Why do lions get the lion's share? : a Hobbesian theory of agreements

Author(s): Esteban, Joan, 1945-;Sákovics, József
Date 2002
Type Working Paper
Series Estudios = Working papers / Instituto Juan March de Estudios e Investigaciones, Centro de Estudios Avanzados en Ciencias Sociales 2002/182
City: Madrid
Publisher: Centro de Estudios Avanzados en Ciencias Sociales

WHY DO LIONS GET THE LION'S SHARE?
A HOBBESIAN THEORY OF AGREEMENTS

Estudio/Working Paper 2002/182
December 2002

Joan Esteban and József Sákovics

Joan Esteban is Research Professor and Director at the *Institut d'Análisi Económica* of the CSIC and Associate Professor at the *Universidad Pompeu Fabra*, Barcelona. This paper is based on a seminar that he presented at the Center for Advanced Study in the Social Sciences, Juan March Institute, Madrid, on 30 October 2001, entitled "Why the Lions get the Lions Share?". József Sákovics is at the: Department of Economics, University of Edinburgh.

Abstract

We present a novel approach to N-person bargaining, based on the idea –inspired in Hobbes– that the agreement reached in a negotiation is determined by how the direct conflict resulting from disagreement would be resolved. The explicit modelling of the *conflict game* directly leads to the observation that the outcome of conflict is a function of the stakes. Thus, our basic building block is the *disagreement function*, which maps each set of feasible agreements into a disagreement point. Using this function and a weakening(!) of the Independence of Irrelevant Alternatives axiom, based on individual rationality, we reach a unique solution. This agreement may be construed as the limit of a sequence of partial agreements, each of which is reached as a function of the parties' relative *power*. We also give an alternative characterisation based on the mere possibility of renegotiation.

«The rich get the law passed by means of force and arms or get it accepted by fear to their might, aren't things this way? » Plato, *Republic*.

«What I am arguing here is that in order to explain the substantive content of social institutions and, therefore, completely explain institutional development and change, our theories must focus primarily on the strategic conflict itself and on the mechanisms by which this conflict is resolved», Knight (1992), p. 123.

1. Introduction[*]

The concept of bargaining power, while acknowledged to be crucial by all, has remained very elusive in the theory of agreements.[1] Practically the only accepted explanation for the "origin" of bargaining power in the asymmetric Nash solution --or in non-cooperative bargaining games, in general-- is based on time preference, as proposed by Rubinstein (1982).[2] In many situations, relative patience is indeed an appropriate characterisation of the parties' differing ability of influencing the agreement. Our contention, however, is that the quest for a complementary explanation should not be given up. As our motivating example, we do not believe that a lion gets a larger share of a carcass because he is more patient than a hyena.

Standard cooperative bargaining theory arrives at solutions in two steps. The first step consists in the reduction of a bargaining situation into the confines of a *bargaining problem*, defined by Nash (1950) as the set S of feasible utility allocations and the threat point d. The latter is meant to be the outcome of some (presumably) non-cooperative interaction that

[*] We are thankful to Salvador Barberà, Jordi Brandts, Yeon-Koo Che, Joe Harrington, Carmen Herrero, Marco Mariotti, Rich McLean, Clara Ponsatí, Debraj Ray and especially to Andreu Mas-Colell, as well as to seminar participants at Alicante, Barcelona Jocs, the Barcelona ESF Exploratory Workshop on Bargaining, CORE, the Kenilworth ESRC Game Theory Meeting, NYU, Rutgers and St. Andrews for most helpful discussions. The first draft was written while J. Esteban visited CREI at Universitat Pompeu Fabra. He also gratefully acknowledges financial support from Fundación Pedro Barrié de la Maza and research grant DGICYT PB96-0897.

[1] This issue was also raised by Svejnar (1986) and Pollak (1994), among others.

[2] Binmore (1998), in his recent work on social contract theory, takes Rubinstein's model to be the natural description of what he calls the Game of Life determining the power of the players.

follows disagreement.[3] Thus, different "disagreement games" –possibly reflecting altered distributions of power among the players– are summarised as different threat points. The second step –which has concentrated the efforts of bargaining theory proper so far– consists in selecting a solution to this simplified problem. The generalised Nash solution permits that the surplus be distributed with some bias in favour of some players, thus admitting the possibility of differential power between them. The important point, however, is that current bargaining theory establishes no link between the power of the players in the first step –determining *d*– and in the second step, when fixing the shares of the surplus over and above *d*.

Inspired in Thomas Hobbes' theory of the social contract,[4] our paper is an attempt at integrating the above two steps in a consistent manner. Specifically, we consider that there is no other source of differential power than the one underlying the non-cooperative game supporting the disagreement point. That is, our interpretation of *bargaining power* is that it coincides with *power* itself in the fully non-cooperative scenario. In order to model this, we need to incorporate more data from the bargaining situation into a (generalised) bargaining problem. Instead of limiting the transmitted information to the solution of the disagreement game, we incorporate into the description a reduced form of the disagreement game itself. As it turns out, all the relevant information can be summarised by the specification of how the outcome of conflict, *d,* varies as a function of the stakes, *S*. This is the game-specific *disagreement function*, *D(.)*, which maps sets of payoffs into the corresponding equilibrium of the disagreement game.[5] A generalised bargaining problem is thus completely described by a pair *(S, D)*.

The nature of disagreement games depends on the problem at hand. In some situations, the underlying disagreement game is so rudimentary that players do not even have

[3] Note that this "conflictual" resolution may take various forms: going to court, to strike, to call an arbitrator, to lobby, to cut prices, to waste time, to fight etc.

[4] See Section 2 for a concise description.

[5] Note that this additional information was already required for the determination of d in the standard context, since the knowledge of the game is necessary to find its equilibrium. Once the game is well defined, it is straightforward to calculate its equilibria under different hypotheses (S's).

a choice over alternative strategies. Consider, for instance, bargaining over the price of an object in the middle of a bazaar. If the players do not reach an agreement, the potential buyer walks out and goes to the next shop. However, such extremely simple situations seem the exception rather than the rule. In the previous case, it is essential that players terminate any future relationship after reaching disagreement. Whenever players do not cease to interact, the non-cooperative disagreement game is necessarily richer. Social relationships are of this type. The fact that we may fail to achieve a particular collective agreement simply means that the future relationship among agents will be non-cooperative. The same can be said of oligopolistic markets, industrial disputes, or simply of individuals litigating over a particular issue of their concern. This is also the case in the international arena, where one cannot modify who are one's neighbours. In all these important instances, players have a menu of potential strategies to follow under the non-cooperative mode.

This paper deals with bargaining situations characterised by truly strategic disagreement games. Our main claim is that for this class of bargaining situations there is a unique, and efficient, solution –the Hobbes solution– that can be completely characterised by means of the associated disagreement function.[6] One simple axiom –essentially positing individual rationality– permits the characterisation of a unique and efficient agreement. The axiom of the Independence of Individually Irrational Alternatives simply states that the agreement should not depend on the availability of alternatives that give to at least one player strictly less than what she would get in disagreement and hence are not individually rational. This axiom is in fact a weakening of Nash's axiom of Independence of Irrelevant Alternatives, though here it is applied in the context of a generalised bargaining game, that increases its bite.

