to see

---

[11] Note that there is a distinction between these coalitional values and those normally analyzed in coalitional game theory. In coalitional game theory, the sum of utilities in a coalition would describe a characteristic function where the actions were chosen to maximize coalitional value. Here, we define coalitional value with respect to an equilibrium set of actions arising from our non-cooperative bargaining game.

that if there were no externalities so that for each i, $u_i(\mathbf{x})$ was independent of x_{jk} for all $\{jk | i \neq j, i \neq k\}$, and a complete network, then maximizing pairwise utilities would result in a maximization of the sum of all utilities of agents linked in the network.

Feasibility

Depending on the nature of the externalities, and the structure of bargaining, an agent may be better off without one of their links, and therefore might unilaterally trigger a breakdown.[12] To make our analysis tractable, we need to restrict the underlying environment to rule this out, in any state of the network (N, L). Stole and Zwiebel (1996) term this *feasibility:*

Definition (Feasible Payoffs). *An equilibrium set of payoffs* $\left\{u_i(\hat{\mathbf{x}}(L)) - \sum_{k\in N} \hat{t}_{ik}(K)\right\}_{i\in N}$ *is* ***feasible*** *if and only if, for any* $K \subseteq L$ *and any* $ij \in K$, $u_i(\hat{\mathbf{x}}(K)) - \sum_{k\in N} \hat{t}_{ik}(K) \geq u_i(\hat{\mathbf{x}}(K-ij)) - \sum_{k\in N} \hat{t}_{ij}(K-ij)$.

In what follows, we simply assume that the primitives of the environment are such that feasibility is assured; after characterizing the equilibrium, we provide a simple sufficient condition for feasibility to hold.[13] However, for any given application, feasibility is something that would have to be confirmed in order to directly apply our equilibrium characterization below. If it did not hold, then our bargaining game will have an equilibrium where not all links would be maintained; resulting in interesting predictions in some environments.

Belief structure

Given that our proposed game involves incomplete information, the game potentially allows for many equilibrium outcomes. We need to impose some structure on 'out of

---

[12] For instance, as Maskin (2003) demonstrates, when an agent may be able to free ride upon the contributions and choices of other agents, that agent may have an incentive to force breakdowns in all their negotiations so as to avoid their own contribution. Maskin shows that this is the case for situations where there are positive externalities between groups of agents (as in the case of public goods).

[13] Although it is always satisfied in environments where there are no externalities.

equilibrium' beliefs that allows us to characterize a unique equilibrium for a given underlying environment. This is an issue that has drawn considerable attention in the contracting with externalities literature (McAfee and Schwartz, 1994; Rey and Vergé, 2004).

It is not our intention to revisit that literature here. Suffice it to say that the most common assumption made about what players believe about actions that they do not observe is the simple notion of "passive" beliefs. We will utilize it below. To define it, let $\{(\hat{x}_{ij}(K),\hat{t}_{ij}(K))_{K\subseteq L}\}_{ij\in L}$ be a set of equilibrium agreements between all negotiating pairs.

Definition (Passive Beliefs). *When i receives an offer from j of* $x_{ij}(K)\neq\hat{x}_{ij}(K)$ *or* $t_{ij}(K)\neq\hat{t}_{ij}(K)$, *i does not revise its beliefs regarding any other unobserved action in the game.*

At one level, this is a natural belief structure that mimics Nash equilibrium reasoning.[14] That is, if *i*'s beliefs are consistent with equilibrium outcomes – as they would be in a perfect Bayesian equilibrium – then under passive beliefs, it holds those beliefs constant off the equilibrium path. At another level, this is precisely why passive beliefs are not appealing from a game-theoretic standpoint. Specifically, if *i* receives an unexpected offer from an agent it *knows* to be rational, a restriction of passive beliefs is tantamount to assuming that *i* makes no inference from the unexpected action (e.g., by signaling). Nonetheless, as we demonstrate here, passive beliefs play an important role in generating tractable and interpretable results from our extensive form bargaining game; simplifying the interactions between different bilateral negotiations.

3. Equilibrium Outcomes: Actions

In exploring the outcomes of this non-cooperative bargaining game, it is useful to focus

---

[14] McAfee and Schwartz (1995, p.252) noted that: "one justification for passive beliefs is that each firm interprets a deviation by the supplier as a tremble and assumes trembles to be uncorrelated (say, because the supplier appoints a different agent to deal with each firm)." Similarly, the passive beliefs equilibrium in this paper is trembling hand perfect in the agent perfect form. A proof of this is available from the authors.

first on the equilibrium actions that emerge before turning to the transfers and ultimate payoffs. Of course, the equilibrium described is one in which actions and transfers are jointly determined. It is for expositional reasons that we focus on each in turn.

Theorem 1. *Suppose that agents hold passive beliefs, and that feasibility holds for each* $K \subseteq L$. *Given* (N, L), *as* $\sigma \to 1$, *any perfect Bayesian equilibrium involves each* $ij \in L$ *taking the bilaterally efficient actions,* $\hat{x}_{ij}(L)$.

This result says that actions are chosen to maximize pairwise utility holding those of others as given. It is easy to see that, in general, the outcome will not be efficient.[15]

The intuition behind the result is subtle. Consider a pair, *i* and *j*, negotiating in an environment where all other pairs have agreed to the equilibrium choices in any past negotiation, there is one more additional negotiation still to come and that negotiation involves *i* and another agent, *k*. Given the agreements already fixed in past negotiations, the final negotiation between *i* and *k* is simply a bilateral Binmore, Rubinstein, Wolinsky bargaining game. That game would ordinarily yield the Nash bargaining solution if *i* and *k* had symmetric information regarding the impact of their choices on their joint utility, $u_i(x_{ij}, x_{ik}, .) + u_k(x_{ij}, x_{ik}, .)$. This will be the case if *i* and *j* agree to the equilibrium $\hat{x}_{ij}$. However, if *i* and *j* agree to $x'_{ij} \neq \hat{x}_{ij}$, *i* and *k* will have different information. Specifically, while under passive beliefs, *k* will continue to base its offers and acceptance decisions on an assumption that $\hat{x}_{ij}$ has occurred, *i*'s offers and acceptances will be based on x'_{ij}. That is, *i* will make an offer, $(t'_{ik}, x'_{ik}(x'_{ij}))$, that maximizes $u_i(x'_{ij}, x_{ik}, .) - t_{ik}$ rather than $u_i(\hat{x}_{ij}, x_{ik}, .) - t_{ik}$ subject to *k* accepting that offer.

