Bargaining, interdependence, and the rationality of fair division

Giuseppe Lopomo*

and

Efe A. Ok**

We consider two-person bargaining games with interdependent preferences and bilateral incomplete information. We show that in both the ultimatum game and the two-stage alternating-offers game, our equilibrium predictions are consistent with a number of robust experimental regularities that falsify the standard game-theoretic model: occurrence of disagreements, disadvantageous counteroffers, and outcomes that come close to the equal split of the pie. In the context of infinite-horizon bargaining, the implications of the model pertaining to fair outcomes are even stronger. In particular, the Coase property in our case generates "almost" 50-50 splits of the pie, almost immediately. The present approach thus provides a positive theory for the frequently encountered phenomenon of the 50-50 division of the gains from trade.

1. Introduction

■ Fair divisions of gains from trade are commonly observed in daily life. They occur even in the case of bargaining among two asymmetrically placed parties, one of whom holds a clear strategic advantage over the other. In addition to the substantial anecdotal evidence to this effect, the numerous experimental studies since the early 1980s have shown that fair outcomes play a truly focal role in bargaining situations. Today most authors appear to agree on at least three major robust, yet unexpected, empirical regularities that arise in bargaining games. First, proposed divisions accumulate around the 50-50 division; the actual outcomes are "more fair" than the usual prediction.[1] Second, rejections, which should never be observed on the equilibrium path, occur in significant numbers. Third, more often than not, subjects who reject an offer make

* Duke University; glopomo@mail.duke.edu.

** New York University; efe.ok@nyu.edu.

The authors would like to thank Jean-Pierre Benoît, Gary Bolton, Kalyan Chatterjee, Faruk Gul, Levent Kockesen, David Pearce, Rajiv Sethi, Ennio Stacchetti, and the seminar participants at Brown, Michigan, NYU, Pompeu Fabra, SUNY at Albany and Stony Brook, and UCSD for helpful comments. We especially acknowledge the critical, yet insightful, suggestions of the Editor Jennifer Reinganum and an anonymous referee of this Journal. Of course, the usual disclaimer applies. Ok gratefully acknowledges financial support from the National Science Foundation through grant no. SBR-9809208. We also thank the C.V. Starr Center for Applied Economics at New York University for its support of this project.

[1] Perhaps the most striking example of this phenomenon is provided by Güth, Schmittberger, and Schwarze (1982), who studied experimentally the two-player ultimatum game and found that the average proposal by first-movers was roughly in the neighborhood of (60%, 40%), with about 20% of the proposals being rejected.

a disadvantageous counteroffer, that is, after rejecting a proposal that would leave them with x dollars, they propose a new division that spares them less than x dollars.[2] Clearly, all of these observations are in contrast with the standard game-theoretic predictions, thereby putting bargaining theory, which has significant applications in the fields of industrial organization and international trade, on a hot seat.

One way of interpreting these findings is to argue that pure expected profit maximization cannot be the only criterion guiding the choices of the players in bargaining games. For instance, among others, Ochs and Roth (1989) have suggested that a notion of "fairness" may be influencing the subjects' behavior. In a complementary study, Bolton (1991) has proposed the idea that "fairness" may guide a subject's behavior *only* when he is getting *less* surplus than his opponent. To account for rejections and disadvantageous counteroffers, Bolton also considered informally a simple model in which players have incomplete information about their opponents' preferences.[3]

In the present article, we advance that the said experimental regularities can be explained by modelling suitably the "fear of rejection" that a bargainer may feel throughout the negotiation process. While this idea is not new (e.g., it is implicit in Bolton (1991)), it has received little attention from the bargaining theorists, probably because the notion of "fear of rejection" is not a primitive of the game but rather implicit in the preferences and beliefs of the players. To bring out reasonable predictions from this notion, then, one needs to alter the bargaining models through the utility functions and/or beliefs of the players. To this end, we investigate here the implications of the possibility that each player's utility depends not only on her absolute level of earnings, but also on her *relative* share of the total surplus.[4] In the resulting bargaining model, each player's *type* is referred to as "independent" if her utility depends exclusively on her absolute earnings, and "(negatively) interdependent" otherwise. The latter kind of preferences relate to the time-honored relative income hypothesis and are thus much studied in various branches of economic literature. In the context of bargaining, on the other hand, a negatively interdependent individual may be thought of as a *competitive* player who does not wish to "lose the game" unless he is sufficiently compensated for it. We shall show that this induces the proposing players to possess possibly significant levels of "fear of rejection," and when it is combined with the realistic hypothesis that one's true degree of competitiveness (interdependence) is private information, it leads to predictions that are consistent with experimental findings on bargaining games. Notably, this is not so simply because "players are spiteful," but because of the much more reasonable hypothesis that "a player might *think* (i.e., assign positive probability to the event) that his opponent may care about her own relative share."[5]

The objective of the present article is, then, twofold. We shall first show that extending the Stahl-type bargaining models by incorporating the notion of "fear of rejection" through the *possible* negative interdependence of individuals leads to equilibrium predictions that accord well with the qualitative regularities obtained in experiments, especially with those that relate to fair divisions of gains from trade.[6] In turn, this observation encourages one to investigate the impli-

[2] See, in particular, Ochs and Roth (1989) and Bolton (1991). Camerer and Thaler (1995) and Roth (1995) provide excellent surveys of the related experimental literature.

[3] Bolton's incomplete-information model is, however, built on restrictive assumptions. For instance, it is assumed *a priori* in this model that the true distribution of players' types is such that the game has a unique equilibrium, in which the initial offer is always accepted. Moreover, the formal examination of the properties of the sequential equilibria of this model is absent in Bolton's otherwise penetrating analysis.

[4] Alternative models that perturb the individual utility functions are considered by Rabin (1993), Daughety (1994), Kirchsteiger (1994), Andreoni and Miller (1996), Levine (1998), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000).

[5] We stress that the notion of "fear of rejection" is not modelled here in an *ad hoc* fashion, but is rather introduced through individual preferences (which are primitives of a game-theoretical model). Put differently, our model does not assume but *predicts* behavior that exhibits a fear of rejection.

[6] The work of Daughety (1994), who considers an incomplete-information setup with positive and negative interdependence, is particularly related to this part of our work. In particular, that article also explains increased sharing in the ultimatum game and provides an example in which disadvantageous counteroffers occur in equilibrium. By contrast, we work here with a more general utility specification (*only* with negative interdependence) and provide a relatively general analysis of finite- and infinite-horizon bargaining models.

cations of the potential negative interdependence of players in infinite-horizon bargaining games. Our second objective is to carry out this task, and we argue formally that such an extension allows one to draw striking conclusions in such models with regard to egalitarian outcomes. There is, of course, a literature that studies the emergence of the equal division as a focal point, but few studies have examined the conditions under which the 50-50 split is actually an equilibrium outcome of a bargaining game played among *rational* individuals.[7] We aim to show that very egalitarian outcomes can indeed be sustained in equilibrium, provided that bargainers have *possibly* negatively interdependent preferences. In fact, it turns out that in a variety of bargaining contexts, the slightest doubt that players may have about their opponents being excessively competitive leads to a very forceful "fear of rejection" and hence induces very fair outcomes. Furthermore, as we shall demonstrate by using a mechanism design approach, the possibility of interdependence leads to such outcomes also in bargaining scenarios that are resolved by an arbitrator.

It is worth noting at the outset that we do not suggest that negatively interdependent preferences could be used to explain the data from a large variety of experiments. Here we wish to model rigorously (and endogenously) the notion of "fear of rejection," and hence, in contrast to Andreoni and Miller (1996), Levine (1998), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000), we focus exclusively on bargaining experiments.[8] Moreover, by no means do we regard the present article as providing evidence for the simplistic statement that "people are spiteful."[9] Instead, we argue that negatively interdependent preferences allow one to extend the usual bargaining model in a way to incorporate the potential competitiveness/aggressiveness of the bargainers and, hence, to bring to the fore the notion of "fear of rejection." In the context of bargaining, such preferences make good intuitive sense, and they produce qualitative predictions that match the data well. What is more, they lead to interesting theoretical results in the context of the usual infinite-horizon bargaining models. Our main message is simply that it is possible to substantiate the theory of bargaining by modelling the bargainers as possibly negatively interdependent agents, at least insofar as the predictions about the egalitarian divisions of gains from trade are concerned.

The rest of the article is organized as follows. In Section 2 we introduce the general formulation of beliefs and preferences of individuals that we shall adhere to throughout. Each of the subsequent four sections will adopt these primitives (with minor technical alterations) but change the rules of bargaining. In particular, in Sections 3 and 4 we examine the standard one- and two-period models, respectively, and prove the existence of perfect Bayesian equilibria that envisage fair outcomes along with rejections and disadvantageous counteroffers. An important finding of Section 4 is that disadvantageous counteroffers cannot occur in a model in which players' *utilities* are discounted, while the monetary surplus remains constant, through time. We therefore claim that disadvantageous counteroffers can only be observed in the experiments in which, for practical reasons, one discounts the size of the pie (or the value of the shares of the players). This is a concrete prediction of the present model: disadvantageous counteroffers cannot take place in finite-horizon bargaining in which discounting is induced by the passage of real time that leaves the physical size of the gains from trade intact. It would be interesting to test experimentally the validity of this claim.

In Section 5 we focus on the implications of *possibly* negatively interdependent preferences

[7] Most articles that examine the origin of fair outcomes in bargaining games are evolutionary in nature. Young (1993), for instance, studies an evolutionary model of the Nash demand game played between two populations who learn adaptively. He shows that the equal split can be the unique stable division depending on the nature of the expected utility function. (See Ellingsen, 1997, for a similar analysis.) Bolton (1997), on the other hand, provides an alternative bargaining game, of which at least one (limit) evolutionarily stable equilibrium results in the equal split.

[8] For example, in the standard "dictator game," where one subject single-handedly decides how to divide a given surplus between himself and another subject, the predictions of our model (like those of the "pure fairness" models) are not really in concert with the experimental findings. (See Forsythe et al., 1994.) We shall elaborate on this issue further in Section 7.

[9] After all, it is not clear why an individual cannot act altruistically in, say, a public-good problem that involves many potential contributors and behave aggressively in a two-person face-to-face bargaining situation. Admittedly, however, endogeneity of preferences is a complicated (but very interesting) issue about which we have little to say at present.

in infinite-horizon bargaining. It turns out that an interesting case in which players have negatively interdependent preferences is essentially identical to the "gap case" of the much-studied buyer-seller bargaining model with risk-neutral players. Therefore, all the results known in that context can be shown to apply to the present setup. In particular, the famous "Coase conjecture," established by Gul, Sonnenschein, and Wilson (1986), implies the following result in the present setting: as the frequency of offers increases arbitrarily, even the smallest doubt in the mind of the player who makes all the offers about his opponent's degree of (negative) interdependence being relatively high, induces him to propose a division relatively close to the 50-50 split, almost immediately. We believe that this result is but an important step toward providing a rational theory of equal division.

