MANAGEMENT SCIENCE
Vol. 53, No. 11, November 2007, pp. 1785–1792
ISSN 0025-1909 | EISSN 1526-5501 | 07 | 5311 | 1785

informs®
DOI 10.1287/mnsc.1070.0736


# Final-Offer Arbitration and Risk Aversion in Bargaining

Eran Hanany
Department of Industrial Engineering, Tel Aviv University, Tel Aviv 69978, Israel,
hananye@post.tau.ac.il

D. Marc Kilgour
Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario N2L 3C5, Canada,
mkilgour@wlu.ca

Yigal Gerchak
Department of Industrial Engineering, Tel Aviv University, Tel Aviv 69978, Israel,
ygerchak@eng.tau.ac.il

Negotiations are often conducted under the stipulation that an impasse is to be resolved using final-offer arbitration (FOA). In fact, FOA frequently is not needed; in Major League Baseball, for instance, more than 80% of the salary negotiations that could go to arbitration instead reach a bargained agreement. We show that the risk aversion of at least one side explains this phenomenon. We then model pay negotiation in baseball by applying a bargaining solution with a variable disagreement outcome representing FOA, studying the existence of pure Nash equilibrium initial offers and their effects on the player's eventual pay, and considering the Nash solution as a special case.

*Key words*: games and group decisions; bargaining; risk aversion; final-offer arbitration; Nash equilibrium
*History*: Accepted by David E. Bell, decision analysis; received June 21, 2005. This paper was with the authors 3 months for 1 revision. Published online in *Articles in Advance* September 21, 2007.

## 1. Introduction

Institutional arrangements under which binding arbitration ends labor-management and other two-sided disputes are now common. An arbitrated settlement is imposed if no agreement is reached by a specified date. Often, this settlement is determined by final-offer arbitration, a procedure that is widely believed to induce bargainers to adopt "reasonable" positions. We study how anticipation of this form of arbitration affects prior bargaining.

Final-offer arbitration (FOA) was first proposed by Stevens in 1966. Under FOA, the arbitrator, or arbitration panel, must choose either the proposal of one side or the proposal of the other; averaging or blending the two sides' positions is not permitted. FOA is now an adjunct to bargaining in the determination of wages of public-sector employees, settlements in workers' compensation cases, and salaries of Major League Baseball (MLB) players who are not yet eligible for free agency (see Commission on Health and Safety and Workers' Compensation 1999, a general bibliography on FOA, and Major League Baseball Players Association 2006, the MLB basic collective agreement for 2007–2011). Because MLB is highly visible and information is readily available, we use the baseball setting for concreteness. (Indeed, in workers' compensation and other domains, FOA is called "baseball arbitration.")

Baseball's experience with FOA has produced an intriguing observation: Of the players who file for arbitration and exchange offers with their teams, only about 15%–20% end up in arbitration—the remainder reach a negotiated agreement (Miller 2005). The stability of this statistic has led many commentators to suggest that the prospect of FOA must induce bargainers to settle (e.g., Gardner 1995). This view is supported by the experimental findings of Manzini and Mariotti (2001) that the availability of arbitration, even if it requires mutual consent, affects negotiation outcomes dramatically.

But exactly how does the possibility of FOA rationally affect bargaining strategy? We model the FOA outcome as determined by the parties' offers and the realization of a random variable representing the arbitrator's preferred settlement.[1] Thus, we ask what offers are best for the parties, given that the same

[1] We assume that each party has the same distribution over the arbitrator's preferred settlement; i.e., the parties have identical information about the arbitrator. This one-dimensional model was apparently initiated by Crawford (1982). A model with differing beliefs could be constructed, based for example on Myerson

offers will be input into FOA if there is no bargaining agreement.

An important insight of Nash (1950) is that the outcome of bargaining depends crucially on the disagreement point. If FOA is the "default" outcome of bargaining, the disagreement point can be thought of as random.[2] It is not difficult to see the effects of this arrangement if both sides are risk-neutral. No matter what the probability that FOA selects one position or the other, any settlement that is better than going to FOA for one party must be worse for the other. Thus, under risk-neutrality, FOA gives the bargainers no extra incentive to reach an agreement.