In this case, the question becomes: will *i* and *j* agree to some $x'_{ij} \neq \hat{x}_{ij}$? If they do, this will alter the equilibrium in subsequent negotiations. *i* will anticipate this, however, the assumption of

---

[15] As noted earlier, it will be efficient if there are no externalities and the network is complete. Consequently, this can be viewed as a generalisation of the efficiency results of Segal (1999, Proposition 3).

passive beliefs means that *j* will not. That is, even if they agreed to $x'_{ij} \neq \hat{x}_{ij}$, *j* would continue to believe that $\hat{x}_{ik}$ will occur. For this reason, *j* will continue to make offers consistent with the proposed equilibrium. On the other hand, *i* will make an offer, (t'_{ij}, x'_{ij}), that maximises $u_i(x_{ij}, x'_{ik}(x_{ij}),.) + t'_{ij}(x_{ij}) - t'_{ik}(x'_{ik}(x_{ij}))$ rather than $u_i(x_{ij}, \hat{x}_{ik},.) - t_{ij} - \hat{t}_{ik}$ subject to *j* accepting that offer. We demonstrate that this is equivalent to *i* choosing:

$$x'_{ij} \in \arg\max_{x_{ij}} u_i(x_{ij}, x'_{ik}(x_{ij}),.) + u_j(x_{ij}, \hat{x}_{ik},.) + u_k(\hat{x}_{ij}, x'_{ik}(x_{ij}),.)$$

which, by the envelope theorem applied to x'_{ik}, has $x'_{ij} = \hat{x}_{ij}$, the bilaterally efficient action. By a similar argument, agents do not find it worthwhile to deviate in a series of several negotiations.

4. Equilibrium Outcomes: Transfers and Payoffs

Turning now to consider equilibrium transfers and payoffs, we demonstrate here that while surplus is determined in a non-cooperative manner (from Theorem 1), under the same passive beliefs assumption, surplus division takes on a form attractively similar to coalitional bargaining outcomes. In particular, depending upon the nature of externalities and the network of bilateral negotiations, the division of whatever surplus is created gives agents a generalization of their Myerson value on that reduced surplus. As such, the division of surplus between players has an appealing coalitional structure even if the surplus is non-cooperatively determined.

It is useful first to consider the Myerson value and related concepts in coalitional game theory. Shapley's famous solution assumed that all agents were linked to each other. Myerson (1977a) generalised that notion by allowing for the possibility that cooperation may be restricted initially to an (exogenously given) graph (*N*, *L*), even before any coalitions have broken links with other players. Jackson and Wolinsky (1996) further extended the restrictions imposed by graphs by allowing the structure of the graph within a coalition itself (e.g., whether agents are linked directly or indirectly) to affect the payoff to a coalition; something that we have permitted here.

The Myerson value is somewhat restrictive in that it is not defined in situations where different groups of agents impose externalities upon one another. In another paper, Myerson (1977b) generalised the Shapley value to consider externalities by defining it for games in partition function form. In this paper, below we define a further generalization of the Myerson value to allow for a partition function space as well as a graph of potential communications (as in Navarro 2007). The "characteristic function" (i.e. the total payoff, *v*, to any given coalition of players) in such an environment depends on the structure of the entire graph, both intra- and inter-coalition.

In order to state the equilibrium payoffs, we need to introduce notation to express partitions of agents. $P=\{P_1,...,P_p\}$ is a partition of the set *N* if and only if (i) $\bigcup_{i=1}^{p} P_i = N$; (ii) $P_i \neq \varnothing$; and (iii) for all $j \neq k$, $P_j \cap P_k = \varnothing$. We define *p* as the cardinality of *P*. The set of all partitions of *N* is P^N. For a given network (*N*, *K*), we can now define a graph, (N, K^P), partitioned by, *P*. That is, $K^P = \{ jk \in K \,|\, \exists i \text{ s.t. } j,k \in P_i \}$. In other words, ($N$, K^P) is a graph partitioned by *P*.

We are now in a position to state our main result.

Theorem 2. *Suppose that agents hold passive beliefs, and that feasibility holds for each* $K \subseteq L$. *Given* (N,L), *as* $\sigma \to 1$, *there exists a unique perfect Bayesian equilibrium with each agent i receiving:*

$$u_i(\hat{\mathbf{x}}(L)) - \sum_{j\in N} \hat{t}_{ij} = \sum_{P\in P^N} \sum_{S\in P} (-1)^{|P|-1}(|P|-1)! \left[\frac{1}{n} - \sum_{\substack{i\notin S'\in P \\ S'\neq S}} \frac{1}{(|P|-1)(n-|S'|)} \right] \hat{v}(S, L^P) \,.$$

The right hand side is, in fact, a generalized Myerson value or Myerson value in partition function space defined on characteristic functions where agents take their bilaterally efficient actions. Thus, in equilibrium, we have a generalized Myerson value type division of a reduced surplus. That surplus is generated by a bilaterally efficient outcome in which each bilateral negotiation maximises the negotiators' own sum of utilities while ignoring the external impact of their choices on other negotiations (as in Theorem 1).[16]

As in Theorem 1, the proof relies upon the agents holding passive beliefs in equilibrium. Without passive beliefs, the equilibrium outcomes are more complex and do not reduce to this simple structure. That simplicity is, of course, the important outcome here. What we have is a bargaining solution that marries the simple linear structure of cooperative bargaining outcomes with easily determined actions based on bilateral efficiency. It is that simplicity that allows it to be of practical value in applied work.

Sufficient Condition for Feasibility

Now that we have derived the payoffs, we can provide sufficient conditions on the structure of externalities for feasibility to hold.

Theorem 3. *Suppose that* $v(N,L)$ *is such that for any set of agents, h, who are connected to each other by L but otherwise not connected to any agents in N/h,* $v(h,K) \geq v(h,K-ij)$, *for any* $K \subseteq L$. *Then the payoffs defined by Theorem 2 will be feasible.*

The proof is in the Appendix. The condition in the proposition implies that any negative externalities within a coalition are counterbalanced by benefits to being part of the coalition, but does not rule out the possibility that other coalitions might experience negative externalities resulting from the actions of the coalition

---

[16] It is easy to demonstrate that when there are no externalities (i.e., $u_i(\mathbf{x}(L))$ is independent of x_{kl} for any *k* and *l* not connected to *i*), this value is equivalent to the Myerson value and, in addition, if it is defined over a complete graph, it is equivalent to the Shapley value.