The key observation driving our result is that once we eliminate the individually non-rational agreements, the bargaining problem becomes a different one –with a new bargaining set reduced to the remainder. Via the disagreement function, the new bargaining set yields a

[6] This is under the assumption that conflict does not exhaust all the surplus and, therefore, the payoffs to conflict always dominate the current status quo. This makes the threat of conflict always credible and pushes the players towards further concessions, until they finally reach an efficient allocation. Observe, that the efficiency of the agreement is derived, not assumed.

new threat point as well. Since our axiom applies to all bargaining problems, it also applies to this new (continuation) one, and further reduces the set of feasible agreements. What we show is that the repeated application of the axiom to the resulting sequence of bargaining games converges to a situation where the disagreement outcome is efficient, thus pinpointing a unique solution.

Our theory carries with it a conceptual novelty from the non-cooperative view as well. This insight relates to the interpretation of the terms: *agreement* and *disagreement*. Recall that the general idea of Rubinstein-type, offer-counteroffer models is that disagreement is temporary –in the sense that the rejection of an offer does not end the negotiation– and that agreement is total –in the sense that at each point in time the players are either in agreement or not, no intermediate possibility is considered. Instead, we make the "dual" assumption: we posit that disagreement is final but possibly partial, while agreements can be temporary, and therefore partial as well. That is, we allow for the possibility that the players agree on the sharing of part of the surplus and either postpone agreement or disagree on the rest. The important observation is that the fact that they did not get to full agreement is not interpreted as a complete failure of the negotiation: the partial agreement can be implemented and the extent (and the efficiency cost) of disagreement is reduced.

To appreciate the degree of the meta-similarity of the dual approaches, note that our enrichment of the bargaining problem with the disagreement function merely corresponds to the incorporation of an exogenous cost of disagreement –over the surplus remaining, conditional on any partial agreement. This is completely parallel to the case where the description of the bargaining problem is augmented with the exogenous parameters of the cost of (temporary) disagreement to each party –following any length of past disagreement. Similarly, our ruling out of a trivial disagreement game corresponds to Rubinstein's ruling out perfectly patient players. Finally, in both cases the sequential story behind the solution is not meant to be actually followed in real time. Rational, fully informed agents will immediately identify which is the unique solution.

To fix ideas, consider the simple example of splitting an inheritance of, say, ten pounds, between two siblings (who do not fancy each other). The siblings can either agree on

a particular split at no cost, or disagree and engage in a costly dispute over the money. Suppose that, if players engaged in conflict, in equilibrium seven pounds would be wasted (on, say, lawyers' fees), while of the remaining three pounds one player would expect to obtain two and the other one. This allocation may reflect the fact that, for instance, one's lawyer is "twice" as influential as the other's.[7] As a result of the expected outcome of conflict, any agreement must give to the siblings at least two and one pounds, respectively. Recognising this, they are willing to get to a partial agreement, which guarantees them these outside payoffs. Consequently, the effective area of dissent shrinks to the remaining seven pounds, which are precisely the benefits from cooperation. On the division of these seven pounds the siblings may again either agree or disagree and engage in a dispute. In the dispute, say, four pounds would be wasted and the strong sibling would obtain two and the weak one. Notice that even if they disagree, both siblings are better off by respecting their partial agreement and restricting the dispute to the distribution of the seven-pound surplus. It thus follows that any agreement must give to the siblings at least four and two pounds, respectively. This observation generates a new partial agreement. Applying the argument repeatedly, we reach the final agreement, where the ten pounds are distributed according to the power of the parties in the conflict game:[8] 20/3 and 10/3.

The argument above provides an attractive interpretation of the negotiation as a *process*[9] where, driven by the fear of a conflictual resolution, the parties accept to gradually narrow down the extent of their dissent.[10] Along each step of this process, it is the relative

[7] For example, the expected division ruled by the court may be 7:3, but the cost of the better lawyer is 5 while the cost of the worse one is only 2.

[8] In contrast, both the Nash (1950) and the Kalai-Smorodinsky (1975) solutions would predict that the seven-pound surplus over and above the (total) disagreement point would be brotherly shared by the two players. They would obtain 5.5 and 4.5 pounds in total, respectively.

[9] This process may be an actual one or just a thought process, which directly leads the players to agreement.

[10] Indeed, we observe that even in the cases in which players do not reach agreement and go into playing the conflict game –think of the extreme case of wars– they do accept restricting its amplitude. Thus, countries accept not to bomb civilian targets or to abstain from the use of particularly harmful weapons. In some cases –think of India and Pakistan– they tacitly agree to keep the conflict as mere border skirmishes. What keeps the conflict from escalation is the separation between the agreement and conflict games: not respecting a (partial) agreement is not a unilateral deviation in the conflict game; instead it is a unilateral deviation provoking

power of the players, as embodied in the disagreement function, that shapes the solution. We prove that for a very rich class of games, perfectly informed, rational agents will accept to reduce the area of their dissent completely: they will reach an agreement.

We also clarify the connection between the Hobbes, the generalised Nash and the Rubinstein solutions. Restricting attention to proportional disagreement functions (what corresponds to fixed discount factors in the Rubinstein-type models) and to the unit simplex as the Pareto frontier, we show that the Hobbes solution coincides with the asymmetric Nash solution, where the ratio of bargaining weights is equal to the proportion of the disagreement utilities.

Finally, we present the derivation of the Hobbes solution from the complementary, strategic perspective. Recall, that we consider a bargaining situation, where the interaction of the players does not stop suddenly in case they reach disagreement. This implies that when we try to implement our solution via a non-cooperative game, we need to incorporate this characteristic to the model. We have chosen to do this by invoking a certain degree of "collective rationality", allowing the players to *renegotiate* any inefficient outcome. In fact, we provide a non-cooperative characterisation of the Hobbes solution *exclusively* based on the possibility of renegotiation. That is, we provide a procedure-free theory of strategic bargaining. Such a theory is all the more desirable, since non-cooperative results are generally not robust to variations in the extensive form.

The paper is structured as follows. In the next section, we discuss the idea of reaching an agreement in the shadow of conflict from an historical perspective. In Section 3 we present our axiomatic analysis. Section 4 contains the alternative characterisation of the Hobbes solution. In the last section before we conclude, we elucidate our theory by contrasting it to the related bargaining literature.

---

a transition to the conflict game. This way such a deviation is observable: the countries foresee each other's reaction to a unilateral deviation. For example, according to our solution, in a complete information Cournot model, two identical firms would each agree to produce half the monopoly quantity, which is indeed the optimal colluding outcome (for them). The Nash equilibrium would correspond to unrestricted conflict (that is, competition) in this case.