Many bargaining situations in the real world do not fit perfectly with the theoretical bargaining models, for they are settled by arbitration. A natural question, then, is whether the potential negative interdependence of preferences would enforce the emergence of fair outcomes in bargaining settlements as well. By adopting a mechanism-design approach and exploiting the intuition that negative interdependence forces fear of rejection in such a context through individual-rationality constraints, we show in Section 6 that this is indeed the case. Section 7 concludes.

2. The bargaining environment and interdependent preferences

■ **The environment.** We consider a bargaining environment in which two players try to agree on how to divide a pie of size $2m$, where $m > 1$. In case of disagreement, each player receives $\varepsilon \in (0, m)$.[10] Thus, the size of the pie, net of the bargainers' holdings, is $2m - 2\varepsilon$. Equivalently, we can assume that the players have an initial level of wealth ε and are bargaining over a pie of size $2m - 2\varepsilon$, which is wasted if disagreement occurs. Without loss of generality, we let $\varepsilon = 1$ so that the set of all feasible divisions of the pie is $X \equiv \{(x_A, x_B) \geq (1, 1) : x_A + x_B \leq 2m\}$, while the set of all efficient divisions is given by

$$Y \equiv \{(x_A, x_B) \in X : x_A + x_B = 2m\}.$$

In this article, we examine the consequences of the *possibility* that the bargainers' welfare depend not only on the absolute gains that they may achieve through bargaining, but also on the *relative* sizes of the slices of the pie that they get. Thus, given an allocation $(x_A, x_B) \in X$, we assume that the utility of player i depends not only on x_i, but also on $x_i/\bar{x}$, where $\bar{x}$ denotes the average payoff in the allocation, i.e., $\bar{x} \equiv (x_A + x_B)/2$. Of course, it is realistic to posit that the players are unsure about their opponents' degree of interdependence, and hence the presence of interdependence forces us to consider an incomplete-information model.[11] To this end, we take an arbitrary $\Theta_i > 0$ and designate a subset $\mathcal{T}_i$ of the interval $[0, \Theta_i]$, with $0 \in \mathcal{T}_i$, as the *type space* of player $i = A, B$. The beliefs of $j \neq i$ will thus be modelled through a distribution function whose support is contained in $\mathcal{T}_i$. In turn, the utility function $u_i : X \times \mathcal{T}_i \to \mathbb{R}$ of player i is expressed as

$$u_i\left(x_A, x_B \mid \theta_i\right) \equiv V_i\left(x_i, \frac{x_i}{\bar{x}} \mid \theta_i\right), \qquad i = A, B, \tag{1}$$

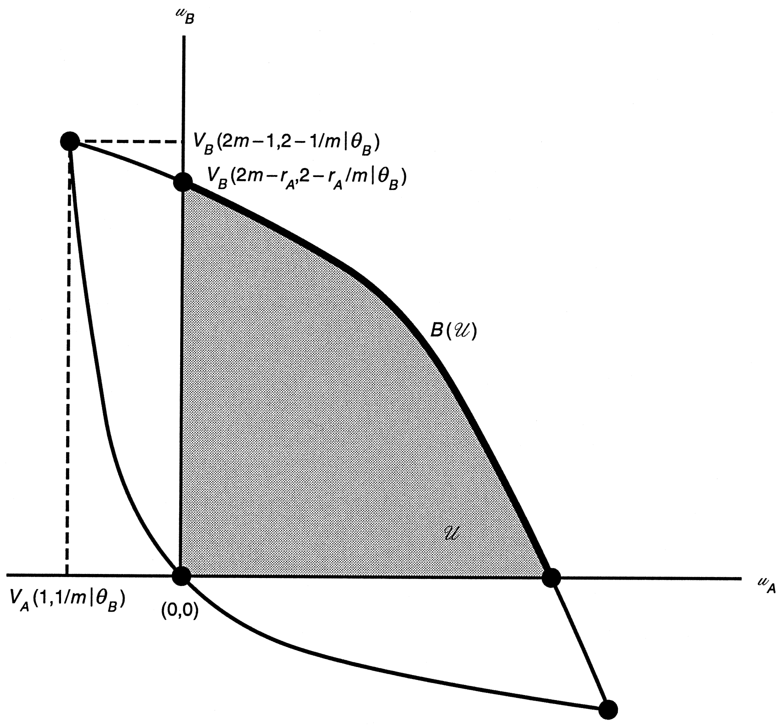
where all functions $V_i(\cdot, \cdot \mid \theta_i)$ for $\theta_i \in \mathcal{T}_i \backslash \{0\}$ are continuous and strictly increasing in both arguments, while $V_i(\cdot, \cdot \mid 0)$ is strictly increasing in its first argument and independent of the second argument. Moreover, in what follows, we impose the following normalization condition:

$$V_i(1, 1 \mid \theta_i) = 0 \qquad \text{for all } \theta_i \in \mathcal{T}_i \backslash \{0\},\ i = A, B. \tag{2}$$

[10] We require $\varepsilon > 0$ in order to avoid the indeterminate form $0/0$.

[11] Bolton (1991, p. 1112), says that "the marginal rate of substitution between absolute and relative money most likely varies by individual, making utility functions private information." A similar point was made also by Kennan and Wilson (1993, p. 93), who argue that the most bargaining experiments in the literature "can be interpreted, in effect, as involving bargaining with private information, as evidently most players did not know the preferences of the opposing bargainer."

FIGURE 1



In words, we model the preferences as (possibly) being *negatively interdependent*: player i of type $\theta_i \in \mathcal{T}_i \backslash \{0\}$ cares not only about her share of the pie x_i, but also about how x_i compares with the average level of earnings $\bar{x}$.[12] By contrast, the present formulation maintains that $\theta_i = 0$ corresponds to the type with (standard) *independent* preferences.

Given a type profile (θ_A, θ_B), the *bargaining problem* (in the sense of Nash) associated with the present setting is $(\mathcal{U}, (0, 0))$, where $(0, 0)$ is the disagreement point and $\mathcal{U}$ is the utility possibility set $\{(V_A(x_A, x_A/\bar{x} \mid \theta_A), V_B(x_B, x_B/\bar{x} \mid \theta_B)) : x \in X\}$. The *bargaining set* $B(\mathcal{U})$ of this problem is, in turn, defined as the set of all Pareto optimal and individually rational utility allocations in $\mathcal{U}$, that is,

$$B(\mathcal{U}) \equiv \left\{ \left(V_A \left(x_A, \frac{x_A}{\bar{x}} \mid \theta_A \right), V_B \left(x_B, \frac{x_B}{\bar{x}} \mid \theta_B \right) \right) \geq (0, 0) : x \in Y \right\}.$$

As Figure 1 illustrates, due to the individual-rationality constraints, this set is typically smaller than the bargaining set of the usual setup (with independent preferences).

□ **Fear of rejection.** The presence of negative interdependence is the main feature that distinguishes this article from the existing literature on bargaining theory. While we do subscribe to the oft-quoted psychological considerations that motivate taking such preferences seriously, we also propose the present model as a particularly suitable one to explicitly account for the *fear of rejection* that a bargainer may have in case she makes an offer that favors herself disproportionately. It is not unreasonable to think of a bargainer A who does not propose an allocation that keeps the entire pie to herself as reasoning along the following lines: "If I offer an allocation that would

[12] There is abundant empirical evidence in support of this hypothesis. (See, for instance, Frank (1987) and Clark and Oswald (1996) and references cited therein.) The particular utility representation we use here is, in turn, axiomatically characterized by Ok and Koçkesen (2000). While simpler, this representation is similar to those employed by Bolton (1991) and Bolton and Ockenfels (2000).

leave B with a very small share of the pie, she may get upset and reject my offer. My share of the pie should then still be higher than hers, but not so large as to offend her." If we assume that the preferences of B are as in (1), it is easy to see that the gist of this reasoning will influence A's behavior when it is her turn to make a proposal. Introducing negatively interdependent preferences into bargaining models may thus be thought of as a non–ad hoc way of modelling (in reduced form) the notion of "fear of rejection," which commands considerable experimental support (Weg and Zwick, 1994; Bolton and Zwick, 1995). We stress that the fear of rejection is not introduced here as a primitive aspect of the utility functions. Instead, as we shall demonstrate shortly, it shows itself in the equilibrium behavior of the (possibly) negatively interdependent bargainers.

Surprisingly, there are only a few studies that investigate how the main insights of the theory of bargaining would be affected by the presence of negatively interdependent preferences. These studies, most notably Bolton (1991) and Kirchsteiger (1996), focus almost exclusively on finite-horizon complete-information models. Therefore, while these authors show that the idea of negatively interdependent preferences passes important preliminary tests in explaining experimental evidence, it is still not clear at this point what general insight can be gained by introducing such preferences into bargaining theory. As noted earlier, the aim of this article is to make use of the general utility specification given in (1) and to argue that the possibility of negative interdependence can explain the occurrence of disagreements, disadvantageous counteroffers, and the outcomes that come very close to the equal split of the pie.

□ **Degree of interdependence.** The connection between the notions of negative interdependence and fear of rejection suggests a natural index for the degree to which a particular type of player i is negatively interdependent: the minimum share of the pie that i (of type θ_i) must be given to be at least as well off as she would be at the disagreement outcome. This share, which we shall call the *reservation amount* $r_i(\theta_i)$, is uniquely determined by the equation

$$V_i\left(r_i(\theta_i), \frac{r_i(\theta_i)}{m} \mid \theta_i\right) \equiv 0.$$

Since $V_i(1, 1/m \mid 0) = V_i(1, 1 \mid 0) = 0$ by the normalization condition (2), the reservation of the independent type is zero, that is, $r_i(0) = 1$. Moreover, $r_i(\theta_i) \in (1, m)$ for any $\theta_i \in \mathcal{T}_i \backslash \{0\}$, since

$$V_i\left(1, \frac{1}{m} \mid \theta_i\right) < V_i\left(1, 1 \mid \theta_i\right) = V_i\left(r_i(\theta_i), \frac{r_i(\theta_i)}{m} \mid \theta_i\right) < V_i\left(m, 1 \mid \theta_i\right).$$

In what follows we wish to think of higher types as more competitive (negatively interdependent) players. Therefore, we invoke the following assumption that associates higher types θ_i with higher levels of reservation amounts:

Assumption 1. $r_i : \mathcal{T}_i \rightarrow [1, m)$ is a strictly increasing function, $i = A, B$.

Clearly, Assumption 1 relates the magnitude of each player's type to her level of interdependence in a natural way: if type θ_i of player i prefers the disagreement outcome $(1, 1)$ to an outcome $x = (x_A, x_B)$, then any type $\theta_i' > \theta_i$ strictly prefers $(1, 1)$ to x. Therefore, this assumption formalizes, in effect, the link between the preferences represented through (1) and the notion of fear of rejection.

3. One-period bargaining

■ In this section we examine briefly the classic ultimatum game with negatively interdependent preferences. The rules of the game are simple: player A proposes an allocation $x \in Y$, and player B either accepts or rejects A's offer. If player B accepts, the proposed allocation is realized. Otherwise, both players receive the disagreement outcome $(1, 1)$.