5. Applications

We now consider how our basic theorems apply in several of specific contexts where multi-agent bilateral bargaining has played an important role.

Stole and Zwiebel's Wage Bargaining Game

Stole and Zwiebel (1996; hereafter SZ) develop a model of wage bargaining between a number of workers and a single firm. The workers cannot negotiate with one another or as a group. Thus, the relevant network has an underlying 'star' graph with links between the firm and each individual worker. A key feature of SZ's model is that bargaining over wages is non-binding; that is, following the departure of any given worker (that is, a breakdown), either the firm or an individual worker can elect to renegotiate wage payments.

Nonetheless, what is significant here is that, when a firm cannot easily expand the set of workers it can employ ex post, there will be a wage bargaining outcome with workers and the firm receiving their Myerson values. This happens even if workers differ in their productivity, outside employment wages, and if work hours are variable. Moreover, if there were many firms, each of whom could bargain with any available worker ex post, each firm and each worker would receive their Myerson value over the broader network. As such, our results demonstrate that a Myerson value outcome can be employed in significantly more general environments than those considered by SZ.

It is instructive to expand on this latter point as it represents a significant generalisation of the SZ model and yields important insights into the nature of wage determination in labor markets. Suppose that there are two identical firms, 1 and 2, each of whom can employ workers from a common pool with a total size of n. All workers are identical with reservation wages normalized here to 0, and supply a unit of value. If, say, firm 1 employs n_1 of them, it produces profits of $F(n_1)$; where $F(.)$ is non-decreasing and concave. The firms only compete in the labor and not the product market.[17]

In this instance, as $F(n_1)$ does not depend on n_2 and vice versa, the actions agreed upon will maximize industry value, defined as $v(n) \equiv \max_{n_1} F(n_1) + F(n-n_1)$. By Theorem 2, each firm receives $\tilde{\pi}(n) = \frac{1}{(n+1)(n+2)}\left(\sum_{i=1}^{n}(i+1)\left(v(i)-2F(i)\right)+(n+2)\sum_{i=1}^{n}F(i)\right)$ and each worker receives $\tilde{w}(n) = \frac{1}{n}v(n) + \frac{2}{n(n+1)(n+2)}\sum_{i=0}^{n}\left((2i-n)F(i)-(i+1)v(i)\right)$.[18] It is straightforward to demonstrate that $\tilde{w}(n)$ is decreasing in n as in the SZ model.

It is interesting to examine the effect of firm competition in the labor market by considering wage outcomes when the two firms above merge.[19] In this case, the bargained wage, $\tilde{w}_M(n)$, becomes $\tilde{w}_M(n) = \frac{1}{n}v(n) - \frac{1}{n(n+1)}\sum_{i=0}^{n}v(i)$. One would normally expect that $\tilde{w}_M(n) < \tilde{w}(n)$ as there is a reduction in competition for workers with a merged firm; pushing wages down. However, this is only the case if:

$$\frac{1}{n(n+1)(n+2)}\sum_{i=0}^{n}(n-2i)\left(v(i)-2F(i)\right) > 0 \tag{1}$$

which does not always hold. For example, suppose that workers can work part time for each firm, then $v(i) = 2F(\frac{1}{2}i)$. In this case, the LHS of (1) becomes $\frac{2}{n(n+1)(n+2)}\sum_{i=0}^{n}(n-2i)\left(F(\frac{1}{2}i)-F(i)\right)$. The terms within the summation move from negative to positive and so if $F(i)-F(\frac{1}{2}i)$ is decreasing in i then the entire expression may be negative so

---

[17] It should be readily apparent that our model here will allow for competing, non-identical firms as well as a heterogeneous workforce.

[18] The complete derivation of these values can be provided by the authors on request.

[19] Stole and Zwiebel (1998) also considered a similar issue but with a small number of heterogeneous workers.

that $\tilde{w}_M(n) > \tilde{w}(n)$. Thus, the simple intuition may not hold.

The model reveals why workers may be able to appropriate more surplus facing a merged firm than two competing ones: if the production function is very flat between 0.5*n* and *n*, a worker has relatively poor outside options. Even if there is another firm to negotiate with, by moving their, their labor adds very little value there, and hence their wage is low.

General Buyer-Seller Networks

Perhaps the most important application of the model presented here is to the analysis of buyer-seller networks. These are networks where downstream firms purchase goods from upstream firms and engage in a series of supply agreements; the joint action between buyer and seller being the amount of input that will be supplied from the seller to the buyer.[20] Significantly, it is often assumed – for practical and antitrust reasons – that the buyers and sellers do not negotiate with others on the same side of the market. Hence, the analysis takes place on a graph with restricted communication and negotiation options.

In this literature, models essentially fall into two types. The first assumes that there are externalities between buyers (as might happen if they are firms competing in the same market) but that there is only a single seller (e.g., McAfee and Schwartz, 1994; Segal, 1999; de Fontenay and Gans, 2004). The second literature assumes that there are multiple buyers and sellers, but assumes that each buyer is in a separate market, so there is no competition in the final-good market (Cremer and Riordan, 1987; Kranton and Minehart, 2001; Inderst and Wey, 2003; Prat and Rustichini, 2003). Our environment here encompasses both of these model types – permitting externalities between buyers (and indeed sellers) as well as not restricting the

---

[20] The transfer payment can be thought of as a lump-sum payment or a per-unit payment. The two are equivalent if quantities are agreed-upon at the same time as price. (But this model excludes environments in which a per-unit price is negotiated, and the downstream firm subsequently orders quantities at that price.)

numbers or set of links between either side of the market. In so doing, we have demonstrated that when there are no spillovers between different agent pairs then industry profits are maximised. Thus, it provides a general statement of the broad conclusion of the buyer-seller network literature. Similarly, we have a fairly precise characterization of outcomes when there are externalities: firms will produce Cournot quantities, in the sense that the contracts of upstream firm *A* with downstream firm 1 will take the quantities sold by *A* to downstream firms 2,…*m* as given; and the quantities sold by *B* to downstream firms 1,…*m* as given.