2. Agreeing in the shadow of conflict

The view we develop here is to a good extent inspired by Hobbes' theory of social agreements. Well before Economics developed the theory of bargaining, Political Philosophy had addressed the question of social agreements in its inquiry about the foundations of the state. Thomas Hobbes (1588-1679) was possibly the first modern political philosopher who formulated an articulated theory of the *social contract*.[11] He viewed the possibility of a collective agreement as a case of "conditional cooperation" (in Taylor's, 1987, words), constrained by what individuals can obtain in the *state of nature*. The state of nature is the outcome that would ensue from a non-cooperative, rule-free interaction among utility maximising, selfish individuals (Hobbes' *first axiom*). The outcome of this interaction is resource consuming and is governed by the differences in endowments across individuals. His *second axiom* asserts that there exist agreements that Pareto dominate the allocation achieved under the state of nature. Finally, according to his *third axiom*, agreements should be conditioned by the allocation resulting in the state of nature: *«... it is a precept, or generall rule of Reason, That every man, ought to endeavour Peace, as farre forth as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre»* (Leviathan, 100, as cited in Taylor, 1987, 131). Therefore, in Hobbes' view, social agreements are not the outcome of an idealistic introspection on how things ought to be, but rather the viable outcome of a process conditioned by the might of the parties.

This view was largely shared by Jean-Jacques Rousseau (1782), for whom the formation of the political society and the establishment of laws *«gave new constraints to the weak and new forces to the rich, irreversibly destroyed natural freedom, established forever property law and inequality»* (p.170-1). Adam Smith (1776) also conceived the state as the creature of the mighty, specifically designed to give stability to the unequal distribution of wealth. In his own words: *«The rich, in particular, are necessarily interested to support that order of things, which can alone secure them in the possession of their own advantages. (...) Civil government, so far as it is instituted for the security of property, is, in reality, instituted*

[11] See Taylor (1987) and Gauthier (1990).

for the defence of the rich against the poor, of those who have some property against those who have none at all». (Book V, Chap. 1, Part II)

Despite the dramatically different normative positions as to what a "social contract" ought to be, all of them coincide in the positive analysis: in actual social agreements the mighty obtain a preferential treatment.[12] That actual social agreements will, at least partly, reflect the distribution of power is to be expected as long as a social contract is to be found acceptable by all parties. Therefore, and this is one of Hobbes' characteristic themes, we cannot develop a theory of social contracts without reference to the power of the parties in the non-cooperative scenario. The state of nature not only determines the size of the potential surplus to be shared, but also the shares themselves.

It is our opinion that standard bargaining theory has been driven to the use of normative axioms because the description of the bargaining problem was so stylised that there were no bases left for a positive derivation of the corresponding agreement.[13] We develop a positive theory of agreements, adopting Hobbes' position that takes the initial conditions as given and focuses on *reachable* social agreements, quite independently of the moral judgement they might deserve. We reserve normative considerations for the "state of nature," the initial conditions under which a particular agreement has been reached.[14] This view is consistent with Roemer's (1996) reservations about the moral content of a bargaining agreement obtained without a prior redistribution of the initial endowments.

[12] One of the lines along which the position held by Rousseau (and by Smith) departs from Hobbes' views is on whether the inequality that forces a biased social contract is innate to humans or is acquired.

[13] Svejnar (1986), Roemer (1988) and, more recently, Chen and Maskin (1999) have also expressed their reservations about the standard description of a bargaining problem, pointing out that Nash's abstraction might be dispensing with essential information.

[14] Consider the parallel case of assigning the gains from exchange. Economics takes a positive stand and investigates the terms of trade that will actually take place, resulting from different market structures and characteristics of the traders. It does not inquire about which would have been the "fair" terms of trade. The normative valuations are reserved for the comparison of the distribution of the characteristics that condition the trade (distribution of endowments, for instance).

We explore whether a solution can be characterized saving on axioms and making a more intensive use of the information contained in the description of the background game. This approach is in line with the growing literature on the explicit modeling of the conflictual resolution of opposing interests. The works by Becker (1983) on pressure groups and Tullock (1980) on rent-seeking, are the predecessors of the more recent papers by Esteban and Ray (1999), Grossman (1991, 1994), Grossman and Kim (1995), Hirshleifer (1991, 1995), Horowitz (1993), and Skaperdas (1992) among many others. The common feature of all these models is that the opposition of interests is resolved via conflict. [15] Players expend resources into trying to make their preferred option prevail. The equilibrium outcome entails waste of resources and the particular allocation reached critically depends on what is at stake as well as on the relative power, among other relevant characteristics, of the players.

In view of this literature, it seems natural to inquire why there is conflict to start with, could not there be a plausible conflict-avoiding agreement in this scenario? An agreement would save resources and, therefore, the crucial issue is how to share this surplus. However, potential agreements are not a central issue for most of these papers. On the other hand, the few who deal with it obtain agreements that are influenced by the power of the parties. This is the case, for instance, of the papers by Grossman (1994) and Horowitz (1993) on land reform. In Grossman (1994), landowners voluntarily give away land in order to decrease the probability of an expropriatory revolution and to save on protective expenditures. The size of the redistribution depends on the effectiveness of each party in rebelling or preventing it, as well as on the initial degree of inequality. Horowitz's (1993) approach is different. Landlords and peasants start from a status quo distribution and reach a sequence of interim agreements. At each stage, if they fail to reach a new interim agreement, either party can expropriate the other with some given probability (reflecting their relative power) or the status quo stays (again with some exogenously given probability). The economy follows a sequence of interim agreements converging to a steady state distribution that exactly reflects the power of

[15] Models of the conflictual resolution of opposing interests have also been developed in areas such us growth, international trade, industrial organization, organizational design, patent races, or economics of litigation, to mention just a few. Conflict models have also been developed for boundedly rational individuals (see, for example, Anderson et al., 1998).

the parties. Our theory of agreements is in accordance with the behavior predicated in this class of conflict models.

3. A disagreement theory of bargaining

In this section, we present our axiomatic analysis. We start by defining our generalised version of the bargaining problem, incorporating into it –via the disagreement function– a reduced form of the conflict game. Having done that, we will proceed to the characterisation of the Hobbes solution.

3.1. Bargaining in the shadow of disagreement

Suppose that there are *N* players, who wish to reach an agreement in $S^0 \in \Sigma$, where Σ is the set of compact subsets of the utility[16] space, $\Re_+^N$. Assume further the existence of a *disagreement function*, *D(.)*, which assigns a disagreement point, *d,* to every compact subset of S^0. That is, if the set of alternatives considered were *S*, the outcome of disagreement would be $d = D(S)$. This mapping is to be interpreted as shorthand for the solution[17] to an underlying *conflict game*. We would like to stress that *D(S)* may depend on additional parameters, especially those related to the players' "strength", which form part of the description of this conflict game. A bargaining problem in the shadow of conflict (BPSC) is then completely described by the pair $\left(S^0, D(.)\right)$. Let B denote the set of all BPSCs. A

[16] Actually, for our analysis it is not necessary that preferences satisfy the von Neumann-Morgenstern axioms. We could directly phrase our model in terms of money, prestige or the like. We elaborate on this issue in the Conclusions.

[17] This solution maybe a unique Nash (subgame-perfect?) equilibrium, but uniqueness of equilibrium is not necessary. In case of multiplicity, the "disagreement outcome" can be defined as the meet of the utilities gained at the different equilibria.

bargaining solution for BPSCs is then a mapping, $f: B \rightarrow \Sigma$, satisfying $f(S^0,D(.)) \subseteq S^0$. That is, the solution selects a subset of the alternatives as acceptable.