We take here that $\mathcal{T}_i = [0, \Theta_i]$ and assume that θ_B is privately known by player B. It is easy to see that the equilibrium in the ultimatum game does not depend on whether player A's type is private information or not. In turn, player A's beliefs about θ_B are represented by a continuous cumulative distribution function $F : \mathcal{T}_i \to [0, 1]$.[13]

To determine the subgame-perfect equilibria of the ultimatum game, we define the function τ_B as $\tau_B(x_B) \equiv r_B^{-1}(x_B)$ for all $x_B \in [1, r_B(\Theta_B)]$. This function specifies a critical level of interdependence $\tau_B(x_B)$ below which player B accepts and above which he rejects any given offer $(2m - x_B, x_B)$. Indeed, by Assumption 1, $\theta_B \lesseqgtr \tau_B(x_B)$ implies that $r_B(\theta_B) \lesseqgtr x_B$, so that $0 = V_B(r_B(\theta_B), r_B(\theta_B)/m \mid \theta_B) \lesseqgtr V_B(x_B, x_B/m \mid \theta_B)$. Consequently, in any subgame-perfect equilibrium, player A of type θ_A chooses $x_B \in [1, m)$ to maximize the following objective function:

$$V_A\left(2m - x_B, \frac{2m - x_B}{m} \mid \theta_A\right) F\left(\tau_B\left(x_B\right)\right). \tag{3}$$

We thus obtain the following result.

Proposition 1. Under Assumption 1 any perfect Bayesian equilibrium outcome of the ultimatum-bargaining game with negatively interdependent preferences has the following structure: Player A proposes a division $(2m - x_B(\theta_A), x_B(\theta_A))$, where $x_B(\theta_A) \in [1, m)$ maximizes the expression in (3). Player B accepts if her true type θ_B is strictly lower than the critical threshold $\tau_B(x_B(\theta_A))$, and rejects if $\tau_B(x_B(\theta_A)) < \theta_B$.

Proposition 1 (the analogs of which are also obtained by Daughety (1994), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000)) establishes that the two major "puzzling" regularities observed in ultimatum-bargaining experiments, namely interior offers and rejections, are consistent with game-theoretic equilibrium behavior. Once one accepts the idea that the competitive nature of bargainers plays a role in ultimatum experiments, it is realistic to assume that the exact degree of a player's interdependence is her private information. In this case, Proposition 1 shows that for generic probability distributions representing player A's beliefs about his opponent's preferences, all equilibrium offers must indeed be "more fair" than $(2m - 1, 1)$ but less fair than (m, m), and rejections occur with positive probability. The intuition is simply that, A knowing that B has possibly negatively interdependent preferences, fears rejection. Clearly, Proposition 1 reinforces the argument that negatively interdependent preferences can be thought of as an indirect way of modelling the presence of fear of rejection on the part of the first movers.

4. Two-period bargaining

■ In this section we investigate the two-period alternating-offer bargaining game in which players have negatively interdependent preferences. It may be argued that two-period bargaining models are not appealing because they do not correspond to any bargaining scenario that is played out in practice. However, such models are studied experimentally by several authors, and just as in the ultimatum-bargaining experiments, a number of striking regularities are identified. Before examining the implications of the interdependence hypothesis in the case of infinite-horizon models, therefore, it is worthwhile to check whether this hypothesis is successful in "explaining" these behavioral regularities. Our aim in this section is to carry out this task.

We focus on the standard Stahl bargaining model. The game is played in two periods. In the first period, player A proposes an allocation in Y that player B either accepts or rejects. In the former case, player A's proposal is realized, while in the latter case the game advances to the second period, in which the players switch roles: player B proposes an allocation (x_A, x_B) and player A either accepts or rejects it, with rejection resulting in disagreement.

---

[13] Continuity of F does not really play a significant role here. It is easy to modify the following analysis to account for discrete distributions.

Arguably, the most paradoxical regularity emerging from the experimental data about this game is that "a substantial percentage of rejected offers were followed by disadvantageous counterproposals" (Ochs and Roth, 1989, p. 376).[14] Interestingly, neither Stahl's original model with independent preferences nor our variation with negatively interdependent preferences can reconcile this evidence with equilibrium behavior. The reason is that the standard Stahl (1972) model posits that if B's proposal is accepted in the second period, player B of type θ_B achieves a utility level of $\delta_B u_B(x_A, x_B \mid \theta_B)$, where $\delta_B \in (0, 1)$ is player B's discount factor. Therefore, by (1), a necessary condition for B to reject A's initial offer (x_A, x_B) in equilibrium is

$$u_B(x_A, x_B \mid \theta_B) = V_B\left(x_B, \frac{x_B}{m} \mid \theta_B\right) \leq \delta_B V_B\left(y_B, \frac{y_B}{m} \mid \theta_B\right) = \delta_B u_B(y_A, y_B \mid \theta_B), \qquad (4)$$

where $(y_A, y_B) \in Y$ is the second-period equilibrium offer of type θ_B. But since sequential rationality ensures $y_B \geq r_B(\theta_B)$, i.e., $u_B(y_A, y_B \mid \theta_B) \geq 0$, the monotonicity of V_B and (4) jointly imply that $y_B \geq x_B$ for all $\delta_B \in [0, 1]$.

This argument, however, depends crucially on the assumption that individuals' utilities are discounted through periods. By contrast, in the experiments, the entire game lasts only a short time; it is rather the *size of the pie* that is discounted through the offer periods. Thus, we can reasonably assume that the subjects incur no cost due to the passage of real time, that is, if the game ends in agreement at the second stage, and the pie shrinks from size k to size δk, the utility of player i is $u_i(x_A, \delta_i k - x_A \mid \theta_i)$ instead of $\delta_i u_i(x_A, k - x_A \mid \theta_i)$.[15]

While inessential under the hypothesis of money-maximizing behavior, the different treatment of time in the standard theoretical model and the game played in the experiments turn out to be crucial for the presence of disadvantageous counteroffers in the present setup. To see why, let us assume that, as in the experiments, if B rejects A's offer, the game enters a second stage, in which the *net* pie shrinks from $2m - 2$ to $\delta(2m - 2)$, and B proposes a division in the set

$$Y(\delta) \equiv \{(y_A, y_B) \geq (1, 1) : y_A + y_B = 2 + \delta(2m - 2)\},$$

where $\delta \in (0, 1]$ stands for the *common* rate at which the pie is discounted.

For simplicity, consider the case in which the player i is either of type 0 or of type $\Theta_i > 0$, that is, $\mathcal{T}_i = \{0, \Theta_i\}$, $i = A, B$.[16] Recall that by Assumption 1 and (2), we have $r_i(\Theta_i) > r_i(0) = 1$. The beliefs of player i about player j are represented by a number $\pi_j \in [0, 1]$ interpreted as $\pi_j \equiv \text{prob}[\theta_j = \Theta_j]$. That is, π_A is the probability B assigns to the event A is negatively interdependent, and similarly for π_B. It is important to note that if the game enters the second stage, and B proposes $(y_A, y_B) \in Y(\delta)$, the utility of player i is $V_i(y_i, y_i/[1 + \delta(m - 1)] \mid \theta_i)$ if A accepts, and zero otherwise.

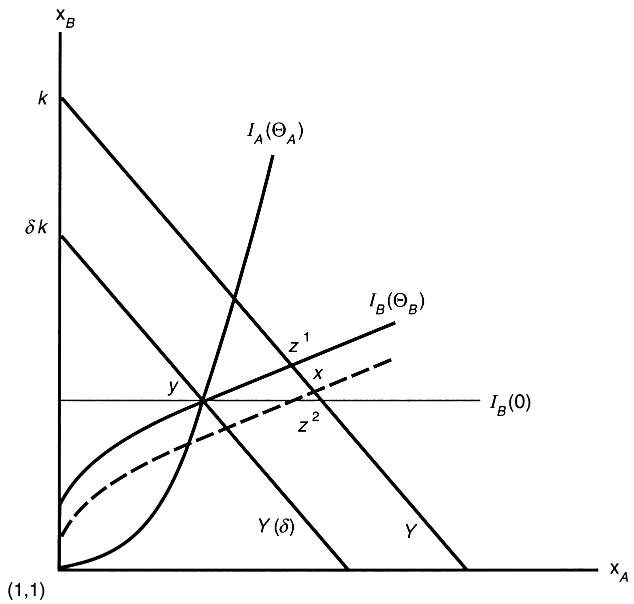
Of course, if $\pi_A = \pi_B = 0$, then the model collapses into a complete-information bargaining game with independent preferences. It is readily verified that in the unique subgame-perfect equilibrium of this game, player A proposes $(1 + (1 - \delta)(2m - 2), 1 + \delta(2m - 2))$ in the first period, and player B accepts. Following Bolton (1991), we shall refer to this allocation as the *pecuniary equilibrium* from now on. In turn, when $\pi_A, \pi_B > 0$, the model carries the basic features of a signalling game, thus admitting a great many perfect Bayesian equilibria. We shall show below that some of these equilibria are perfectly consistent with all the experimental regularities and, in particular, with the occurrence of disadvantageous counteroffers.

[14] To be precise, we note that the 125 first-offer rejections in the Ochs-Roth experiments were followed by 101 disadvantageous responses.

[15] While discounting the size of the pie is quite common in experimental studies, sometimes the designers have discounted the value of the share of a player (Ochs and Roth, 1989). In this case, if the game ends in agreement at the second stage, the utility of player i is $u_i(\delta_A(k - x_A), \delta_B(k - x_B) \mid \theta_i)$. The analysis of this section applies, without modification, to this case as well.

[16] The analysis of the model and the results reported below would remain true with inessential modifications, if we allowed for different discount rates and more than two types. We invoke these two assumptions here only to simplify the exposition.

FIGURE 2



Before stating the formal results, we provide a heuristic argument that shows how disadvantageous counteroffers may arise in equilibrium. Consider Figure 2, in which the bargaining sets Y and $Y(\delta)$ are plotted with some indifference curves for both players. The independent type of each player does not care about his opponent's share of the surplus: hence A's indifference curves are vertical if $\theta_A = 0$, and B's curves are horizontal if $\theta_B = 0$. The indifference curves of the negatively interdependent type of each player instead are upward sloping, since their utility decreases if their opponent's share increases. For instance, type Θ_A of player A is indifferent between receiving the disagreement outcome $(1, 1)$ and the allocation y. Similarly, type Θ_B of B is indifferent between the allocations y and z^1.

Now consider an allocation like x in Figure 2, and suppose that both types of A offer x in the first period. If B is pessimistic enough, i.e., π_A is high enough, then in any perfect Bayesian equilibrium, he will propose the division $y \in Y(\delta)$ in the second stage, whatever his level of interdependence, and both types of A would accept. Therefore, it will be optimal for type 0 of B to accept the proposal x. But is it optimal for type Θ_B of B to accept this offer when π_A is high? The answer is no, for if B rejects x and offers y in the second period, she is certain that both types of A will accept her offer. Since type Θ_B of B strictly prefers y to x (y is on a higher indifference curve than is x), type Θ_B will indeed reject A's offer x and counteroffer y in the second stage. But notice that y is a disadvantageous counteroffer for B; that is, $x_B > y_B$!