Ultimately, the framework here allows one to characterize fully the equilibrium outcome in a buyer-seller network where buyers compete with one another in downstream market. The key advantage is that the cooperative structure of individual firm payoffs makes their computability relatively straightforward. For example, consider a situation with *m* identical downstream firms each of who can negotiate with two (possibly heterogeneous) suppliers, *A* and *B*. In this situation, applying Theorem 2, *A*'s payoff is:

$$\begin{aligned}\Upsilon_A = &\sum_{x=0}^{m}\binom{m}{x}\left(\sum_{i=0}^{x}\frac{(-1)^{x-i}}{m-i+2}\binom{x}{i}\right)\hat{v}_{A,B}(m-x)\\ &+\sum_{x_A=0}^{m}\sum_{x_B=0}^{m-x_A}\binom{m}{x_A}\binom{m-x_A}{x_B}\left(\begin{array}{l}\left(-\sum_{i=0}^{m-x_A-x_B}\frac{(-2)^{m-x_A-x_B-i}}{m-i+2}\binom{m-x_A-x_B}{i}\right)\left(\hat{v}_A\left(x_A\mid x_B\right)+\hat{v}_B\left(x_B\mid x_A\right)\right)\\ +\frac{(-1)^{m-x_A-x_B}}{m-x_B+1}\hat{v}_A\left(x_A\mid x_B\right)\end{array}\right)\end{aligned} \tag{2}$$

where $\hat{v}_{A,B}(m-x)$ is the bilaterally efficient (i.e., Cournot) surplus that can be achieved when both suppliers can both supply *m*–*x* downstream firms and $\hat{v}_A(x_A \mid x_B)$ is the bilaterally efficient surplus generated by *A* and x_A downstream firms when those x_A downstream firms can only be supplied by *A* and there are x_B downstream firms that can only be supplied by *B* (with no downstream firms able to be supplied by both). Thus, with knowledge of $\hat{v}_{A,B}(m-s)$, $\hat{v}_A(x_A \mid x_B)$

and $\hat{v}_B(x_A|x_B)$, using demand and cost assumptions to calculate Cournot outcomes, it is a relatively straightforward matter to compute each firms' payoff.

One implication of Theorem 2 is the parsimony of the structure: relatively few terms impact on the final payoff. These payoffs do not include the surplus created by industry environments in which some firms are linked to both upstream firms and some firms are linked to only one upstream firm, even though such environments are possible, and are considered by the players in their bargaining.

Significantly, this solution can be used to analyze the effects of changes in the network structure of a market. The linear structure makes comparisons relatively simple. For example, Kranton and Minehart (2001) explore the formation of links between buyers and sellers while de Fontenay and Gans (2005) explore changes in those links as a result of changes in the ownership of assets. The cooperative game structure of payoffs – in particular its linear structure – makes the analysis of changes relatively straightforward. It is also convenient for analyzing the effect of non-contractible investments (e.g., Inderst and Wey, 2003).

6. Conclusion and Future Directions

This paper has analyzed a non-cooperative bilateral bargaining game that involves agreements that may impose externalities on others. In so doing, we have demonstrated that the generation of overall surplus is likely to be inefficient, as a result of these externalities, but surplus division results in payoffs that are the weighted sums of surplus generated by different coalitions. As such, there exists an equilibrium bargaining outcome that involves a cooperative division of a non-cooperative surplus. This is both an intuitive outcome but also one that provides a tractable foundation for applied work involving interrelated bilateral exchanges.

7. Appendix

Proof of Theorem 1

As there is a probability that there is an exogenous breakdown between a pair prior to any offer being made by them, with $\sigma < 1$, there is a non-zero chance that no agreements will be reached. In a perfect Bayesian equilibrium, agents hold consistent beliefs along the equilibrium path; note that, every sub-network will appear on the equilibrium path, albeit, ultimately, with arbitrarily small probability. Because agents hold passive beliefs, when they observe a breakdown between other agents and consequently a new subgame, they assume that the breakdown was due to this $(1-\sigma)$-improbable event rather than due to a deviation from equilibrium, and play their equilibrium strategies in the subgame. The agents involved in the breakdown never play against each other again, and are forward-looking in their dealings with other agents. Hence behavior in each sub-network is independent of how that sub-network was reached.

We focus attention first on situations where there is an arbitrary σ and on offers that are made if there is an initial chance to make them. We note here that as σ goes to 1, the probability that an offer is not made falls to zero and the equilibrium outcome will reflect that.

Without loss in generality, therefore, let the current state of the network be *L*, and let $\{\hat{x}_{ij}(L),\hat{t}_{ij}(L)\}_{ij\in L}$ be the conjectured equilibrium outcome *and* also agents' passive beliefs regarding unobserved actions. We need only consider the incentives for one player, *i*, to deviate.

Suppose *i* is involved in *k* negotiations, and re-name the agents that *i* negotiates with as "1 to *k*". Suppose that *i* is considering deviating in the negotiation with *k*. If *i* makes the first offer, *i* solves the following problem:

$$\max_{x_{ik},t_{ik}} u_i(x_{ik},\hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\})-t_{ik}-\sum_{s\in N\setminus\{i,k\}}\hat{t}_{is}(L)$$

$$\text{subject to } u_k(x_{ik},\hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\})+t_{ik}-\sum_{s\in N\setminus\{i,k\}}\hat{t}_{sk}(L)\geq\sigma\Upsilon_k(L)+(1-\sigma)\Upsilon_k(L-ik)$$

Here, $\hat{\mathbf{x}}(L)$ is the vector of conjectured equilibrium actions, $\Upsilon_k(L)$ is *k*'s expectation of their payoff if it makes a counter-offer, and $\Upsilon_k(L-ik)$ is *k*'s equilibrium payoff if there is a breakdown in negotiations between *i* and *k* and a renegotiation subgame is triggered. As discussed above, both agents have consistent expectations about equilibrium actions and transfers in the sub-network; thus, in this negotiation, they both take $\Upsilon_k(L-ik)$ as given.

The incentive constraint reflects the passive beliefs of both players: Player *i* implicitly assumes that if *k* were to reject an offer and make a counter-offer, *k* would make the equilibrium counter-offer. And *i* assumes that *k* will not change behavior in subsequent negotiations (although such deviations will make the offer even more profitable for *k*). *k* believes that *i* has not deviated in prior negotiations, and that if this out-of-equilibrium offer is refused, *i* will still accept the equilibrium counter-offer.[21]

The transfer payment provides a degree of freedom that allows *i* to make the constraint bind; therefore:

$$t_{ik} = \sigma\Upsilon_k(L) + (1-\sigma)\Upsilon_k(L-ik) - \sum_{s\in N\setminus\{i,k\}} \hat{t}_{ik}(L) - u_k(x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\})$$

and *i* solves (omitting terms that do not depend on x_{ik}):

$$\max_{x_{ik}} u_i(x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\}) + u_k(x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\})$$

Hence, unless $\hat{x}_{ik}(L)$ is bilaterally efficient relative to all other equilibrium actions, a profitable deviation exists.