But things are different if one party, or both, is risk-averse. Several authors (including Wittman 1986, Curry and Pecorino 1993, and Kilgour 1994, sec. 4) argue that at least one party—in the baseball context, the player—should be modeled as risk-averse, because for him the arbitration is an extraordinary, perhaps once-in-a-lifetime, event. These studies focus on whether a more risk-averse party wins FOA more often, and at what cost. Our interest is different: We explore the consequences of imposing a formal negotiation stage between the initial offers and FOA; negotiation might produce an agreement, and then the arbitration stage would be redundant.

Part of our explanation of why FOA is not invoked very often is that risk aversion implies that there are bargains the parties could reach that would be Pareto improvements over FOA. Later, we show how such a bargain could be in equilibrium. We use a model that opens with a risk-averse player stating his demand while, simultaneously, a (possibly risk-averse) team states its offer. Were the parties actually to reach FOA, their expected utilities would depend on their offers as elaborated in the literature.[3] We first show formally, and illustrate, that for all likely pairs of initial offers—including Nash equilibrium offers—there is a range of settlements preferred by both sides to FOA. In other words, FOA creates an incentive to reach a bargained agreement. Then, assuming that bargaining is resolved according to a reasonable general solution, we obtain bounds on pure-strategy equilibrium offers. Finally, we suppose that the bargaining solution is the Nash solution (Nash 1950) and derive comparative statics.

Our model is related to, but different from, earlier work. Nash (1953) introduces the *Nash threat game*, a noncooperative game model of the selection of a disagreement point in a cooperative game. This game is equivalent to a zero-sum game, a fact that Nash uses to find "optimal threat strategies."[4] Adapting the Nash Threat Game to FOA, Crawford (1982) analyzes how utilities are related to arbitrator characteristics at equilibrium.

Our model is different from the Nash threat game and from Crawford's model of FOA and bargaining, but closer to the latter. The fact that disagreement payoffs are not linearly related to strategies reflects the FOA structure of the inner game. Because mixed-strategy offers are difficult to interpret, we focus on pure-strategy equilibria and the settlements they imply. The analysis is more intricate because we must deal directly with nonlinearities. Thus, our model addresses new issues of theory; as well, our conclusions inform a central aspect of an important application.

## 2. Risk Aversion and Incentives to Bargain

In our model, the only issue is the annual pay of a player, $P$, who is under contract to a team, $T$. Suppose that $T$ offers $a$ and $P$ demands $b$, and that $P$'s preference is described by a utility function $u_P(\cdot)$. (Later we introduce a utility for $T$ as well.) Now suppose that FOA is carried out by an arbitrator whose assessment of $P$'s market value is unknown, and that both $T$ and $P$ believe that this assessment follows a continuous cumulative distribution function, $F$. If $a < b$, the arbitrator will choose the closer of $a$ and $b$ to the realized value of $F$. Then $T$'s expected wealth is equal to a constant plus

$$d_T(a,b) \equiv (-a)F\left(\frac{a+b}{2}\right) + (-b)\left[1 - F\left(\frac{a+b}{2}\right)\right], \quad (1)$$

while $P$'s expected utility is

$$d_P(a,b) \equiv u_P(a)F\left(\frac{a+b}{2}\right) + u_P(b)\left[1 - F\left(\frac{a+b}{2}\right)\right]. \quad (2)$$

If $a \geq b$ then $P$'s pay is set at the midpoint of $a$ and $b$, so $d_T(a,b) = -(a+b)/2$ and $d_P(a,b) = u_P((a+b)/2)$. To simplify notation, we write $d_T$ instead of $d_T(a,b)$, etc.

We now assume that $P$ is risk-averse (and, of course, nonsatiated) and ask whether, given $a$ and $b$ with $a < b$, there is a compromise, $x$, that both parties

(1984), but because offers would reflect differential information this model would be less likely to explain why agreements are so common. Besides, different beliefs seem unlikely in the baseball context, where relevant information (playing statistics and salaries) is widely available (Greenstein et al. 2004). Other models of arbitrator behavior, e.g., Farmer and Pecorino (1998), do not seem to lend themselves to this kind of analysis.

[2] Bargaining with a random disagreement point appears in Chun and Thompson (1982), but there the randomness has no strategic aspect, whereas here it reflects the bargainers' positions.

[3] For example, see Kilgour's analysis of FOA with risk-averse parties in Kilgour (1994, sec. 4).