Note that, in principle, we need not impose any structure on $D(.)$, since it is meant to be a positive description of some real underlying conflict situation and therefore it cannot be freely chosen by the modeller. Nevertheless, to make the negotiation meaningful, we assume that disagreement can never be Pareto optimal: for all $S \in \Sigma$, $\exists s \in S$, such that $s \geq D(S)$, with strict inequality for some $i \in \{1, 2, ..., N\}$.

3.2. The Hobbes solution

We require the Hobbes solution to satisfy a single axiom, based on the fundamental concept of individual rationality. In the context of a bargaining game that requires consensus to reach agreement, individual rationality implies that any solution should weakly Pareto dominate the disagreement outcome, since otherwise at least one player would prefer to provoke disagreement. The complement of the set of individually rational alternatives is then known not to be "eligible" for an agreement, so it is natural to expect that the shape/extension of this set should not affect the solution. Indeed, this is the only assumption we make.

Let $S_x = \{s \in S \mid s \geq x\}$. That is, S_x is the subset of S which weakly Pareto dominates x. We impose the following axiom:

Independence of Individually Irrational Alternatives (IIIA):
$f(S,D(.)) = f(S_{D(S)},D(.))$ *for all* $(S,D(.)) \in B$.

That is, the axiom requires that eliminating the feasible agreements which do not (weakly) Pareto dominate the disagreement point should not change the solution. Conceptually, IIIA is much weaker than Nash's Independence of Irrelevant Alternatives (IIA) axiom, since it only eliminates a subset of his "irrelevant alternatives" and the definition of

this subset makes no reference to the final solution. Correspondingly, in a standard bargaining problem (SBP), IIIA would simply eliminate the alternatives that do not weakly dominate the disagreement point. However, when applied to a BPSC, IIIA has a recursive effect: once we eliminate the individually irrational alternatives, the application of the disagreement function to the remaining set results, in general, in a different disagreement point than before. To this new BPSC the axiom also applies (note that, if $(S,D(.)) \in B$ then $(S_{D(S)},D(.)) \in B$ as well). Thus, as long as $D(.)$ is not constant (as in a SBP), the application of IIIA generates new BPSCs which, in turn, also have to satisfy the axiom. In view of all this, should we still find IIIA a plausible axiom? We certainly think so. The point of all "irrelevant alternatives" type axioms is to provide some consistency between solutions of the same underlying bargaining situation but with different sets of available agreements. In our view, the appropriate description of the bargaining situation should not be confined to a fixed disagreement point, since the outcome of disagreement is likely to depend on the alternatives available. Therefore, what should be kept fixed when carrying out the "consistency check" is the disagreement *function*, just as it is done in IIIA. That is, our assumption compares bargaining situations where the same set of players are bargaining in the shadow of the same conflict game but with different sets of feasible utility pay-offs.

We do not want to impose any further restrictions on our solution:

Definition 1 *The Hobbes solution assigns to each BPSC the maximal set that is consistent with IIIA.*

Let us look at the implications of this –implicit– definition. Let $f^H(.,.)$ be a bargaining solution, $S^0 \in \Sigma$ an arbitrary bargaining set and $D(.)$ a disagreement function. IIIA implies that $f^H(S^0,D) = f^H(S^0_{d^0},D)$, where $d^0 = D(S^0)$. The disagreement point corresponding to the set $S^0_{d^0}$, however, is not d^0 but it is given by $d^1 = D(S^0_{d^0})$. Thus, the application of IIIA results in a new set, S^1. Repeatedly eliminating the individually irrational alternatives, for the t-th iteration we will have

$$S^t = \{u \in S^{t-1} \mid u \geq d^{t-1}\}.$$

IIIA requires exactly that for all the sets of this sequence, when coupled with *D(.),* the solution be the same. In other words, a bargaining solution satisfies IIIA if and only if $f^H(S^0,D) \subseteq S^* = lim_{T\to\infty} \bigcap_{t=0}^{T} S^t$.[18] Thus the Hobbes solution is defined as the limit set, *S**.

Our first result shows that the requirement imposed on the solution is not too stringent –that is, for every BPSC there exists a non-empty set of agreements consistent with IIIA.

Proposition 1 *There exists a unique Hobbesian bargaining solution.*[19] *Moreover, the set of Hobbesian agreements is always non-empty.*

Proof: Note that, given the assumption that there always exist non-negative gains from agreement, the sets S^t are compact and nested. Therefore their intersection is uniquely defined and, by Tychonov's theorem, it is non-empty as well. Q.E.D.

Proposition 1 shows that, independently of the exact form it takes, just the conceptual increase in the informational content of the description of the bargaining problem is sufficient to provide us with a set of "acceptable" agreements. In general, these agreements need not be unique. Whether the solution is determinate or not depends on the nature of the disagreement game. We shall now prove that for strategic disagreement games the above result can be strengthened: the Hobbes solution singles out a unique, Pareto efficient agreement.

The assumptions we need to make are the following:

[18] Note that, unless S* is itself a member of the sequence, IIIA does not require that fH(S*,D) ⊆ SD(S*).

[19] Recall that we defined bargaining solutions to be set valued. Uniqueness here refers to the set, which without further assumptions cannot be guaranteed to be a singleton.

Assumption 1 *D is continuous in the Hausdorff topology: if a sequence of elements of Σ converges to S in the Hausdorff topology, then the corresponding sequence of disagreement points converges to D(S).*

Assumption 2 *Unless S is singleton, the disagreement outcome is strictly preferred to her worst agreement in S by at least one player: for all $S \subseteq S^0$*, such that *$S \in \Sigma$, there exists $z \in S$ such that $z_i < D_i(S)$ for some $i \in \{1,2,...,N\}$.*

Assumption 1 is straightforward: it posits that small changes in the set of feasible utility allocations should not provoke major changes in the outcome of disagreement. Assumption 2 imposes that there exists some agreement to which at least one player strictly prefers the conflict outcome. That is, it requires that disagreement/conflict do not destroy all what is at stake, but leave some positive part of the surplus for the players. We will return to the relevance and meaning of this assumption after the proof of Proposition 2.

Note that, for every $S^0 \in \Sigma$, the set of *D(.)* satisfying Assumptions 1 and 2 is non-empty.

Proposition 2 *When Assumptions 1 and 2 hold, the Hobbesian bargaining solution selects a unique and efficient agreement.*

Proof: To see that *S** has a unique element, note that, by the continuity of *D(.)*, $\lim_{t\to\infty} D(S^t) = D(S^*)$, and thus $S^* = S^*_{D(S^*)}$. Suppose that *S** is not a singleton. Then, by Assumption 2, *D(S*)* does dominate some points in *S**. Contradiction.

By construction, each set S^t contains the points of the weak Pareto frontier of S^0 that dominate $D(S^{t-1})$. Therefore, the point *S** is on the frontier of S^0. This proves the efficiency of the solution. Q.E.D.

In view of their critical role, let us discuss our assumptions on the disagreement function in more detail. Note first that without continuity, even in the presence of Assumption

2, the Hobbes solution could be set valued, since the sequence of disagreement points starting from d^0, might converge to an interior point of S. On the other hand, if we imposed that IIIA had to apply to the limit set, S^*, as well (c.f. footnote 18), we could drop the continuity assumption. However, as a principle, we prefer to put more structure on the (empirically testable) disagreement game rather than to increase our normative requirements (no matter how reasonable) on the solution.