This argument suggests that disadvantageous counteroffers should not be considered paradoxical when we allow for potentially negatively interdependent preferences. Indeed, we have the following:

Proposition 2. Under Assumption 1, there exist $\pi_A, \pi_B \in (0, 1)$ such that, for all $\delta \in (0, 1)$, the two-period alternating-offer bargaining game described above admits a continuum of many pure-strategy perfect Bayesian equilibria that have the following features:

(i) both types of A make the same initial offer;

(ii) type 0 of B accepts while type Θ_B of B rejects and makes a disadvantageous counteroffer in the second period, which both types of A accept;[17]

[17] Daughety (1994) has anticipated this result by means of providing a concrete example.

(iii) there exists $\bar{\delta} \in [0, 1)$ such that A's initial offer deviates from the pecuniary equilibrium in the direction of 50-50 division for all $\delta \geq \bar{\delta}$; and

(iv) we have

$$\lim_{\delta \to 1} x(\delta, \theta) = (r_A(\Theta_A), 2m - r_A(\Theta_A)) \qquad \text{for all } \theta \in \{0, \Theta_A\} \times \{0, \Theta_B\},$$

where $x(\delta, \theta)$ denotes the equilibrium allocation at state θ.

Proof. See the Appendix.

Proposition 2 is interesting in that it shows that the bargaining model at hand is capable of predicting the rejection of first-period offers that are followed by disadvantageous counteroffers in equilibrium.[18] Moreover, this result shows that one should not be surprised to see opening offers in the experiments that deviate toward the 50-50 division when δ is sufficiently high. Of course, not all equilibria possess these properties; there are many other equilibria of the game than those mentioned in Proposition 2. The point is that when we model the bargaining experiments by using negatively interdependent preferences with one's degree of interdependence being private information, what seems like paradoxical play becomes perfectly reasonable equilibrium behavior. In a qualitative sense, then, we would like to argue that the present model fits the data fairly well.

Two objections can be raised at this point about the equilibria identified by Proposition 2. In these equilibria, all counteroffers are disadvantageous and are always accepted in the second stage. Indeed, experimental results show that this is not the case: both advantageous and disadvantageous counteroffers are observed, and both kinds are turned down, albeit infrequently, forcing the game to end in this agreement. We note that the only reason why Proposition 2 falls short of predicting these two additional features in equilibrium is our parsimonious way of modelling incomplete information here.

To drive this point home, consider the following slightly richer scenario in which player A has four types, $\theta_A^3 > \cdots > \theta_A^0 = 0$, and player B has three types, $\theta_B^2 = \theta_B^1 > \theta_B^0 = 0$. We assume that all four types of A have the same beliefs about B's type, while types θ_B^2 and θ_B^1 of B hold different beliefs. (So the beliefs, not the preferences, distinguish these types from each other.)[19] The beliefs of types θ_B^0 and θ_B^1 about A's type are given by a probability vector $\pi_A \equiv (\pi_A^0, \ldots, \pi_A^3)$, where $\pi_A^j \equiv \text{prob}[\theta_A = \theta_A^j]$. In turn, the beliefs of type θ_B^2 are given by another probability vector $\rho_A \equiv (\rho_A^0, \ldots, \rho_A^3)$, where ρ_A^j is the probability assigned by θ_B^2 to the event that A is of type θ_A^j. Finally, the beliefs of all types of A are represented by a probability vector $\pi_B \equiv (\pi_B^0, \pi_B^1, \pi_B^2)$, where $\pi_B^j \equiv \text{prob}[\theta_B = \theta_B^j]$. In this setup, we can improve Proposition 2 to the following:

Proposition 3. Under Assumption 1, there exist probability vectors π_A, ρ_A, and π_B such that for all $\delta \in (0, 1)$, the two-period alternating-offer bargaining game described above admits a continuum of many perfect Bayesian equilibria in which both advantageous and disadvantageous counteroffers are made and rejected with positive probability.

Proof. See the Appendix.

We note that the outcome-fairness properties of the equilibria mentioned in Propositions 2 and 3 are in line with the main thesis of the article. Negative interdependence endogenizes the fear of rejection a bargainer may have, and this, in turn, leads to relatively fair outcomes. Since the presence of discounting shadows the potential of making comparisons with the pecuniary equilibrium, however, here we can make such a comparison only for large δ. In this case, all

[18] As for belief-based refinement properties of these equilibria, we should note that they satisfy the intuitive criterion of Cho and Kreps (1987).

[19] The careful reader will notice that this formulation departs from the unidimensional way we have modelled the types of agents so far in the article. Indeed, apart from Proposition 3, all of our results are proved in a context where a "type" corresponds to a particular utility function, whereas for Proposition 3, we distinguish between two types on the basis of their beliefs. However, this is obviously a minor point. We have chosen here to model the type spaces as subsets of the real line in the general development only to achieve clarity of the exposition.

equilibria in Proposition 2 envisage a deviation of both the opening offers and the equilibrium allocations toward the 50-50 division. This is seen most clearly by comparing part (iv) of Proposition 2 with the fact that the pecuniary equilibrium converges to $(1, 2m - 1)$ as $\delta \to 1$.

In closing, let us pose the following important question: How serious is the possibility of observing disadvantageous counteroffers in the real world? The present model points out that this may depend on the nature of discounting. Recall that Propositions 2 and 3 are obtained under the assumption that time discounts the pie, not the utilities. So in bargaining contexts in which rejections result in the physical reduction of the size of the gains from trade (as would be the case if rejections could result in strikes), the occurrence of disadvantageous counteroffers may well be expected. By contrast, in a variety of actual bargaining situations, the passage of real time affects the utilities of the bargainers instead of the physical size of the gains from trade. As shown earlier, in this case disadvantageous offers cannot arise in any equilibrium of the present setup. This, in turn, yields a sharp prediction: In a finite-horizon bargaining environment with utility discounting (which would take place through real time), disadvantageous counteroffers will not be observed. It will be very interesting to evaluate the validity of this prediction through experiments that use real time by design.

5. Infinite-horizon bargaining

■ The findings of the previous two sections suggest that enriching bargaining games with the potential presence of negatively interdependent preferences can indeed be useful. However, while these findings allow us compare the predictions of the proposed models with those of the experiments, they do not apply to more realistic bargaining scenarios, with potentially infinite duration. It is important to study infinite-horizon bargaining games because in many situations that involve substantial gains from trade, it would be unrealistic to postulate that bargainers could credibly commit themselves not to trade just because a certain amount of time has elapsed.

Consequently, in this section we shall extend our treatment to the context of a standard infinite-horizon bargaining model, the so-called one-sided offers model with one-sided private information.[20] The basic primitive of this setting is the following *concession game*: player A, whose type θ_A is common knowledge, makes all the offers, while player B, who has private information about his type θ_B, can only accept or reject. If no offer is ever accepted, then the outcome of the game is the disagreement outcome, and if a proposal $x \in Y$ is accepted in period t, the utility of player i is $\delta^{t-1} u_i(x_A, x_B \mid \theta_i)$, where $\delta \in (0, 1)$ is the common discount rate.[21] As is standard, we assume in what follows that $\mathcal{T}_B = [0, \theta_B]$.

Unfortunately, the complexity of this game does not allow for a telling analysis in the case of the general specification of preferences as given in (1). For this reason, we shall adopt here the following simple one-parameter specification:

$$u_i\left(x_A, x_B \mid \theta_i\right) \equiv (1 - \theta_i)\, x_i + \theta_i \frac{x_i}{(x_A + x_B)/2} - 1, \qquad i = A, B, \tag{5}$$

where $(x_A, x_B) \in X$ and $0 \leq \theta_i < 1$. To see the usefulness of this specification (which is easily checked to satisfy Assumption 1), observe that since the equilibrium is invariant under linear transformations of u_i, we may instead take the utility function of individual i of type θ_i as

$$v_i\left(x_A, x_B \mid \theta_i\right) \equiv r(\theta_i)\, u_i\left(x_A, x_B \mid \theta_i\right),$$

[20] Interestingly, the qualitative nature of our findings remains valid in the classical Rubinstein (1982) model. That is, in this model as well, allowing for the possibility of interdependent preferences generates results consistent with the commonly observed high frequency of "fair divisions." For the formal analysis of the Rubinstein model with interdependent preferences, we refer the reader to Lopomo and Ok (1998).

[21] Again we assume that the players have the same discount factor for simplicity. Modifying the analysis in the case of distinct discount factors is straightforward.

where

$$r(\theta_i) \equiv (1 - \theta_i + \theta_i/m)^{-1} \in (1, m)$$

is the reservation amount of individual i of type θ_i. It follows from (5) that

$$v_i\left(x_A, x_B \mid \theta_i\right) = x_i - r(\theta_i) \qquad \text{for all } (x_A, x_B) \in Y.$$

Thus, given (5), the present model is essentially identical to the standard buyer-seller bargaining model. Indeed, if we think of player A as the owner of an indivisible object with value $r(\theta_A)$, and of player B as the potential "buyer" with value $2m - r(\theta_B)$, we can interpret any division $(x_A, 2m - x_A)$ in our model as a "sale" at price x_A. Conveniently, this allows us to bring to bear a number of major results established in the bargaining literature. While a similar correspondence between the models also exists in the case of specifications more general than (5), these lead us to buyer-seller models with nonlinear utilities, the analysis of which is well known to be extremely difficult. This is the main reason why we use here the linear specification given in (5).

As usual, player A's beliefs about B's type are modelled through a distribution function $F : [0, \Theta_B] \to [0, 1]$, where $0 < \Theta_B < 1$. We assume that F is common knowledge and posit the following technical condition:

Assumption 2. $\liminf_{\theta \to \Theta_B} \dfrac{1 - F(\theta)}{\Theta_B - \theta} > 0.$

This assumption is a very weak regularity condition that requires the slope of F to be bounded away from zero near Θ_B. For instance, if F is left-differentiable at Θ_B and $F'_{-}(\Theta_B) > 0$, or if F has a mass point on Θ_B, then it satisfies Assumption 2. Intuitively speaking, all the Assumption 2 requires is that player A assign strictly positive probability to the event that player B is arbitrarily close to being the most negatively interdependent type possible.

Given Assumption 2, it can be shown that a perfect Bayesian equilibrium of the game at hand exists. In fact, under this assumption, one equilibrium may differ from another only with respect to the first-period proposal of player A; the equilibrium is *generically* unique. What is more, as the time interval between offers gets very small (i.e., as the offers take place very quickly), all equilibrium first-period offers converge to a unique allocation:

Proposition 4. Given Assumption 2, consider any perfect Bayesian equilibrium of the concession game defined above, and denote the corresponding equilibrium sequence of offers made by player A (conditional on the discount rate and the type space of B by $\{x^t(\delta, \Theta_B)\}_{t=1}^{\infty} \in Y^{\infty}$. We have

$$\lim_{\delta \to 1} x^1(\delta, \Theta_B) = (2m - r(\Theta_B), r(\Theta_B)).$$

Proof. See the Appendix.