Now we consider what happens if *i* considers deviating in offers to both *k* and $k-1$. Let us assume, for simplicity, that *i* always gets to make the first offer, noting that, if this were not the case, as σ approaches 1, player *i* would simply reject any different offer.

Having concluded agreements with 1 through *k* – 2, *i*'s offers to $k-1$ and *k* solve:

$$\max_{x_{i,k-1},x_{ik},t_{i,k-1},t_{ik}} u_i(x_{i,k-1}, x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{i,k-1}(L), \hat{x}_{ik}(L)\}) - t_{i,k-1} - t_{ik} - \sum_{s\in N\setminus\{i,j,k\}} t_{is}(L)$$

subject to:

$$u_{k-1}(x_{i,k-1}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{i,k-1}(L)\}) + t_{i,k-1} - \sum_{s\in N\setminus\{i,k-1\}} t_{s,k-1}(L) \geq \sigma\Upsilon_{k-1}(L) + (1-\sigma)\Upsilon_{k-1}(L-(i,k-1)) \tag{3}$$

$$u_k(x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\}) + t_{ik} - \sum_{s\in N\setminus\{i,k\}} t_{sk}(L) \geq \sigma\Upsilon_k(L) + (1-\sigma)\Upsilon_k(L-ik) \tag{4}$$

Note that, because of passive beliefs, $k-1$ does not infer that a deviation will change *i*'s preferred x_{ik} offer to *k*; instead $k-1$ expects the equilibrium $\hat{x}_{ik}(L)$.

When the transfers $t_{i,k-1}$ and t_{ik} are chosen to make constraints (3) and (4) bind, the choice of $x_{i,k-1}$ and x_{ik} is equivalent to solving:

$$\max_{x_{i,k-1},x_{ik}} u_i(x_{i,k-1}, x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{i,k-1}(L), \hat{x}_{ik}(L)\}) + u_{k-1}(x_{i,k-1}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{i,k-1}(L)\}) + u_k(x_{ik}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ik}(L)\})$$

By iteration on agents from $k-2$ to 1, *i*'s optimal offers of actions $\{x_{ij}\}_{j=1}^{k} = \{x_{i1}, x_{i2}, \ldots x_{ik}\}$ to agents 1 to *k* will satisfy:

$$\max_{\{x_{ik}\}_{j=1}^{k}} u_i(\{x_{ij}\}_{j=1}^{k}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ij}(L)\}_{j=1}^{k}) + \sum_{j=1}^{k} u_j(x_{ij}, \hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ij}(L)\})$$

If there is an interior solution to this problem, it is clear that the first-order conditions for each action the first-order condition for bilateral efficiency and to this problem are the same.[22] Because of assumed compactness, continuity and differentiability, applying Milgrom and Segal

---

[21] *k* maintains these beliefs even if *i* refuses the equilibrium counter-offers for many rounds. Thus, there is no possibility of credible (costly) signalling.

[22] If each u_i is concave in its joint actions, as we have assumed, then this maximization problem is concave. The proof is available from the authors on request.

(2002, Corollary 4), the results hold even if the optimal value of x_{ij} is a corner solution.[23] Thus, we can conclude that the equilibrium values of $\hat{\mathbf{x}}(L)$ are bilaterally efficient, otherwise a profitable deviation exists.

Proof of Theorem 2

The proof of this theorem has two parts. First, we consider the set of conditions that characterize the unique coalitional bargaining allocation in a partition function environment when the communication structure is restricted to a graph. Second, we will demonstrate that the equilibrium of our non-cooperative bargaining game considered in Theorem 1 satisfies these conditions.

Part 1: Conditions Characterising the Generalized Myerson Value

Beginning with Myerson (1977a), a way of demonstrating a coalitional bargaining allocation was to state characteristics of that allocation that themselves determine that an allocation satisfying them was unique. Then one would demonstrate that a particular allocation satisfied those characteristics. Hence, it could be concluded that that allocation was the unique outcome of the coalitional bargaining game.

Myerson (1977a) used this approach and Jackson and Wolinsky (1996) extended it to demonstrate that the Myerson value was the outcome of a graph-restricted coalitional game. Stole and Zwiebel (1996) used this to prove Shapley value equivalence for their wage bargaining game. Myerson (1977b) defines a cooperative value for a game in partition function space but does not consider the possibility of a restricted communication structure nor does he provide a characterisation of that outcome based on conditions such as fair allocation and component balance.

Let $v(S, K^P)$ be the underlying coalitional value of a game in partition function form with total number of agents (S) and graph of communication (K). Here are some definitions important for the results that follow. Some definitions:

Definition (Connectedness). *Agents i and j are connected in network L if there exists a sequence of agents* $(i_1, i_2, ..., i_t)$ *such that* $i_1 = i$ *and* $i_t = j$ *and* $\{i_l, i_{l+1}\} \in L$ *for all* $l \in \{1, 2, ..., t-1\}$. *i is directly connected to j if* $ij \in L$.

Definition (Component). *A set of agents* $h \subset N$ *is a component of N in L if (i) all* $i \in h, j \in h, i \neq j$ *are connected in* (N,L)*; and (ii) for any* $i \in h, j \notin h$, *i and j are not connected. The set of all components of* (N,L) *is* $C(L)$.

Definition (Allocation Rule). *An allocation rule is a function that assigns a payoff vector,* $\mathbf{Y}(N, v, L) \in \mathrm{R}^N$, *for a given (N, v, L).*

---

[23] Applying the iterative process in the proof, if i's optimal offer was $x'_{i,k-1}$ taking into account an offer to k of x'_{ik}, the total derivative of the objective function with respect to $x_{i,k-1}$ is equal to the partial derivative of the objective function holding x_{ik} constant at x'_{ik}, even if that function is not differentiable in x_{ik}. Iterating back to 1, this version of the envelope theorem can accommodate optimal values of actions that may be corner solutions.

Definition (Component Balance). *An allocation rule,* $\mathbf{Y}$, *is component balanced if* $\sum_{i\in h}\Upsilon_i(N,v,L)=v(h,L)$ *for every* $h\in C(L)$, *where* $v(h,L)=\sum_{i\in h}u_i$.