As for Assumption 2, it is satisfied in most settings. As we argued earlier, this assumption is satisfied by all non-cooperative games in which there is at least one player that has a choice over a set of possible strategies and that in equilibrium is not indifferent to all of them. It is plain that in such type of strategic games at least this player obtains in equilibrium a payoff that is strictly higher than the worse feasible payoff. Examples abound: in pre-trial bargaining the lawyer's fees are often set as a percentage of the amount under dispute; in collusive agreements in a market setting, even if there is cut-throat Bertrand competition, unless the firms are identical, there are always positive profits for the more efficient firm; in conflict models with endogenous choice of effort there is usually a unique interior Nash equilibrium, etc. [20] [21]

3.3. Hobbes and conventional arbitration

One of the most straightforward interpretations of the disagreement function is that it actually represents an arbitrator's expected decision, modified by the costs that the bargainers need to incur if they use the services of an arbitrator. In this scenario the question that begs to be asked is: what conditions need the arbitrator's theory and cost allocation rule satisfy to guarantee that the Hobbesian agreement coincides with the arbitrator's theory? In other

[20] Esteban and Ray (1999) show that for a generalised version of the rent-seeking model, there always exists a unique Nash equilibrium and at this equilibrium each contending party expends strictly positive amounts of resources. It is straightforward to show that the disagreement point generated by the Nash equilibrium satisfies our Assumptions 1 and 2.

[21] Notice that the familiar case of bargaining over the price to be paid for an object to be traded violates in principle Assumption 2.

words, what distributional rules used by the arbitrator can be consistently implemented by Hobbesian negotiation under voluntary arbitration? The adverb "consistently" is of course crucial here. We are not interested in how can the arbitrator misrepresent her preferences in order to induce her favourite bargaining outcome. Rather, we would like to know which preferences are such that if stated truthfully, they are respected by the Hobbesian solution.

Definition 2 *An arbitrator's solution is* consistent *if and only if it assigns the same payoffs as the Hobbesian agreement in the shadow of this very arbitration rule.*

To make this exercise meaningful, we need to restrict the set of cost allocation rules, so that they do not take over the role of the distribution theory itself. A sensible cost allocation rule satisfying this requirement is that the arbitrator "charges" each player in proportion (λ) to his gain over his disagreement payoff. Let us denote the arbitrator's solution to the standard[22] bargaining problem (S,d) by A(S,d). Then, we have that $D(S_{D(S)}) = A(S,D(S)) - \lambda[A(S,D(S)) - D(S)]$. To keep everything simple, we assume that A(.,.) satisfies IIIA: $A(S,D(S)) = A(S_{D(S)},D(S))$. As we remarked earlier, for a standard bargaining problem the IIIA is a very weak requirement. Let us define now a stronger requirement, in the spirit of IIIA for Generalised Bargaining Problems.

Definition 3 *A standard bargaining solution satisfies Generalised IIIA, if it is robust to the elimination of points not Pareto dominating any given point that is Pareto dominated by the solution. That is, $A(S,d) = A(S_d,d)$, for any $d << A(S,d)$.*

Note that this axiom is still weaker than IIA, since it only eliminates a certain type of irrelevant alternatives.

Proposition 3 *An arbitration scheme is consistent if and only if the arbitrator's solution satisfies GIIIA.*

[22] Since the arbitrator's solution is an exogenous concept, it should not be defined over the generalised bargaining problem.

Proof: Note that $A(S_{D(S)},D(S)) = A(S,D(S)) = H(S,D(.)) = H(S_{D(S)},D(.)) = A(S_{D(S)},D(S_{D(S)}))$. The first equality follows from IIIA, the second from consistency, the third from the derivation of the Hobbes solution, while the last one again from consistency. The equality of the first and last terms, implies that GIIIA has to be satisfied at the sequence of partial agreements leading to the Hobbes solution. Since the requirement on the distributive theory cannot depend on the Hobbes solution, we need GIIIA to hold in general.

To see sufficiency, note that the proportional cost rule means that the disagreement points are always on the straight line connecting the arbitrator's solution with the first disagreement point. But this means that their sequence must converge to this solution, while the limit of this sequence is, by definition the Hobbesian agreement. QED.

Corollary *The Nash solution with proportional costs is a consistent arbitration scheme.*

Proof: Just note that IIA implies GIIIA. QED.

Corollary *The Kalai-Smorodinsky solution cannot be part of a consistent arbitration scheme.*

Proof: The Kalai-Smorodinsky solution clearly violates GIIIA. QED.

3.4. Hobbes' as an asymmetric Nash solution

Recall that the asymmetric Nash solution (see Harsányi and Selten, 1972) results from the constrained maximisation of a social welfare function where the individual welfare weights are supposed to embody the differential (bargaining) power of the players: $W(x,d) = \prod_{i=1}^{N}(x_i - d_i)^{\gamma_i}$. We shall now discuss the relationship between the vector γ and the power of the parties as embodied in the disagreement function.

Since the Hobbes solution selects a unique point on the Pareto frontier, it can obviously be interpreted as an asymmetric Nash solution. As is well known, this solution can be characterised as the point on the Pareto frontier, where the pair-wise elasticity of this frontier is equal to the corresponding ratio of the bargaining weights. In general, one needs to calculate the Hobbes solution, in order to derive the associated bargaining weights. However, restricting attention to a relevant subset of disagreement functions, these weights can be directly given.

Let us then make the simplifying assumption that the disagreement function satisfies *proportionality*, i.e. $D(\lambda S) = \lambda D(S)$ for all $\lambda > 0$.[23] Note that this scenario is equivalent in "richness" to the one analysed by Rubinstein (1982), in the sense that in both models at each step of the process, the pie remaining in dispute decreases at some given proportion.[24] As our next proposition shows, in this setting the Hobbes solution is very simple and intuitive: the players distribute utilities (efficiently) in the same proportion as the disagreement function does. Consequently, the pair-wise ratio of bargaining weights corresponds to the elasticity of the Pareto frontier at the point where the utilities are distributed in the same proportion as in the disagreement point.

Proposition 4 *Let the disagreement function be proportional. Then, the Hobbes solution satisfies* $\frac{f_i^H(S,D)}{D_i(S)} = \frac{f_j^H(S,D)}{D_j(S)}$ *for i, j = 1,2,..., N.*

Proof: Recalling the proof of Proposition 1, we only need to prove that $\frac{d_i^t}{d_j^t} = \frac{d_i^{t+1}}{d_j^{t+1}}$.

Note that $d^{t+1} = d^t + D(S^t)$, by definition. Therefore,

[23] This condition is somewhat weaker than homogeneity. It is easy to show that the endogenous contest model of Esteban and Ray (1999) mentioned earlier, satisfies this assumption whenever the Pareto frontier of the bargaining set is linear.