Proposition 4 says that if player A assigns positive probability, however small, to the event that B is maximally negatively interdependent, then, as the interval between offers becomes small, his strategy becomes close to the strategy that he would use were he certain that B's reservation amount is near $r(\Theta_B)$, that is, his first offer becomes close to the allocation that will be accepted even by the maximally negatively interdependent type, and the probability that the game ends immediately converges to one. In particular, if player A thinks that B may be the most competitive type who is almost indifferent between 50-50 division and the disagreement outcome (i.e., if Θ_B is close to 1), then the equilibrium outcome is (almost) the 50-50 split, no matter how unlikely A may think this event really is:

$$\lim_{\Theta_B \to 1} \lim_{\delta \to 1} x^1(\delta, \Theta_B) = (m, m).$$

This shows, again, the effectiveness of using negatively interdependent preferences to incorporate the reasoning of a proposer who fears rejection, which, in turn, forces fair outcomes. We would like to stress, however, that this is a deeper observation than the findings of the previous two sections. It may not be terribly surprising that spiteful preferences might lead to fair outcomes. But this result shows that in a variety of contexts, even the slightest doubt in the mind of the proposer about the excessively spiteful nature of the responder would force fair divisions to materialize with probability one. Very fair outcomes may thus be rational, after all.

Proposition 4 is essentially a corollary of a well-known result in bargaining theory. Indeed, as noted earlier, the concession game described above can be shown to correspond to the single-sale bargaining model with one-sided incomplete information. The limit result stated in Proposition 4 then corresponds to the famous Coase conjecture, which states that the seller's expected gain from trade tends to its lowest possible value when the frequency of price offers becomes arbitrarily large. But the validity of this conjecture (as a unique equilibrium outcome) depends on whether or not it is common knowledge that mutually beneficial agreements exist. Interestingly, here we have $r(\theta_A) + r(\Theta_B) < 2m$ for all $\theta_A \in [0, 1]$, so our model corresponds to the so-called gap case of the single-sale model (see Fudenberg and Tirole, 1991), in which beneficial trade exists with certainty. Consequently, the related results of Fudenberg, Levine, and Tirole (1985) and Gul, Sonnenschein, and Wilson (1986) that establish the uniqueness of equilibrium and the validity of the Coase conjecture in the gap case of the single-sale model entail our Proposition 4. (The details of this claim are found in the Appendix.)

In passing, we note that extending the present model to allow for two-sided offers and/or two-sided incomplete information introduces a vast multiplicity of equilibria. For instance, in any alternating-offers game with incomplete information, many allocations can be sustained as equilibrium outcomes, simply by making each player's beliefs concentrated on the independent type of his opponent off the equilibrium path. (Under these beliefs, the only constraint acting on the equilibrium strategies is the sequential rationality on the equilibrium path.) Nevertheless, by suitably refining the sequential equilibria, one may identify those equilibria that have the property that A's initial offer converges to $(2m - r(\Theta_B), r(\Theta_B))$ as $\delta \to 1$. Indeed, in both the alternating-offers game with one-sided private information (Gul and Sonnenschein, 1988), and in the concession game with two-sided private information (Cho, 1990), all equilibria that satisfy certain monotonicity and stationarity properties exhibit the Coase property.

6. Bargaining by arbitration

■ In this section we depart from the "positive" analysis of bargaining among negatively interdependent players and adopt a normative approach. This is of interest because in many instances of real-world bargaining, negotiations are brought to an end by a settlement guided by a third party. A natural question, then, is this: If an arbitrator realizes that players may possess negatively interdependent preferences, how should she resolve the pie-division problem? More specifically, we wish to see if there is any reason to suspect that the potential interdependence of preferences would enforce the emergence of egalitarian outcomes in bargaining settlements. To address this issue, we adopt a mechanism-design approach.

As in the previous sections, we consider $\mathcal{T}_i \subseteq [0, \Theta_i]$ as the type space of individual i for some $\Theta_i > 0, i = A, B$, that is, $\mathcal{T}_i$ stands for the set of all preferences that the arbitrator conceives as admissible for individual i. The product $\mathcal{T} \equiv \mathcal{T}_A \times \mathcal{T}_B$ then corresponds to the set of all states of nature.

A social-choice function is any function that assigns to a state of nature a particular division of the pie. Formally, we define a social-choice function (SCF) on $\mathcal{T}$ as any function $f : \mathcal{T} \to Y$. Thus, $f(\theta_A, \theta_B)$ stands for the (efficient) division of the pie that the arbitrator would choose, had she known that the true utility (type) of player i was θ_i.[22] As is usual, we say that f is individually rational if f does not allocate to any type a share that is strictly below its reservation amount, i.e.,

[22] Alternatively, we may define a SCF on $\mathcal{T}$ as mapping a type profile (θ_A, θ_B) to a 3-tuple $(x, t) \in Y \times \mathbf{Z}_+$, where (x, t) denotes the allocation awarded to players at period t. In this case, our definition of a SCF would postulate implicitly the property of *ex post* efficiency.

if

$$u_i(f(\theta_A, \theta_B) \mid \theta_i) \geq 0 \qquad \text{for all } \theta_i \in \mathcal{T}_i,\ i = A, B.$$

Needless to say, individual rationality is a participation constraint that needs to be satisfied by any reasonable settlement.

We now ask if it is possible here to implement a SCF in dominant strategies.[23] While it is well known that this is in general impossible (the Gibbard-Satterthwaite theorem), it is easily seen that the restricted domain we consider here admits individually rational SCFs that are dominant strategy implementable. For instance, the SCF that assigns to *any* type profile the equal split of the pie is individually rational and dominant strategy implementable. Our next result provides a characterization of all such SCFs.

Proposition 5. A SCF f on $\mathcal{T}$ is individually rational and dominant strategy implementable if and only if

$$f(\theta_A, \theta_B) \geq \left(\sup_{\theta_A \in \mathcal{T}_A} r(\theta_A),\ \sup_{\theta_B \in \mathcal{T}_B} r(\theta_B) \right) \qquad \text{for all } (\theta_A, \theta_B) \in \mathcal{T}.$$

Proof. See the Appendix.

Proposition 5 shows that if the domain of a SCF includes highly negatively interdependent preferences, then that SCF must choose highly egalitarian outcomes at all states of nature. For instance, if $\mathcal{T}_i = [0, \Theta_i]$, Assumption 1 holds, and $r_i(\Theta_i) = m$, then Proposition 5 yields that the only SCF f on $\mathcal{T}$ that is individually rational and dominant strategy implementable is the constant function defined as

$$f(\theta_A, \theta_B) = (m, m) \qquad \text{for all } (\theta_A, \theta_B) \in \mathcal{T}.$$

Put differently, if the class of all negatively interdependent preferences considered by the arbitrator is sufficiently rich, then the only possible SCF is the one that assigns the 50-50 division regardless of the state of nature.[24] This observation provides a rigorous normative rationale for the equal-split solution, the applications of which arise abundantly in daily life. For instance, consider a parent who has to divide the last slice of the pie among two siblings. It would not be unreasonable to expect that the parent will divide the pie equally (to the best of her abilities) to avoid any possible conflict between the siblings that may arise due to envious feelings. Proposition 5 suggests a rigorous foundation precisely for this decision rule. The intuition is the same one that underlies all of our results. Interdependence entails a fear of rejection, in the present case on the part of the arbitrator through the individual-rationality constraints. This, in turn, forces a fair settlement.

7. Concluding comments

■ Admitting the possibility of negatively interdependent preferences may not be the only way one can modify standard bargaining models to obtain predictions consistent with experimental evidence. Nevertheless, the combination of negatively interdependent preferences and private information does provide a non–ad hoc way of modelling the notion of fear of rejection that is likely to be present in the actual strategic reasoning of a bargainer. We have shown here that this, in turn, yields a useful model that is capable of accommodating the observed experimental regularities in bargaining games and provides a rational theory of fair outcomes. Moreover, the idea of negatively interdependent preferences, which is already in use in other areas of economics,

[23] A mechanism is any list $(\{Z_A, Z_B\}, h)$ where Z_i is an arbitrary message space and $h : Z_A \times Z_B \rightarrow Y$ is an arbitrary outcome function. Such a mechanism is said to implement the SCF f on $\mathcal{T}$ in dominant strategies if and only if the two-person normal-form game $(\{Z_i, u_i(h(\cdot \mid \theta_i)\}_{i=A,B})$ has a dominant strategy equilibrium $z(\theta_A, \theta_B) \in Z_A \times Z_B$ such that $h(z(\theta_A, \theta_B)) = f(\theta_A, \theta_B)$, for all $(\theta_A, \theta_B) \in \mathcal{T}$.

[24] One certainly does not need the type space of player i to be "too large" for this result to hold. If, for instance, $\mathcal{T}_i$ contains all of the affine utility functions considered in Section 5, then the result goes through.

seems particularly appropriate in bilateral bargaining contexts. Indeed, it is hardly unreasonable to think that competitive feelings may arise and influence the players' choices in such situations.

Having said this, however, let us note that we have considered in this article only negative interdependence, ignoring completely the possibility of positive interdependence. While this modelling strategy is useful in focusing on matters related to fear of rejection, it falls short of letting us cover in a sensible way many interesting games that are not of the bilateral bargaining form. For instance, if the dictator game were played by negatively interdependent players, the unique outcome would be none other than the dictator sparing the entire surplus to himself. Since in the related experiments more than half of the individuals are observed to make interior divisions (with several 50-50 splits), one may argue that such experiments refute the negative interdependence hypothesis.

The validity of this conclusion depends on what one really means by the "negative interdependence hypothesis." If this hypothesis is taken to suggest that individuals behave in all of their strategic (or otherwise) encounters under the guidelines of spite, then clearly it cannot be taken seriously; the experimental findings shoot down this universal interpretation with ease. From this global perspective, then, the present contribution is meaningful only in a *ceteris paribus* sense. A complete model that would yield good predictions in nonbargaining environments as well should presumably model individuals in a way that includes both positive *and* negative interdependence.[25]

On the other hand, if one subscribes to the negative interdependence hypothesis, as we do, in a narrower sense that applies only with regard to a certain "class" of games (namely, the bargaining games), the conceptual standpoint of our contribution is stronger. From this viewpoint, the basic hypothesis acts as a useful means to model the competitive nature of the bargainers that may well surface in face-to-face bargaining situations in which each party may hurt the other. This in turn yields a good number of theoretical results that may provide a basis for a rational theory of fair outcomes.[26]

Provided that one adopts this narrower viewpoint, then—and admittedly only then—the dictator game does not provide a suitable testing ground, precisely because there is no second-mover in this game; the modelled scenario does not really correspond to a bargaining environment. As noted earlier, we view negatively interdependent preferences as a means to model the competitive nature of the bargainers that may surface in face-to-face bargaining situations. The only upshot of the present article is to suggest formally that modelling agents as negatively interdependent enriches the standard model by including, endogenously, the notion of fear of rejection. Since this notion is absent in the dictatorship game, there is no reason to even apply the negatively interdependent model to this nonstrategic situation. Our approach makes sense only in contexts where fear of rejection is likely to be inherent to one's strategic behavior, as in bargaining games: nothing more, nothing less. Consequently, we contend that the right experimental test of the negative interdependence hypothesis must be conducted in a bargaining setting. Whether such tests will eventually falsify the theory we advance here is a question that should be addressed in future experimental work.