Definition (Fair Allocation). *An allocation rule,* $\mathbf{Y}$, *is fair if* $\Upsilon_i(N,v,L)-\Upsilon_i(N,v,L-ij)=\Upsilon_j(N,v,L)-\Upsilon_j(N,v,L-ij)$ *for every* $ij\in L$.

The method of proof will be the following. First, Lemma 1 establishes that under component balance and fair allocation, there is a unique allocation rule. Second, we show that the generalized Myerson value satisfies fair allocation and component balance. Thus, using Lemma 1, this implies that the generalized Myerson value is the unique allocation rule for this type of cooperative game.

First, we note the following result from Navarro (2007):

Lemma 1 (Navarro, 2007). *For a given cooperative game* (*N, v, L*)*, under component balance and fair allocation, there exists a unique allocation rule.*

Next, in both an earlier version of this paper and the working paper version of Navarro (2007), there is a proof that the generalized Myerson value where agent *i* receives:

$$\Upsilon_i(N,L)=\sum_{P\in P^N}\sum_{S\in P}(-1)^{p-1}(p-1)!\left[\frac{1}{|N|}-\sum_{\substack{i\notin S'\in P\\ S'\neq S}}\frac{1}{(p-1)(|N|-|S'|)}\right]v(S,L^P)$$

satisfies component balance and fair allocation. Navarro (2007) states a theorem to that effect:

Lemma 2 (Navarro, 2007). *The generalized Myerson value for the game* (*N, v, L*) *in partition function form satisfies component balance and fair allocation.*

Part 2: The non-cooperative bargaining game satisfies these conditions

We want to show that the non-cooperative bargaining game satisfies fair allocation and component balance over a cooperative game with value function $\hat{v}(N,L)$ as determined by bilateral efficiency. Note that Theorem 1 demonstrates that the unique equilibrium of the bargaining game under passive beliefs involves achieving bilateral efficiency. This defines an imputed value function. We now want to show that, for this equilibrium, the two conditions are satisfied for the game with this value function.

When *i* and *j* negotiate, the current state of the network is *L*. When *i* and *j* bargain together, let t_{ij}^i be the transfer that *i* offers, which would give a payoff $\hat{v}_i^i$ and $\hat{v}_j^i$ to *i* and *j* respectively; *j*'s offer t_{ij}^j would, if accepted, lead to payoffs $\hat{v}_i^j$ and $\hat{v}_j^j$ respectively. Given that the transfers are chosen to make the incentive constraint bind, the offers satisfy:

$$\begin{aligned}\hat{v}_i^j&=\sigma\left(\tfrac{1}{2}\hat{v}_i^j+\tfrac{1}{2}\hat{v}_i^i\right)+(1-\sigma)\Upsilon_i(N,L-ij)\\ \hat{v}_j^i&=\sigma\left(\tfrac{1}{2}\hat{v}_j^j+\tfrac{1}{2}\hat{v}_j^i\right)+(1-\sigma)\Upsilon_j(N,L-ij)\end{aligned}\tag{5}$$

where $\Upsilon_i(N,L-ij)$ is the payoff to *i* after a breakdown with *j*. (Recall that if an offer is rejected, the order of offers is randomized again; so either *i* or *j* may make the next offer, with 0.5

probability each).

The payoff of a player, $\hat{v}_i$, is simply their utility from the actions taken plus equilibrium transfers $\hat{t}_{ki}$ from other players (which may be negative): $\hat{v}_i^i = \hat{u}_i^i - \hat{t}_{ij}^i + \sum_{k\in N\backslash j} \hat{t}_{ki}$ and $\hat{v}_i^j = \hat{u}_i^j - \hat{t}_{ij}^j + \sum_{k\in N\backslash j} \hat{t}_{ki}$ (where transfer t_{ki} is zero if *i* and *k* do not have a bargaining link). Also, the total amount that *i* and *j* have to divide is given by the other bargaining relationships: if $\hat{t}_{ki}$ is the equilibrium transfer from *k* to *i*:

$$\hat{v}_i^i + \hat{v}_j^i = \hat{v}_i^j + \hat{v}_j^j = \hat{u}_i + \sum_{k\in N\backslash\{i,j\}} \hat{t}_{ki} + \hat{u}_j + \sum_{k\in N\backslash\{i,j\}} \hat{t}_{kj} \qquad (6)$$

Using (5) to substitute out $\hat{v}_i^j$ and $\hat{v}_j^i$ in the first part of (6):

$$\hat{v}_i^i + \tfrac{\sigma}{2-\sigma}\hat{v}_j^j + \tfrac{2-2\sigma}{2-\sigma}\Upsilon_j(N, L-ij) = \tfrac{\sigma}{2-\sigma}\hat{v}_i^i + \tfrac{2-2\sigma}{2-\sigma}\Upsilon_i(N, L-ij) + \hat{v}_j^j$$
$$\Rightarrow \tfrac{2-2\sigma}{2-\sigma}\ \hat{v}_i^i + \tfrac{2-2\sigma}{2-\sigma}\Upsilon_j(N, L-ij) = \tfrac{2-2\sigma}{2-\sigma}\hat{v}_j^j + \tfrac{2-2\sigma}{2-\sigma}\Upsilon_i(N, L-ij)$$
$$\Rightarrow \hat{v}_i^i + \Upsilon_j(N, L-ij) = \hat{v}_j^j + \Upsilon_i(N, L-ij)$$

Note from (5) that in the limit, as σ tends towards 1, payoffs $\hat{v}_i^i$ and $\hat{v}_i^j$ become the same payoff $\hat{v}_i$, and therefore, $\hat{v}_i + \Upsilon_j(N, L-ij) = \hat{v}_j + \Upsilon_i(N, L-ij)$ which is the fair allocation condition.

Now consider condition (6) and its analogue for every bargaining link in the component that includes *i* and *j*. In the limit, as σ tends towards zero, the condition becomes $\hat{v}_i = u_i + \sum_{k\in N\backslash i} \hat{t}_{ki}$ for each *i*, where transfer t_{ij} is zero if *i* and *j* do not have a bargaining link. Therefore, for a given component, *h*:

$$\sum_{i\in h} \hat{v}_i = \sum_{i\in h}\left(u_i + \sum_{k\in N\backslash i} \hat{t}_{ki}\right) = \sum_{i\in h}\left(u_i + \sum_{k\in h\backslash i} \hat{t}_{ki}\right) = \sum_{i\in h} u_i$$

because there are no transfers to agents that you do not bargain with. The non-zero transfers in this summation term are all between agents in *h*, and, therefore, the summation includes both $\hat{t}_{ij}$ and ($-\hat{t}_{ij}$), which cancel out. This demonstrates component balance.