[24] Recall, however, that Rubinstein also assumes that there are only two players and the Pareto frontier of the bargaining set is linear (with slope –1).

$$\frac{d_i^{t+1}}{d_j^{t+1}} = \frac{d_i^t + D_i\left(S^t\right)}{d_j^t + D_j\left(S^t\right)} = \frac{d_i^t\left(1 + \dfrac{D_j\left(S^t\right)}{d_j^t}\right)}{d_j^t\left(1 + \dfrac{D_j\left(S^t\right)}{d_j^t}\right)} = \frac{d_i^t}{d_j^t} \ ,$$

where the second equality follows by the hypothesis, $\frac{d_i^t}{d_j^t} = \frac{D_i\left(S^t\right)}{D_j\left(S^t\right)}$. Q.E.D.

In Rubinstein's alternating-offer bargaining model the unique subgame-perfect equilibrium yields an agreement as a function of the discount factors (δ_i) and the selection of the first mover. As the time between offers shrinks to zero this solution converges to the same outcome as the asymmetric Nash solution –with bargaining weights[25] $\gamma_1 = \log \delta_2$ and $\gamma_2 = \log \delta_1$– independently of the identity of the first mover. Assuming that the Pareto frontier is the unit simplex, as Rubinstein does, we can prove a similar result for the Hobbes solution, without having to resort to taking limits. That is, the Hobbes solution will exactly coincide with the asymmetric Nash solution, while Rubinstein's does so only in an approximate sense.

Proposition 5 *When the Pareto frontier is the unit simplex and the disagreement function is proportional, the bargaining weights corresponding to the Hobbes solution are* $\gamma_i = D_i(S)$, *i = 1,2.*

Proof: When the Pareto frontier is the unit simplex, the marginal rate of substitution is 1, everywhere. Consequently the elasticity of the Pareto frontier is equal at every point to the ratio of the utilities at that point. By Proposition 4, this ratio is equal to the ratio of the disagreement utilities. Q.E.D.

The following corollary is now immediate.

[25] See Binmore (1987a,b) and Binmore et al. (1986). Wilson (2000), has obtained the same result in a model with a mediator who makes random proposals.

Corollary *Under the Rubinstein assumptions (including the proportionality of the disagreement function), the Hobbes solution and the Rubinstein solution coincide if and only if* $\frac{D_1(S)}{D_2(S)} = \frac{\log \delta_2}{\log \delta_1}$.

When the disagreement function is not restricted to be proportional, our model still resembles somewhat a Rubinstein-like model, where the discount rates are not stationary (see Binmore, 1987b, for a detailed discussion of these games). Both models are still equivalent to some asymmetric Nash solution. However, the bargaining weights –just as the actual solutions– are no longer easily computable. In terms of computability, the Hobbes solution has a significant advantage over the Rubinstein-like one: each step in the calculation of the Hobbes solution improves the precision of the current estimate, and this precision is known. In contrast, to calculate the subgame-perfect equilibrium of a Rubinstein-like game, one has to work backwards from the solution, trying to end up at the disagreement point. At no point in the process, can one have a precise idea about how good the approximation is.

4. Renegotiation-proof bargaining

To complement our axiomatic analysis, in this section, we provide non-cooperative foundations for the Hobbes solution, based exclusively on the possibility to renegotiate the disagreement outcome. A crucial difficulty in the strategic implementation of our solution concept is to capture the high degree of collective rationality present in the cooperative formulation within a non-cooperative context. In particular, while it is easy to see that all the players would prefer a (partial) agreement at the disagreement outcome to outright disagreement, in a strategic game, we also would need to argue that it is preferred to a partial agreement at some lower payoffs. One way around this problem could be to rule out the undesired strategies by a technical assumption. However, this method would defeat the original purpose of shedding more light on the issue. Therefore, we will proceed in a way that captures the intuition behind the result. We assume a certain amount of collective rationality that is widely used in non-cooperative game theory: the players will renegotiate any inefficient outcome.

When the players cannot commit not to renegotiate a contract, this necessarily becomes incomplete. We take this observation to the extreme and actually assume that there is no contract signed. While it has been shown that such a "null contract" can actually be optimal from a mechanism design perspective (see Hart and Moore, 1999), our motivation is quite different, as we explain below.

To date, renegotiation has always built upon bargaining theory, since it was considered (even in its etymology) as something that is posterior, more evolved than that. We invert this order of hierarchy. We derive a theory of bargaining from the mere possibility of renegotiation. At first blush, this may sound a bit circular: how can we have a theory of *re*-negotiation, before we have one of negotiation? Note, however, that there is an important difference between the two concepts: renegotiation by its very *raison d'être* implies that the players want to move towards efficiency. In our case this means that they want to move away from the disagreement outcome. The question is: where to? The answer comes from the absence of a theory of negotiation: the only way to ensure a Pareto improvement over and above an inefficient outcome is to assign everybody at least this payoff. Since we have no indication how much more than that, this should be negotiated: thus, let's give everybody his disagreement payoff and let's restart the negotiation about the remainder. This necessarily yields a Pareto improvement, without directly biasing the division of the remaining surplus.

Take any strategic game of negotiation, with the only restriction that each player should be able to unilaterally and costlessly provoke total disagreement (that would then be renegotiated). By the renegotiation method outlined above, it is straightforward to see that – by constantly provoking disagreement – each player can guarantee herself her Hobbesian payoff, and consequently this is the only subgame-perfect equilibrium of the renegotiable bargaining game.

Proposition 6 *In any renegotiable extensive form game of negotiation where any player can unilaterally and costlessly provoke total disagreement, the unique subgame-perfect equilibrium prescribes the Hobbes agreement.*

Proposition 6 translates the known mechanics of disagreement into a theory of agreement. That is, having identified a disagreement function, the simple fact that we assume that players can renegotiate any "outcome" of some non-cooperative game of negotiation identifies a unique candidate for an agreement, which will not be renegotiated.

The generality of Proposition 6 comes from the fact that for many extensive forms the Hobbesian agreement would arise through renegotiation. One might wish for a bargaining procedure that yields agreement just in the shadow of renegotiation but without actually recurring to it. Similarly, one might expect –in a Hicksian manner– that the agreement should be immediate. Note that in a standard game with commitment to offers this is not possible, since a deviation by the player(s) whose offers cannot be observed before some other players need to offer is always profitable conditional on the others following the (hypothetical) equilibrium strategy. Consider, for example, a Nash demand game. If the other players are offering me my Hobbesian share, I am always better off accepting it, not giving them anything and provoke a disagreement over the remainder. Thus, we need a procedure where the players can make conditional concessions: I give up x if you give up y etc. In practice this is often achieved by a process of ratification.

For an example, consider quantity-setting, non-differentiated oligopolists who are trying to collude in a market. Here, offers are self-imposed quantity caps, while disagreement is Cournot competition. In this setup a producer always has time to react to any increase in production of his competitors[26] before the market closes, exogenously ratifying the quantities. Consequently, by dividing up (equally) the monopoly quantity, the Hobbesian agreement is directly implementable, since all the producers know that by unilaterally increasing their production they would trigger a response by their competitors, making the deviation unprofitable.

[26] This argument is reminiscent of Sweeney's kinked demand curve model, with the important difference that we do have a theory where the kink should be.

5. A comparative analysis

In this section, we clarify our theory by contrasting it to the most related papers and ideas.

i) Hobbes and the theory of bargaining.