Even if one agrees with our interpretation of the negative interdependence hypothesis, there are several other directions for future research. First, it will be interesting to design experiments to identify the predictive limitations of the finite-horizon bargaining models examined here. Second, the infinite-horizon model with nonlinear negatively interdependent utility functions remains to be investigated. This case is difficult since, unlike the model of Section 5, it is not necessarily isomorphic to the standard buyer-seller bargaining model. In addition, the way the negatively interdependent agents' attitudes toward risk influence the bargaining outcomes needs to be further

[25] Daughety (1994), for instance, provides precisely this sort of a model.

[26] There is a caveat here. If the modeller is free to choose preferences in a game-dependent manner, then the model clearly loses much of its predictive power. However, this does not mean that one "must" therefore focus on universal models of preferences. There is a compromise approach in which one examines the suitability of certain preference structures for certain *classes* of games. Provided that these classes are large enough, one may have sufficient predictive power along with a tractable model of decision-making. It is this strategy that we follow in the present article by focusing on the large class of all bilateral bargaining games.

explored. Third, one should study how the alternative interdependence hypotheses (based on "positive and negative interdependence" as in Daughety (1994), on "inequality aversion" as in Fehr and Schmidt (1999), or on "equity focus" as in Bolton and Ockenfels (2000)) would alter the results reported here. Finally, the analysis of multilateral bargaining with possibly negatively interdependent preferences is left for future research.

Appendix

■ This Appendix contains the proofs of the main results reported in the text.

Proof of Proposition 2. Let k denote the size of the pie net of the bargainers' holdings, that is, set $k \equiv 2m - 2$. Since V is strictly increasing and continuous, (1) and $V(1, 1 \mid \Theta_A) = 0$ imply that there exists a unique $\alpha \in (0, 1)$ such that

$$u_A \left(1 + (1 - \alpha)\, \delta k,\, 1 + \alpha \delta k \mid \Theta_A\right) = 0.$$

Clearly, α is the maximum share for B that both types of A will accept. Throughout this proof we shall let

$$(y_A, y_B) \equiv (1 + (1 - \alpha)\delta k, 1 + \alpha\delta k)\,.$$

We define next the functions $\omega : \{0, \Theta_B\} \to (0, 1)$ and $\hat{\omega} : \{0, \Theta_B\} \to (0, 1)$ by

$$u_B \left(1 + (1 - \omega(\theta_B))\, k,\, 1 + \omega(\theta_B)\, k \mid \theta_B\right) = u_B \left(y_A, y_B \mid \theta_B\right) \tag{A1}$$

and

$$u_B \left(1 + (1 - \hat{\omega}(\theta_B))\, k,\, 1 + \hat{\omega}(\theta_B)\, k \mid \theta_B\right) = u_B \left(1, 1 + \delta k \mid \theta_B\right), \tag{A2}$$

respectively. One can think of $\omega(\theta_B)$ as the minimum share for B that type θ_B will accept in stage 1, provided that B has sufficiently pessimistic beliefs about A's type (i.e., π_A is high enough). In contrast, $\hat{\omega}(\theta_B)$ is the minimum share for B that type θ_B will accept in stage 1, provided that B is sufficiently optimistic about A's type.

Claim A1. $\alpha\delta = \omega(0) < \omega(\Theta_B)$ and $\delta = \hat{\omega}(0) < \hat{\omega}(\Theta_B)$.

Proof. Since $u_B(\cdot, \cdot \mid 0)$ is independent of its first argument, equation (A1) readily yields $\omega(0) = \alpha\delta$. But then, by (1) and (A1),

$$V\left(1 + \omega(\Theta_B)\, k,\, \frac{1}{m}\, (1 + \omega(\Theta_B)\, k) \mid \Theta_B\right) = V\left(1 + \omega(0)\, k,\, \frac{1}{1 + \delta(m - 1)}\, (1 + \omega(0)\, k) \mid \Theta_B\right).$$

Since $m > 1 + \delta(m - 1)$ for all $m > 1$ and $\delta \in [0, 1)$, by strict monotonicity of V, we must then have $\omega(\Theta_B) > \omega(0)$. The second part of the claim is proved similarly. *Q.E.D.*

Claim A2. There exists an $\omega^* \in (\omega(0), \min\{\omega(\Theta_B), \hat{\omega}(0)\})$ such that

$$u_A \left(1 + \left(1 - \omega^*\right) k,\, 1 + \omega^* k \mid \Theta_A\right) > 0.$$

Proof. Since $1 - \alpha\delta > (1 - \alpha)\delta$ and $\omega(0) = \alpha\delta$, by definition of α, we have

$$\begin{aligned} u_A(1 + (1 - \omega(0))k, 1 + \omega(0)k \mid \Theta_A) &= u_A(1 + (1 - \alpha\delta)k, 1 + \alpha\delta k \mid \Theta_A) \\ &> u_A(1 + (1 - \alpha)\delta k, 1 + \alpha\delta k \mid \Theta_A) \\ &= 0. \end{aligned}$$

So, since $\min\{\omega(\Theta_B), \hat{\omega}(0)\} > \omega(0)$ by Claim A1, the result follows by continuity. *Q.E.D.*

Take any $\omega_0 \in (\omega(0), \omega^*)$, where ω^* is as found in Claim A2. We propose the following assessment as a candidate for a pure-strategy perfect Bayesian equilibrium.

Strategy of A. Both types of A offer

$$(x_A, x_B) \equiv (1 + (1 - \omega_0)k, 1 + \omega_0 k) \in Y$$

in stage 1. In case of rejection, each type of A responds optimally to B 's offer in stage 2, that is, type 0 accepts any feasible offer $z \in Y(\delta)$, while type Θ_A accepts $z = (z_A, z_B) \in Y(\delta)$ if $z_A \geq r_A(\Theta_A)$ and rejects if $z_A < r_A(\Theta_A)$.

Strategy of B. If A offers $(1 + (1 - \omega)k, 1 + \omega k)$ in stage 1, then type 0 of B

$$\begin{cases} \text{accepts} & \text{if } \omega = \omega_0 \text{ or if } \omega_0 \neq \omega > \hat{\omega}(0), \\ \text{rejects and proposes } (1, 1 + \delta k) & \text{if } \omega_0 \neq \omega \leq \hat{\omega}(0) \end{cases}$$

and type Θ_B of B

$$\begin{cases} \text{rejects and proposes } (y_A, y_B) & \text{if } \omega = \omega_0, \\ \text{accepts} & \text{if } \omega_0 \neq \omega > \hat{\omega}(\Theta_B), \\ \text{rejects and proposes } (1, 1 + \delta k) & \text{if } \omega_0 \neq \omega \leq \hat{\omega}(\Theta_B). \end{cases}$$

Beliefs. After any off-equilibrium offer $z \neq (x_A, x_B)$, the beliefs of both types of B are degenerate on $\theta_A = 0$.[27] On the equilibrium path, i.e., after A offers (x_A, x_B), Bayes' rule applies, hence B still believes that $\theta_A = \Theta_A$ with probability π_A.

In what follows we shall show that there exist $\pi_A, \pi_B \in (0, 1)$ such that the above assessment is a perfect Bayesian equilibrium. It is readily verified that this will complete the proof of Proposition 4. In particular, notice that the above assessment specifies a disadvantageous counteroffer for B on its equilibrium path since $\alpha\delta = \omega(0) < \omega_0$.

To establish sequential rationality, take any $\pi_A \in (0, 1)$ such that

$$\pi_A > 1 - \max_{\theta_B \in \{0, \Theta_B\}} \frac{u_B\left(y_A, y_B \mid \theta_B\right)}{u_B\left(1, 1 + \delta k \mid \theta_B\right)}. \tag{A3}$$

Given the strategy of player A in the second stage, the decision problem of type θ_B of B in the second stage is

$$\max_{\gamma \in [\alpha, 1]} \begin{cases} (1 - \pi_A) u_B(1 + (1 - \gamma)\delta k, 1 + \gamma \delta k \mid \theta_B) & \text{if } \gamma \in (\alpha, 1], \\ u_B(y_A, y_B \mid \theta_B) & \text{if } \gamma = \alpha. \end{cases}$$

Therefore, (A3) ensures that the optimal offer of both types of B is $(y_A, y_B) \in Y(\delta)$ in the second stage.

Sequential rationality for B. Since $\alpha\delta = \omega(0) < \omega_0$, we have

$$u_B\left(1 + (1 - \alpha)\delta k, 1 + \alpha\delta k \mid 0\right) < u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid 0\right).$$

Thus, accepting A's offer is optimal for type 0 of B. On the other hand, by (A1) and since $\omega_0 < \omega(\Theta_B)$,

$$\begin{aligned} u_B(y_A, y_B) &= u_B\left(1 + (1 - \omega(\Theta_B))k, 1 + \omega(\Theta_B)k \mid \Theta_B\right) \\ &> u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \Theta_B\right), \end{aligned}$$

so that it is optimal for type Θ_B of B to reject A's offer and to counteroffer (y_A, y_B) in the second stage.

It is also clear that B's strategy is sequentially rational for both types $\theta_B \in \{0, \Theta_B\}$ after any off-equilibrium offer $z \neq (x_A, x_B)$, since the beliefs about A's type become degenerate on $\theta_A = 0$.

Sequential rationality for A. We need to show that offering $(x_A, x_B) \equiv (1 + (1 - \omega_0)k, 1 + \omega_0 k)$ in stage 1 is optimal for both types of A. Suppose A proposes $(1 + (1 - \omega')k, 1 + \omega' k)$, where $\omega_0 \neq \omega' \in [0, 1]$. Then, given the belief structure, if $\omega' < \hat{\omega}(0)$, the offer is rejected and A is offered $(1, 1 + \delta k)$ in stage 2. Thus, both types of A earn zero utility in this case. But since $\omega_0 < \omega^*$, by Claim A2 above we have $u_A(x_A, x_B \mid \theta_A) > 0$ for any $\theta_A \in \{0, \Theta_A\}$; moreover, $u_A(y_A, y_B \mid \theta_A) \geq 0$ for any $\theta_A \in \{0, \Theta_A\}$. Thus both types of A obtain strictly positive expected utility on the equilibrium path.

Suppose now that $\omega' \in [\hat{\omega}(0), \hat{\omega}(\Theta_B))$. In this case, type 0 of player B accepts A's offer, while type Θ_B rejects and counteroffers $(1, 1 + \delta k)$ in the second stage; hence A's expected utility from offering $\omega' \in [\hat{\omega}(0), \hat{\omega}(\Theta_B))$ is $(1 - \pi_B)u_A(1 + (1 - \omega')k, 1 + \omega' k \mid \theta_A)$. Since $\omega_0 < \omega^* < \hat{\omega}(0) \leq \omega'$ and $u_A(y_A, y_B \mid \theta_A) \geq 0$, we have

$$(1 - \pi_B) u_A\left(x_A, x_B \mid \theta_A\right) + \pi_B u_A\left(y_A, y_B \mid \theta_A\right) > (1 - \pi_B) u_A\left(1 + \left(1 - \omega'\right)k, 1 + \omega' k \mid \theta_A\right)$$

for any $\theta_A \in \{0, \Theta_A\}$. Hence, given the belief structure, offering $(1 + (1 - \omega')k, 1 + \omega' k)$ is not a profitable deviation for either type of A in this case.