Binding Contingent Contracts

We now extend Theorems 1 and 2 by demonstrating that they apply for the game with binding contingent contracts. Let an arbitrary order of negotiations be fixed, and suppose the order of negotiations is known to all players. In the negotiation between *i* and *j*, *i* and *j* negotiate over all possible contingencies that may still occur.

The proof will show that the equilibrium actions and transfers – consistent with Theorems 1 and 2 – for the nonbinding contracts bargaining game also form a unique equilibrium of the contingent contract bargaining game.

Suppose that, when any *i* makes an offer to any *j*, their equilibrium offer is composed of:

- an offer of the bilaterally efficient actions $\hat{x}_{ij}(K)$ for each contingency *K* in which *i* and *j* are still linked;
- an offer of the transfers $\hat{t}_{ij}(K)$ that satisfy (5) for each contingency *K* in which *i* and *j* are still linked.

Suppose that *i* and *j* are the first pair to negotiate, in network *L*. They expect all other pairs to negotiate the agreements described above. Therefore, actions $\hat{x}_{ij}(L)$ and $\hat{t}_{ij}(L)$ satisfy (5), and hence, are the outcome of bilateral bargaining between *i* and *j*.

As σ approaches 1, *i* and *j* are indifferent as to the actions and transfers negotiated in other contingencies. Notice, however, that if *i* and *j* assign any positive probability to any contingency other than *L* (or a number of other contingencies), these contingent offers automatically satisfy conditions (5) and (6). Suppose, for instance, that they assign probability μ to one other contingency *K*. To satisfy the above conditions, *i*'s offer must satisfy

$$\max_{\substack{x_{ij}(K),t_{ij}(K)\\ x_{ij}(L),t_{ij}(L)}} \mu\left(u_i(x_{ij}(K),\hat{\mathbf{x}}(K)\setminus\{\hat{x}_{ij}(K)\}) - \sum_{k\in N\setminus\{i,j\}} \hat{t}_{ik}(K) - t_{ij}(K) \right)$$
$$+(1-\mu)\left(u_i(x_{ij}(L),\hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ij}(L)\}) - \sum_{k\in N\setminus\{i,j\}} \hat{t}_{ik}(L) - t_{ij}(K) \right)$$

$$\text{subject to} \quad \begin{array}{c} \mu\left(u_j(x_{ij}(K),\hat{\mathbf{x}}(K)\setminus\{\hat{x}_{ij}(K)\}) + t_{ij}(K) + \sum_{k\in N\setminus\{i,j\}} \hat{t}_{jk}(K) \right) \\ +(1-\mu)\left(u_j(x_{ij}(L),\hat{\mathbf{x}}(L)\setminus\{\hat{x}_{ij}(L)\}) + t_{ij}(L) + \sum_{k\in N\setminus\{i,j\}} \hat{t}_{jk}(L) \right) \end{array} \geq \begin{array}{c} \mu\left(\sigma\hat{v}_j^j(K) + (1-\sigma)\Upsilon_j(K-ij)\right) \\ +(1-\mu)\left(\sigma\hat{v}_j^j(L) + (1-\sigma)\Upsilon_j(L-ij)\right) \end{array}$$

Clearly the equilibrium offers from the non-binding game satisfy these conditions.

Would *i* and *j* have an incentive to deviate and negotiate different contingent contracts, in order to influence the other negotiations? For example, would *i* and *j* want to negotiate a contract that is very favorable to *i* in the event of a breakdown between *i* and *k*, in order to improve *i*'s bargaining power in negotiations with *k*? Given that we are assuming passive beliefs by all agents, a deviation by these two would not change *k*'s equilibrium beliefs, and, therefore, would not improve *i*'s bargaining position with *k*.

Now let us consider a negotiation that is further down in the line of negotiations. Suppose that *i* and *j* negotiate after players *a* and *b*. Then, if *a* and *b* have not had a breakdown in negotiations, *i* and *j* do not negotiate over the contingencies in which *a* and *b* have a breakdown, for instance, as that will clearly not occur. However, if *a* and *b* have indeed had a breakdown, they negotiate over those contingencies, and not over any contingencies in which *a* and *b* are still linked. From *a* and *b*'s point of view, therefore, *i* and *j* behave in the same way as in the non-binding contract game: they negotiate a contract in whatever contingency they find themselves in, not constrained by any earlier agreement. Therefore, they expect them to reach the agreements described in the non-binding contract game.

Proof of Theorem 3

The condition in the proposition is formally 'component superadditivity':

Definition (Component Superadditivity). *For a given set of agents, h, and graph K, suppose that* $v(h,K) \geq v(h,K-ij)$. *If this condition holds, for each component* $h \in C(K)$ *and each graph,* $K \subseteq L$, *then* $v(N,L)$ *satisfies component superadditivity.*

The strategy of the proof of this is to demonstrate first that we can map our game and its payoffs with externalities to one that is equivalent but is a coalitional game without externalities with Shapley values corresponding to the payoffs in the original game. We can then use that equivalence, to apply the results of Myerson (1977a) and show that no agent has an incentive to cause a breakdown in bilateral negotiations.

Owen (1986) shows that every coalitional game, Γ, that assigns a payoff $\gamma(T)$ to each subset *T* of *N* can be represented as a **unique** linear combination of unanimity games. A unanimity game u^S assigns the following payoffs:

$$u^S(T) = \begin{cases} 1 & \text{if} \quad S \subseteq T \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Given that there are 2^N possible subsets, this implies that there are 2^N parameters that define any particular game, Γ.

Shapley values satisfy three properties: linearity, symmetry, and no payments to "carriers," that is, players who add nothing to value. Those last two properties imply that the Shapley value payoff to the players of a unanimity game must be 1/*s* for each of the *s* members of *S*, and 0 for everyone else. So the payoff to players in a unanimity game is pinned down uniquely by the properties.

Linearity means the following: Suppose game Γ is a linear combination of two games, Γ_1 and Γ_2. Then the Shapley value of each player in game Γ, θ_i^Γ, has the same linear relation to the Shapley values of games Γ_1 and Γ_2: If, for any set *T*, $\gamma(T) = a\gamma_1(T) + b\gamma_2(T)$, then $\theta_i^\Gamma = a\theta_i^{\Gamma_1} + b\theta_i^{\Gamma_2}$. Linearity implies that if each game Γ is a unique linear combination of unanimity games, then its Shapley value is uniquely defined: If, for any *T*, $\gamma(T) = \sum_{S \subseteq N} a_S^\Gamma u^s(T)$ then $\theta_i^\Gamma = \sum_{S \subseteq N} a_S^\Gamma \theta_i^{u^S}$, for all *i*. So the Shapley value of agents in game Γ are uniquely defined by 2^N coefficients.