Binmore (1994, 1998) has also invoked Hobbes in support of bargaining theory. However, he identified Hobbes' state of nature with the *status quo* point, not with the disagreement point (as we do). The question here is not who is right and who is wrong. Simply, the different interpretations correspond to different social situations. Binmore has in mind a bargaining problem that is about possible improvements over an already existing contract. In that case, if the agents do not reach agreement, they continue respecting the old contract. We, on the other hand, are thinking of an incomplete contract scenario, where there is no fall-back option and thus the conflict of interests must be resolved: either by consensus or by conflict.

ii) Endogenous determination of the disagreement point.

In his 1953 paper, Nash proposed a generalisation of his original model of 1950. In this game, known as the "variable threat" model of bargaining, the players choose threats before the actual bargaining phase, of which they serve as the disagreement point. At first blush, our model may seem just like Nash's one, with a specific, well-motivated threat game. Actually, however, our contribution goes well beyond that. There are two important differences between the models that we would like to underline:

a) Nash needs to employ an "umpire" to oblige the players to carry out their threats (in case of disagreement). We do without a $n+1^{st}$ party. The underlying reason for this is quite relevant. Nash thinks of the threat phase as one preceding the Nash bargaining game. Therefore, this phase has no interpretation on its own, it is simply a –perhaps realistic– way to make the bargaining game more detailed. In contrast, we think of our conflict subgame as one posterior to bargaining. By invoking sequential rationality, we can then analyse the

players' optimal behaviour in that subgame without any additional commitment device. Apart from the obvious difference in philosophy, the technical difference is also apparent, since in Nash's game by a well-chosen threat (which she would prefer not to carry out) a player can improve her share, without her bluff ever being called. Thus, even if we used our conflict game as the threat game, the equilibria would differ, since the players, in general, would not use a threat that forms part of an equilibrium of the conflict game.

b) When Nash's players generate a disagreement point, he considers the bargaining problem properly defined and proceeds to its solution (according to his 1950 paper). In contrast, we argue that they have simply arrived at a new bargaining situation, where they might wish to employ different threats than before. To put it another way: while in the Nash model the demand phase depends on the outcome of the threat phase, in our model the conflict game is supposed to depend on the demands made (when they are not compatible).

iii) Step-by-step resolution.

Kalai (1977) introduced the axiom of decomposability. This assumption requires that if we break up the set of available agreements, S, into two subsets, X and Y, then using the solution of (either) one of these as a partial agreement to subsequently bargain over the rest, $(S\text{-}f(X, d)) \bigcap \Re_+^N$, should give the same result as applying the solution directly. Note that Kalai's model agrees with ours in the idea that partial agreements are only renegotiated if this yields a Pareto improvement. On the other hand, Kalai does not propose a well-defined solution: he only establishes that the solution should be "proportional," without identifying what should be these proportions. In addition, Kalai's model has two caveats, first pointed out by Ponsati and Watson (1997). The first of these is that when agreeing on the first sub-problem, the bargainers of Kalai are not supposed to take into account the effect of today's agreement on tomorrow's one. This is not true in our model. Second, there seems to be an inconsistency between the assumption that the agreement on the first subproblem is binding, but at the same time can be renegotiated –since the second sub-problem is not $S \backslash X = Y$ but $(S\text{-}f(X, d)) \bigcap \Re_+^N$. In our model, however, these two sets coincide so we avoid any confusion.

Wiener and Winter (1999) propose a solution for bargaining problems where the feasible set is exogenously divided up into smaller pieces. Their solution is equivalent to agreeing step-by-step on each "crumb" according to the Nash solution, using the result of the previous step as the new disagreement point. This procedure is similar to ours, but we use the disagreement function to determine the new *status quo* and we do not need the arbitrary division.

iv) An auxiliary function, which maps each bargaining problem into a point in the utility space.

Thomson (1981) introduced the concept of a *reference function*, the purpose of which is to summarise the relevant features of a bargaining problem. This function maps every bargaining problem into a reference point, which is then used to calibrate the bargaining power of the players. While, at first blush, our disagreement function may sound just like a special case, actually the two approaches are diametrically opposed. The role of a reference function is to summarise information that is already present in the bargaining problem. In contrast, our approach complements the originally available information with the outcome of conflict, which possibly depends on additional factors.

v) Bargaining under the threat of some outside enforcement mechanism.

This topic has been extensively dealt with in the applied literature (pre-trial negotiations, strikes, arbitration etc.). Perhaps, the piece closest to our approach is Powell (1996). Powell sets up a non-cooperative bargaining game where the players can choose to force a (probabilistic) settlement at some cost. The important difference with respect to our approach is that, in his model, forcing the settlement is equivalent to taking an *outside option*. However, outside options do not determine, in general, the outcome of a bargaining game. Therefore, Powell needs to rely on the solution to the bargaining game, which would come about in the absence of outside options. In our case, in contrast, the solution of the game cannot be dissociated from the underlying conflict situation.

vi) Recursive solutions.

We are not the first ones to use a recursive application of some rule in bargaining theory. Let us mention just a couple. Raiffa (1953) proposes a method where the players first pocket half of their most preferred allocation, then half of their most preferred allocation in the remainder... etc. While in (its recursive) structure his procedure is very much like ours, the important difference is that he has no justification other than some vague consideration of "fairness" for the fifty percent rule. van Damme (1986) considers a recursivity axiom which imposes that if the players are making demands according to some individual theories, then in every step of the iteration, as a function of these demands some subset of S is to be discarded, and the negotiation resumed. Technically, the IIIA assumption is very similar, with the important difference that we only invoke individual rationality for discarding "irrelevant alternatives."

vii) The local shape of the Pareto frontier matters.

The Hobbes solution relaxes Nash's Independence of Irrelevant Alternatives (IIA) axiom to a large extent, since an infinite number of (endogenously determined) points of the Pareto frontier affect it. While most non-Nash bargaining solutions also relax IIA, the one that comes closest to ours in this respect is the Perles and Maschler (1981) solution. According to this concept, the players start at their most preferred outcome and trace the Pareto frontier by simultaneously lowering their demands at the speed that corresponds to the slope of the Pareto frontier at their current proposal. Our main criticism of the Perles-Maschler solution is that, while the rate of concession equalling the "marginal rate of substitution" is an appealing idea, it continues to be arbitrary.

viii) Disagreement modelled as a non-cooperative game.

Lundberg and Pollak (1993) replace divorce by a non-cooperative equilibrium within marriage, as the disagreement point in a model of marital bargaining. While they implicitly recognise that the forces determining the threat point are the same ones that influence the

bargaining process, they do not make this connection explicit, and simply use the Nash solution.

6. Concluding remarks

In this paper we have presented a new approach to the theory of negotiation and have introduced the corresponding agreement concept. The cornerstone of our theory is the more efficient use of information that was already necessary for the standard theory: the description of the non-cooperative resolution of conflict. Indeed, we use not only the utility allocation in a particular equilibrium (the disagreement point), but we make full use of the primitives behind this equilibrium. In fact, we have shown that the disagreement function contains sufficient information to derive a unique agreement when coupled with a mild generalisation of individual rationality. Our results thus prove the power of focusing on the "state of nature" in order to understand social agreements, as proposed by Hobbes.