[27] This degenerate specification of beliefs is not necessary, but convenient. All we need to require here is that B's off-equilibrium beliefs be sufficiently optimistic.

Finally, assume that $\omega' \geq \hat{\omega}(\Theta_B)$. In this case, both types of B accept A's offer. Therefore, to complete the proof we need to show that

$$(1-\pi_B)\, u_A\left(x_A, x_B \mid \theta_A\right) + \pi_B u_A\left(y_A, y_B \mid \theta_A\right) \geq u_A\left(1 + \left(1-\omega'\right)k, 1 + \omega' k \mid \theta_A\right) \tag{A4}$$

holds for any $\theta_A \in \{0, \Theta_A\}$. Since $\omega' > \omega_0$, we again have

$$u_A\left(x_A, x_B \mid \theta_A\right) > u_A\left(1 + (1-\omega')k, 1 + \omega' k \mid \theta_A\right).$$

Thus, if $(1-\alpha)\delta \geq 1-\omega'$, then $u_A(y_A, y_B \mid \theta_A) \geq u_A(1+(1-\omega')k, 1+\omega' k \mid \theta_A)$ and (A4) must hold for all $\theta_A \in \{0, \Theta_A\}$.[28] Assume then that $(1-\alpha)\delta < 1-\omega'$, so that $u_A(y_A, y_B \mid 0) < u_A(1+(1-\omega')k, 1+\omega' k \mid 0)$. There are two possibilities to consider: (i) $u_A(y_A, y_B \mid \Theta_A) < u_A(1+(1-\omega')k, 1+\omega' k \mid \Theta_A)$, and (ii) otherwise. In case (i), (A4) is established for all θ_A by choosing any $\pi_B \in (0, 1)$ such that

$$\pi_B \leq \min_{\theta_A \in \{0, \Theta_A\}} \frac{u_A(x_A, x_B \mid \theta_A) - u_A(1+(1-\hat{\omega}(\Theta_B))k, 1+\hat{\omega}(\Theta_B)k \mid \theta_A)}{u_A(x_A, x_B \mid \theta_A) - u_A(y_A, y_B \mid \theta_A)},$$

since $\omega' \geq \hat{\omega}(\Theta_B)$. In case (ii), on the other hand, the proof is completed upon choosing any $\pi_B \in (0, 1)$ such that

$$\pi_B \leq \frac{u_A(x_A, x_B \mid 0) - u_A(1+(1-\hat{\omega}(\Theta_B))k, 1+\hat{\omega}(\Theta_B)k \mid 0)}{u_A(x_A, x_B \mid 0) - u_A(y_A, y_B \mid 0)}.$$

Q.E.D.

Proof of Proposition 3. Proceeding as in the proof of Proposition 2, define α_j through the equation

$$u_A\left(1 + \left(1-\alpha_j\right)\delta k, 1 + \alpha_j \delta k \mid \theta_A^j\right) = 0 \tag{A5}$$

for $j = 0, \ldots, 3$. Clearly, α_j is the maximum share of the net pie that B can get in the second period if A's type is θ_A^j. Letting $(y_A^j, y_B^j) \equiv (1+(1-\alpha_j)\delta k, 1+\alpha_j \delta k)$; for $j = 0, \ldots, 3$, we have

$$u_B(y_A^2, y_B^2 \mid \theta_B)\left[1-\pi_A^3\right] \geq \max\left\{u_B(y_A^j, y_B^j \mid \theta_B)\sum_{i=0}^{j}\pi_A^i : j = 0, \ldots, 3\right\}, \qquad \theta_B = \theta_B^0, \theta_B^1 \tag{A6}$$

and

$$u_B\left(y_A^1, y_B^1 \mid \theta_B^2\right)\left[\rho_A^0 + \rho_A^1\right] \geq \max\left\{u_B\left(y_A^j, y_B^j \mid \theta_B^2\right)\sum_{i=0}^{j}\rho_A^i : j = 0, \ldots, 3\right\}, \tag{A7}$$

provided that ρ_A^1 and π_A^2 are sufficiently close to one. Thus, if A's initial proposal does not reveal any additional information about θ_A, it is optimal in the second stage for the type θ_B^2 of B to propose (y_A^1, y_B^1) and for θ_B^0 and θ_B^1 to propose (y_A^2, y_B^2).

The first-period reservation share $\omega(\theta_B^j)$ of type θ_B^j is then defined by

$$u_B\left(1 + \left(1-\omega(\theta_B^j)\right)k,\ 1 + \omega(\theta_B^j)k | \theta_B^j\right) = u_B\left(y_A^2, y_B^2 | \theta_B^j\right), \qquad j = 0, 1,$$

whereas

$$u_B\left(1 + \left(1-\omega(\theta_B^2)\right)k, 1 + \omega(\theta_B^2)k \mid \theta_B^2\right) = u_B\left(y_A^1, y_B^1 \mid \theta_B^2\right).$$

Finally, we define the off-equilibrium reservation shares $\widehat{\omega}(\theta_B^j)$, $j = 0, 1, 2$, by

$$u_B\left(1 + \left(1-\widehat{\omega}(\theta_B^j)\right)k, 1 + \widehat{\omega}(\theta_B^j)k \mid \theta_B^j\right) = u_B\left(1, 1+\delta k \mid \theta_B^j\right).$$

Clearly, $\widehat{\omega}(\theta_B^j)$ is the minimum share that θ_B^j accepts if she is convinced that A is independent.

The following two claims can be established by mimicking the steps of Claims A1 and A2 in the proof of Proposition 2.

[28] Therefore, for high δ we do not have to put any restriction on the beliefs of A. In particular, for such δ, we may let $\pi_A = \pi_B$.

Claim A3. $\alpha_2\delta = \omega(\theta_B^0) < \omega(\theta_B^1)$, and $\delta = \widehat{\omega}(\theta_B^0) < \widehat{\omega}(\theta_B^1)$.

Claim A4. There exists ω^* such that $\omega(\theta_B^0) < \omega^* < \min\{\omega(\theta_B^1), \widehat{\omega}(\theta_B^0), \delta\alpha_1\}$ and

$$u_A\left(1 + \left(1 - \omega^*\right)k, 1 + \omega^* k \mid \theta_A^j\right) > 0, \qquad \text{for each } j = 1, 2, 3.$$

Take any $\omega_0 \in (\omega(\theta_B^0), \omega^*)$, where ω^* is as found in Claim A4. We propose the following assessment as a candidate for a perfect Bayesian equilibrium.

Strategy of A. Each type θ_A^j, $j = 0, \ldots, 3$, proposes $(1 + (1 - \omega_0)k, 1 + \omega_0 k)$ in the first stage, and, in case of rejection, accepts $(1 + (1 - \alpha)\delta k, 1 + \alpha\delta k)$ if and only if $\alpha \leq \alpha_j$, where α^j is as defined in (A5).

Strategy of B. If A proposes $(1 + (1 - \omega_0)k, 1 + \omega_0 k)$ in stage 1, then type θ_B of B

$$\begin{cases} \text{accepts} & \text{if } \theta_B = \theta_B^0, \\ \text{rejects and proposes } (y_A^2, y_B^2) & \text{if } \theta_B = \theta_B^1, \\ \text{rejects and proposes } (y_A^1, y_B^1) & \text{if } \theta_B = \theta_B^2. \end{cases}$$

If A's initial offer is different from $(1 + (1 - \omega_0)k, 1 + \omega_0 k)$, then all types of B reject and counteroffer $(1, 1 + \delta k)$.

Beliefs. After any off-equilibrium initial offer, the beliefs of all types of B are degenerate on $\theta_A = \theta_A^0$. On the equilibrium path, i.e., after A proposes $(1+(1-\omega_0)k, 1+\omega_0 k)$, each type of B keeps his beliefs unchanged in accordance with Bayes' law.

Note that (y_A^2, y_B^2) is a disadvantageous counteroffer, since $y_B^2 = 1+\omega(\theta_B^0)k < 1+\omega_0 k$, while (y_A^1, y_B^1) is advantageous since $1 + \omega_0 k < 1 + \omega^* k < 1 + \alpha_1\delta k = y_A^1$.

We now show that the assessment described above constitutes a perfect Bayesian equilibrium if ρ_A^1, π_A^2, and π_B^0 are sufficiently close to one. As in the proof of Proposition 2, this amounts to verifying the sequential rationality of each player's strategy given their opponent's strategy and the beliefs specified above.

Sequential rationality for B. The independent type of B accepts $(1+(1-\omega_0)k, 1+\omega_0 k)$ because $u_B(1+(1-\omega_0)k, 1+\omega_0 k \mid \theta_B^0) > u_B(y_A^2, y_B^2 \mid \theta_B^0)$ implies

$$u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \theta_B^0\right) > u_B\left(a_2, Y - a_2 \mid \theta_B^0\right)\left[1 - \pi_A^3\right].$$

Type θ_B^1 rejects and counteroffers (y_A^2, y_B^2), because

$$u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \theta_B^1\right) < u_B\left(y_A^2, y_B^2 \mid \theta_B^1\right)$$

and hence

$$u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \theta_B^1\right) < u_B\left(y_A^2, y_B^2 \mid \theta_B^1\right)\left[1 - \pi_A^3\right],$$

provided that π_A^2 is sufficiently close to one. Finally, type θ_B^2 rejects and counteroffers (y_A^1, y_B^1) since

$$u_B\left(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \theta_B^1\right) < u_B\left(y_A^1, y_B^1 \mid \theta_B^1\right)\left[\rho_A^0 + \rho_A^1\right],$$

provided that ρ_A^1 is sufficiently close to one.