The next step is to take a given network state and determine whether the payoffs are feasible. If this is done for any arbitrary network state, the proof will carry over to all possible states. Consider the set of players *h* (including *i* and *j*) who form a component in the current form of the graph, *g* (where $ij \in g$); the players in *h* are not necessarily a component in the graph $g - ij$.

Now let's imagine two games with no externalities and no graph structure, Γ and Γ', which are defined **only** for those *k* players in *h*. The Shapley values for game Γ are called θ_i^Γ, and for game Γ' are called $\theta_i^{\Gamma'}$ (for each $i \in h$). Let v_Γ and $v_{\Gamma'}$ be the characteristic functions of the coalitional games Γ and Γ', respectively. No restrictions are imposed on the games except the following:

$v_{\Gamma} = v(h, g)$ (1 equation)

$v_{\Gamma'} = v(h, g - ij)$ (1 equation)

For every i, $\theta_i^{\Gamma} = \Upsilon_i(g)$ (N equations)

For every i, $\theta_i^{\Gamma'} = \Upsilon_i(g - ij)$ (N equations)

where $\Upsilon_i(g)$ is i's payoff in the original game.

Therefore there are N+1 constraints imposed on the values of each game and of its Shapley values. Any game Γ and its Shapley value are uniquely defined by 2^N parameters, the coefficients a_S^{Γ} that define the relationship to unanimity games. Therefore, so long as $N + 1 \le 2^N$, which always holds, there are enough degrees of freedom to write down games Γ and Γ' that satisfy these conditions.

Component superadditivity of the game with externalities (which holds by assumption) implies that $v(h, g) \ge v(h, g - ij)$, and, therefore, that the same holds for the game without externalities. Without specifying additional, notation, this condition also has to hold for an augmented game including possible sub-graphs. This requires another 2^N equations or inequalities to satisfy. We still have enough degrees of freedom so long as $2N + 2 + 2^N \le 2(2^N)$, which holds whenever $N \ge 3$. (When $N < 3$, the original game has no externalities by definition and so feasibility is not an issue.)

Finally, following the proof in Myerson (1977a), we define a final game $\Gamma'' = \Gamma - \Gamma'$. Since, by construction, Γ'' has value of 0 for any coalition that does not include i and j, and weakly positive value for any set that includes i and j, the Shapley value of i for game Γ'' is positive, by the representation of the Shapley value as an expected marginal contribution. Thus, $\theta_i^{\Gamma''} \ge 0$, for all i, and thus, by the linearity of Shapley values, $\theta_i^{\Gamma} \ge \theta_i^{\Gamma'}$, and therefore $\Upsilon_i(g) \ge \Upsilon_i(g - ij)$.

8. References

Binmore, K., A. Rubinstein and A. Wolinsky (1986), "The Nash Bargaining Solution in Economic Modelling," *RAND Journal of Economics*, 17, 176-188.

Cremer, J. and M. Riordan (1987), "On Governing Multilateral Transactions with Bilateral Contracts," *RAND Journal of Economics*, 18 (2), pp.436-451.

de Fontenay, C.C. and J.S. Gans (2004), "Can Vertical Integration by a Monopsonist Harm Consumer Welfare?" *International Journal of Industrial Organization*, 22 (3), pp.821-834.

de Fontenay, C.C. and J.S. Gans (2005), "Vertical Integration in the Presence of Upstream Competition," *RAND Journal of Economics*, 36 (3), pp.544-572.

Hart, O. and J. Moore (1990), "Property Rights and the Theory of the Firm," *Journal of Political Economy*, 98 (6), pp.1119-1158.

Inderst, R. and C. Wey (2003), "Bargaining, Mergers and Technology Choice in Bilaterally Oligopolistic Industries," *RAND Journal of Economics*, 34 (1), pp.1-19.

Jackson, M.O. and A. Wolinsky (1996), "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71 (1), pp.44-74.

Jehiel, P. and B. Moldovanu (1995), "Cyclical Delay in Infinite Horizon Bargaining with Externalities," *Review of Economic Studies*, 62 (4), pp.619-639.

Kranton, R.E. and D.F. Minehart (2001), "A Theory of Buyer-Seller Networks," *American Economic Review*, 91 (3), pp.485-508.

Maskin, E. (2003), "Bargaining, Coalitions and Externalities," Presidential Address to the Econometric Society, *Econometrica* (forthcoming).

McAfee, R.P. and M. Schwartz (1994), "Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity and Uniformity," *American Economic Review*, 84 (1), 210-230.

McAfee, R.P. and M. Schwartz (1995), "The Non-Existence of Pairwise-Proof Equilibrium," *Economic Letters*, 49, pp.251-259.

Milgrom, P.R. and I. Segal (2002), "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, 70 (2), pp.583-601.

Myerson, R. (1977a), "Graphs and Cooperation in Games," *Mathematics of Operations Research*, 2, pp.225-229.

Myerson, R. (1977b), "Values of Games in Partition Function Form," *International Journal of Game Theory*, 6 (1), pp.23-31.

Navarro, N. (2007), "Fair Allocation in Networks with Externalities," *Games and Economic Behavior*, 58, pp.354-364.

Prat, A. and A. Rustichini (2003), "Games Played Through Agents," *Econometrica*, 71 (4), pp.989-1026.

Rey, P. and T. Verge (2004), "Bilateral Control with Vertical Contracts," *RAND Journal of Economics*, 35 (4), pp.728-746.

Segal, I. (1999), "Contracting with Externalities," *Quarterly Journal of Economics*, 114 (2), pp.337-388.

Segal, I. and M.D. Whinston (2003), "Robust Predictions for Bilateral Contracting with Externalities," *Econometrica*, 71 (3), pp.757-791.

Stole, L. and J. Zwiebel (1996), "Intra-firm Bargaining under Non-binding Contracts," *Review of Economic Studies*, 63 (3), 375-410.

Stole, L. and J. Zwiebel (1998), "Mergers, Employee Hold-Up and the Scope of the Firm: An Intrafirm Bargaining Approach to Mergers," *mimeo.*, Stanford.