We consider our theory to be complementary to the one based on time preferences. In scenarios where delay costs (and the risk of breakdown) are negligible with respect to the stakes of negotiation, like political disputes; or where disagreement leads into conflict which generates inefficiencies that are not related to delay, our approach seems to be more appropriate. In addition, the Hobbes solution yields a unique solution for an arbitrary number of negotiators, while the alternating-offers models usually generate multiple equilibria for more than two players.

Finally, we should emphasise that we have presented our model based on cardinal preferences only to minimise our departure from standard theory. It is easy to see that we need not restrict attention to the utility space in order to derive our results. Any underlying space of bargaining outcomes, together with a complete preference relation, would suffice. In other words, our theory is one based on ordinal preferences, an unreachable goal for solutions to the standard bargaining problem.

REFERENCES

Anderson, S., Goeree, J. and C. Holt (1998), "Rent Seeking with Bounded Rationality: An Analysis of the All-Pay Auction," *Journal of Political Economy* 106(4), 828-853.

Aumann, R. and M. Maschler (1985), "Game Theoretic Analysis of a Bankruptcy Problem from the Talmud," *Journal of Economic Theory* 36, 195-213.

Becker, G. (1983), "A Theory of Competition among Pressure Groups for Political Influence," *Quarterly Journal of Economics* 98, 371-400.

Binmore, K. (1987a), "Nash Bargaining Theory II," Chapter 4 in *The Economics of Bargaining* (eds. Binmore and Dasgupta) Basil Blackwell, Oxford.

Binmore, K. (1987b), "Perfect Equilibria in Bargaining Models," Chapter 5 in *The Economics of Bargaining* (eds. Binmore and Dasgupta) Basil Blackwell, Oxford.

Binmore, K. (1994), *Game Theory and the Social Contract: Playing Fair*, The MIT Press, Cambridge (Mass.).

Binmore, K. (1998), *Game Theory and the Social Contract: Just Playing*, The MIT Press, Cambridge (Mass.).

Binmore, K., Rubinstein, A. and A. Wolinsky (1986), "The Nash Bargaining Solution and Economic Modelling," *RAND Journal of Economics* 17(2), 176-188.

Chen, M. and E. Maskin (1999), "Bargaining, Production, and Monotonicity in Economic Environments," *Journal of Economic Theory* 89, 140-147.

van Damme, E. (1986), "The Nash Bargaining Solution is Optimal," *Journal of Economic Theory* 38, 78-100.

Esteban, J. and D. Ray (1999), "Conflict and Distribution," *Journal of Economic Theory* 87, 379-415.

Forsythe, R., Horowitz, J., Savin, N. and M. Sefton (1994), "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior* 6, 347-369.

Gauthier, D. (1990), *Moral Dealing. Contracts, Ethics and Reason*, Cornell University Press, Ithaca N.Y..

Grossman, H. (1991), "A General Equilibrium Model of Insurrections," *American Economic Review* 81, 912-921.

Grossman, H. (1994), "Production, Appropriation and Land Reform," *American Economic Review* 84, 705-712.

Grossman, H. and M. Kim (1995), "Swords or Plowshares? A Theory of the Security of Claims to Property," *Journal of Political Economy* 103(6), 1275-1288.

Harsányi, J. and R. Selten (1972), "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information," *Management Science* 18, 80-106.

Hart, O. and J. Moore (1999), "Foundations of Incomplete Contracts," *Review of Economic Studies* 66(1), 115-138.

Hirshleifer, J. (1991), "The Paradox of Power," *Economics and Politics* 3, 177-200.

Hirshleifer, J. (1995), "Anarchy and its Breakdown," *Journal of Political Economy* 103, 26-52.

Horowitz, A. (1993), "Time Paths of Land Reform: A Theoretical Model of Reform Dynamics," *American Economic Review* 83(4), 1003-1010.

Kalai, E. and M. Smorodinsky (1975), "Other Solutions to Nash's Bargaining Problem," *Econometrica* 43, 513-518.

Kalai, E. (1977), "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons," *Econometrica* 45(7), 1623-1630.

Knight, J. (1992), *Institutions and Social Conflict*, CUP Cambridge.

Levine, D. (1998), "Modeling Altruism and Spitefulness in Experiments," *Review of Economic Dynamics* 1, 593-622.

Lundberg, S. and R. Pollak (1993), "Separate Spheres Bargaining and the Marriage Market," *Journal of Political Economy* 101(6), 988-1010.

Nash, J. (1950), "The Bargaining Problem," *Econometrica* 18, 155-162.

Nash, J. (1953), "Two Person Cooperative Games," *Econometrica* 21, 128-140.

Perles, M. and M. Maschler (1981), "A Super-Additive Solution for the Nash Bargaining Game," *International Journal of Game Theory* 10, 163-193.

Pollak, R. (1994), "For Better or Worse: The Roles of Power in Models of Distribution within Marriage," *American Economic Review* 84 (Papers and Proceedings), 148-152.

Ponsati, C. and J. Watson (1997), "Multiple-Issue Bargaining and Axiomatic Solutions," *International Journal of Game Theory* 26, 501-524.

Powell, R. (1996), "Bargaining in the Shadow of Power," *Games and Economic Behavior* 15, 255-289.

Rabin, M. (1993), "Incorporating Fairness into Game Theory and Economics," *American Economic Review* 83, 1281-1302.

Raiffa, H. (1953), "Arbitration Schemes for Generalized Two-Person Games," in Kuhn and Tucker (eds.) *Contributions to the theory of games II*, Annals of Mathematics Studies #28. Princeton University Press.

Roemer, J. (1988), "Axiomatic Bargaining Theory on Economic Environments," *Journal of Economic Theory* 45, 1-31.

Roemer, J. (1996), *Theories of Distributional Justice*, Harvard University Press, Cambridge (Mass.).

Rousseau, J.J. (1782), *Discours sur l'Origine et les Fondements de l'Inégalité parmi les Hommes*, London.

Rubinstein, A. (1982), "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50, 97-109.

Skaperdas, S. (1992), "Cooperation, Conflict, and Power in the Absence of Property Rights," *American Economic Review* 82, 720-739.

Smith, A. (1776), *The Wealth of Nations*.

Svejnar, J. (1986), "Bargaining Power, Fear of Disagreement, and Wage Settlements: Theory and Evidence from U.S. Industry," *Econometrica* 54(5), 1055-1078.

Taylor, M. (1987), *The Possibility of Cooperation*, Cambridge University Press.

Thomson, W. (1981), "A Class of Solutions to Bargaining Problems," *Journal of Economic Theory* 25, 431-441.

Tullock, G. (1980), "Efficient Rent Seeking," in J.M. Buchanan, R.D. Tollison and G. Tullock (eds.) *Toward a Theory of the Rent-Seeking Society*, College Station: Texas A&M University Press, 97-112.

Wiener, Z. and E. Winter (1999), "Gradual Bargaining," mimeo, Hebrew University, Jerusalem, March.

Wilson, C. (2000), "Mediation and the Nash Bargaining Solution," *Review of Economic Design*, forthcoming.