Sequential rationality for A. Recall that we have chosen the out-of-equilibrium beliefs so that if A's first offer is $(1 + (1 - \omega)k, 1 + \omega k)$ with $\omega \neq \omega_0$, then all types of B become convinced that A is independent. Therefore A's expected utility is at most

$$\begin{cases} u_A((1 + (1 - \widehat{\omega}(\theta_B^0))k, 1 + \widehat{\omega}(\theta_B^0)k) \mid \theta_A) & \text{if } \widehat{\omega}(\theta_B^0) \leq \omega \leq 1, \\ \pi_B^0 u_A(1 + (1 - \omega_0)k, 1 + \omega_0 k \mid \theta_A) & \text{if } \widehat{\omega}(\theta_B^1) \leq \omega < \widehat{\omega}(\theta_B^0), \\ 0 & \text{if } 0 \leq \omega \leq \widehat{\omega}(\theta_B^1). \end{cases}$$

Consequently, type θ_A^j of A would have no incentive to deviate from $(1 + (1 - \omega_0)\,k,\, 1 + \omega_0 k)$, provided that

$$\pi_B^0 \geq \frac{u_A\left(\left(1 + \left(1 - \widehat{\omega}(\theta_B^0)\right)k,\, 1 + \widehat{\omega}(\theta_B^0)k\right) \mid \theta_A^j\right)}{u_A\left(1 + (1 - \omega_0)\,k,\, 1 + \omega_0 k \mid \theta_A^j\right)}, \qquad \text{for each } j = 0, \ldots, 3.$$

This completes the proof. *Q.E.D.*

Proof of Proposition 4. Let $r(\theta) \equiv (1 - \theta + \theta/m)^{-1}$ and $b(\theta) \equiv 2m - r(\theta)$ for all $\theta \in [0, \Theta_B)$. Define

$$v_A(x \mid \theta_A) \equiv r(\theta_A)u_A(x \mid \theta_A) \quad \text{and} \quad v_B(x \mid \theta_B) \equiv b(\theta_B)u_B(x \mid \theta_B)$$

for all $x \in X$. It follows from (5) that

$$v_A(x \mid \theta_A) = x_A - r(\theta_A) \quad \text{and} \quad v_B(x \mid \theta_B) \equiv b(\theta_B) - x_B \qquad \text{for all } x \in Y.$$

Let $\underline{b} \equiv b(\Theta_B) > m$ and $\overline{b} \equiv b(0) = 2m - 1 > \underline{b}$. Define next the distribution function $P : [\underline{b}, \overline{b}] \to [0, 1]$ as $P \equiv 1 - (F \circ b^{-1})$. P is the distribution function of the random variable $b(\cdot)$:

$$P(b) = \text{prob}(b(\theta) \leq b) = 1 - F(b^{-1}(b)), \qquad \underline{b} \leq b \leq \overline{b}.$$

The equilibrium outcomes of our model are, therefore, equivalent to those of the standard single-sale model in which the seller with the known cost $r(\theta_A)$ makes (all) offers to the buyer whose valuation b is distributed on the interval $[\underline{b}, \overline{b}]$ according to the distribution function P (see Fudenberg and Tirole, 1991). Moreover, since $\Theta_B < 1$, this bargaining model corresponds to the gap case: for all $\theta_A \in [0, 1]$,

$$r(\theta_A) \leq m < 2m - r(\theta_B) = \underline{b}.$$

Consequently, if we can show that P^{-1} is Lipschitz continuous at zero, we may then apply Theorem 3 (and Remark 6.2) of Gul, Sonnenschein, and Wilson (1986) to complete the proof. To see this, we use Assumption 2 to find a $\theta^* \in (0, \Theta_B)$ and $K_0 > 0$ such that $1 - F(\theta) > K_0(\Theta_B - \theta)$ for all $\theta \in [\theta^*, \Theta_B)$. Since r is easily checked to be convex, we thus have

$$1 - F(\theta) > K_0(\Theta_B - \theta) = K_1 r'(\Theta_B)(\Theta_B - \theta) \geq K_1(r(\Theta_B) - r(\theta)), \qquad \theta^* \leq \theta < \Theta_B,$$

where $K_1 \equiv K_0/r'(\Theta_B)$. Therefore,

$$1 - F(b^{-1}(b(\theta))) > K_1(b(\theta) - b(\Theta_B)), \qquad \theta^* \leq \theta < \Theta_B,$$

that is, $P(b) > K_1(b - \underline{b})$, $\underline{b} < b \leq b^* \equiv b(\theta^*)$. Since $P(\underline{b}) = 0$, we have $q > K_1(P^{-1}(q) - P^{-1}(0))$ for all $q \in (0, P(b^*)]$. Letting $q^* \equiv P(b^*)$ and $K \equiv 1/K_1$, therefore, we find

$$\left|P^{-1}(0) - P^{-1}(q)\right| < Kq, \qquad 0 < q \leq q^*,$$

that is, P^{-1} is Lipschitz continuous at zero. *Q.E.D.*

Proof of Proposition 5. Suppose that f on $\mathcal{T}$ is individually rational and implementable in dominant strategies. By the revelation principle, there must then exist a direct mechanism that truthfully implements f. This implies that we must have

$$u_i(f(\theta_A, \theta_B) \mid \theta_i) \geq u_i(f(\theta_i^*, \theta_{-i}) \mid \theta_i) \qquad \text{for all } \theta_i, \theta_i^* \in \mathcal{T}_i \text{ and } \theta_{-i} \in \mathcal{T}_{-i}$$

for any $i = A, B$, and therefore

$$V_i\left(f_i(\theta_A, \theta_B),\, \frac{f_i(\theta_A, \theta_B)}{m} \mid \theta_i\right) \geq V_i\left(f_i(\theta_i^*, \theta_{-i}),\, \frac{f_i(\theta_i^*, \theta_{-i})}{m} \mid \theta_i\right)$$

for all $\theta_i, \theta_i^* \in \mathcal{T}_i$ and $\theta_{-i} \in \mathcal{T}_{-i}$. Thus, by monotonicity of V_i,

$$f_i(\theta_A, \theta_B) \geq f_i(\theta_i^*, \theta_{-i}) \qquad \text{for all } \theta_i, \theta_i^* \in \mathcal{T}_i \text{ and } \theta_{-i} \in \mathcal{T}_{-i}.$$

Notice that f_i must then be independent of its ith component, that is, $f_i(\theta, \theta_{-i}) = f_i(\theta', \theta_{-i})$ for any $\theta, \theta' \in \mathcal{T}_i$ and any $\theta_{-i} \in \mathcal{T}_{-i}$. Consequently, we may write $f_i(\theta_A, \theta_B) = \varphi_i(\theta_{-i})$ for all $(\theta_A, \theta_B) \in \mathcal{T}$ for some function $\varphi_i : \mathcal{T}_{-i} \to Y$. But by individual rationality, we must have $\varphi_i(\theta_{-i}) \geq r(\theta_i)$ so that $f_i(\theta_A, \theta_B) = \varphi_i(\theta_{-i}) \geq \sup_{\theta_i \in \mathcal{T}_i} r(\theta_i)$ for all $(\theta_A, \theta_B) \in \mathcal{T}$, which completes the proof of the "only if" part. The validity of the "if" part of the proposition is self-evident. *Q.E.D.*

References

ANDREONI, J. AND MILLER, J. "Giving According to GARP: An Experimental Study of Rationality and Altruism." SSRI Working Paper no. 9601, Department of Economics, University of Wisconsin, 1996.

BOLTON, G. "A Comparative Model of Bargaining: Theory and Evidence." *American Economic Review*, Vol. 81 (1991), pp. 1096–1136.

———. "The Rationality of Splitting Equally." *Journal of Economic Behavior and Organization*, Vol. 32 (1997), pp. 365–381.

——— AND OCKENFELS, A. "E.R.C.: A Theory of Equity, Reciprocity and Competition." *American Economic Review*, Vol. 90 (2000), pp. 166–193.

——— AND ZWICK, R. "Anonymity Versus Punishment in Ultimatum Bargaining." *Games and Economic Behavior*, Vol. 10 (1995), pp. 95–121.

CAMERER, C.F. AND THALER, R.H. "Ultimatums, Dictators and Manners." *Journal of Economic Perspectives*, Vol. 9 (1995), pp. 209–219.

CHO, I.-K. "Uncertainty and Delay in Bargaining." *Review of Economic Studies*, Vol. 57 (1990), pp. 575–596.

——— AND KREPS, D. "Signalling Games and Stable Equilibria." *Quarterly Journal of Economics*, Vol. 102 (1987), pp. 179–221.

CLARK, A.E. AND OSWALD, A.J. "Satisfaction and Comparison of Income." *Journal of Public Economics*, Vol. 61 (1996), pp. 359-381.

DAUGHETY, A. "Socially-Influenced Choice: Equity Considerations in Models of Consumer Choice and in Games." Mimeo, Department of Economics, University of Iowa, 1994.

ELLINGSEN, T. "The Evolution of Bargaining Behavior." *Quarterly Journal of Economics*, Vol. 112 (1997), pp. 581–602.

FEHR, E. AND SCHMIDT, K.M. "A Theory of Fairness, Competition and Cooperation." *Quarterly Journal of Economics*, Vol. 114 (1999), pp. 817–868.

FORSYTHE, R., HOROWITZ, J.L., SAVIN, N.E., AND SEFTON, M. "Fairness in Simple Bargaining Experiments." *Games and Economic Behavior*, Vol. 6 (1994), pp. 347–369.

FRANK, R. *Choosing the Right Pond*. New York: Oxford University Press, 1987.

FUDENBERG, D. AND TIROLE, J. *Game Theory*. Cambridge, Mass.: MIT Press, 1991.

———, LEVINE, D., AND TIROLE, J. "Infinite-Horizon Models of Bargaining with One-Sided Incomplete Information." In A.E. Roth, ed., *Game-Theoretic Models of Bargaining*. New York: Cambridge University Press, 1985.

GUL, F. AND SONNENSCHEIN, H. "On Delay in Bargaining with One-Sided Uncertainty." *Econometrica*, Vol. 56 (1988), pp. 601–611.

———, ———, AND WILSON, R. "Foundations of Dynamic Monopoly and the Coase Conjecture." *Journal of Economic Theory*, Vol. 39 (1986), pp. 155–190.

GÜTH, W., SCHMITTBERGER, R., AND SCHWARZE, B. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior and Organization*, Vol. 3 (1982), pp. 367–388.

KENNAN, J. AND WILSON, R. "Bargaining with Private Information." *Journal of Economic Literature*, Vol. 31 (1993), pp. 45–104.

KIRCHSTEIGER, G. "The Role of Envy in Ultimatum Games." *Journal of Economic Behavior and Organization*, Vol. 25 (1994), pp. 373–389.

LEVINE, D.K. "Modeling Altruism and Spitefulness in Experiments." *Review of Economic Dynamics*, Vol. 1 (1998), pp. 593–622.

LOPOMO, G. AND OK, E.A. "Bargaining, Interdependence, and the Rationality of Fair Division." C.V. Starr Center Working Paper no. 98-19, Department of Economics, New York University, 1998.

OCHS, J. AND ROTH, A.E. "An Experimental Study of Sequential Bargaining." *American Economic Review*, Vol. 79 (1989), pp. 355–384.

OK, E.A. AND KOÇKESEN, L. "Negatively Interdependent Preferences." *Social Choice and Welfare*, Vol. 17 (2000), pp. 533–558.

RABIN, M. "Endogeneous Preferences in Games." *American Economic Review*, Vol. 83 (1993), pp. 1281–1302.

ROTH, A.E. "Bargaining Experiments." In J.H. Kagel and A.E. Roth, eds., *Handbook of Experimental Economics*. Princeton: Princeton University Press, 1995.

RUBINSTEIN, A. "Perfect Equilibrium in a Bargaining Model." *Econometrica*, Vol. 50 (1982), pp. 97–109.

STAHL, I. *Bargaining Theory*. Stockholm: Stockholm School of Economics, 1972.

WEG, E. AND ZWICK, R. "Toward the Settlement of the Fairness Issues in Ultimatum Games." *Journal of Economic Behavior and Organization*, Vol. 24 (1994), pp. 19–34.

YOUNG, H.P. "An Evolutionary Model of Bargaining." *Journal of Economic Theory*, Vol. 59 (1993), pp. 145–168.