[4] On the Nash threat game, see Luce and Raiffa (1957, pp. 140–143) and Myerson (1991, pp. 385–389).

prefer to FOA. Both parties would be better off if $P$'s pay were equal to such a *mutually improving* value $x$.

To be specific, our assumptions on $u_P$ are that $u'_P > 0$ and $u''_P < 0$. We are searching for values of $x$ such that $-x > d_T$, which is equivalent to

$$u_P(x) < u_P(-d_T), \tag{3}$$

because $u_P$ is strictly increasing, and

$$u_P(x) > d_P. \tag{4}$$

It is evident that there are values of $x$ that satisfy both (3) and (4) if and only if

$$d_P < u_P(-d_T). \tag{5}$$

But, as (1) and (2) make clear, $-d_T$ is the expected value of the lottery representing FOA, so $u_P(-d_T)$ is the utility of that expected value, while $d_P$ is $P$'s expected utility for that lottery. Jensen's inequality applies because $u_P$ is concave so (5) holds provided $0 < F((a+b)/2) < 1$, i.e., the lottery is nondegenerate. We conclude that the interval $(u_P^{-1}(d_P), -d_T)$ is nonempty and that every $x$ within it satisfies (3) and (4). Thus, if one party is risk-averse there are mutually improving settlements, which are preferred by both parties to FOA.

If *both* parties are risk-averse, $P$ with utility function $u_P$ and $T$ with utility function $u_T$, then mutually improving settlements $x$ satisfy two conditions,

$$u_T(-x) > u_T(-a)F\left(\frac{a+b}{2}\right) + u_T(-b)\left[1 - F\left(\frac{a+b}{2}\right)\right], \tag{6}$$

$$u_P(x) > u_P(a)F\left(\frac{a+b}{2}\right) + u_P(b)\left[1 - F\left(\frac{a+b}{2}\right)\right]. \tag{7}$$

Define the function $g$ by $g(\alpha) \equiv u_P[-u_T^{-1}(-\alpha)]$. Then $g$ is increasing and concave, and $u_P(x) = g[-u_T(-x)]$. Now (6) and (7) can be rewritten as

$$u_P(x) < g\left\{-u_T(-a)F\left(\frac{a+b}{2}\right) - u_T(-b)\left[1 - F\left(\frac{a+b}{2}\right)\right]\right\} \tag{8}$$

$$u_P(x) > g\{-u_T(-a)\}F\left(\frac{a+b}{2}\right) + g\{-u_T(-b)\}\left[1 - F\left(\frac{a+b}{2}\right)\right]. \tag{9}$$

It follows that a mutually improving settlement, $x$, exists iff the right side of (9) is less than the right side of (8), which can occur only if $0 < F((a+b)/2) < 1$. The required relation again follows from Jensen's inequality, which applies because $g$ is concave. As above, the mutually improving settlements form a nonempty open interval.[5]

We illustrate with an example in which only the player, $P$, is risk-averse.

Example 1.

$$F \sim \text{Uniform}(10, 14),$$
$$u_P(x) = \sqrt{x},\ 20 - b \le a < b \le 28 - a.$$

Applying (1), the condition $-x > d_T$ is equivalent to

$$x < a\left(\frac{a+b-20}{8}\right) + b\left[1 - \left(\frac{a+b-20}{8}\right)\right].$$

Using (2) and (4), the second condition is

$$\sqrt{x} > \sqrt{a}\left(\frac{a+b-20}{8}\right) + \sqrt{b}\left[1 - \left(\frac{a+b-20}{8}\right)\right].$$

Our argument above shows that it is always possible to find values of $x$ that satisfy both inequalities.

To illustrate, we have solved for and tabulated the intervals of mutually improving settlements for various $(a, b)$ pairs in Example 1:

| $a$ | $b$ | Mutually improving settlements |
|---|---|---|
| 10 | 14 | $11.916 < x < 12$ |
| 11 | 13 | $11.979 < x < 12$ |
| 11 | 14 | $12.083 < x < 12.125$ |
| 10 | 13 | $11.829 < x < 11.875$ |

For instance, if $T$ offers to pay $P$ 10 but $P$ demands 14, both parties would prefer to settle for any pay level strictly between 11.916 and 12, rather than go on to FOA. In this sense, adjoining FOA to bargaining creates incentives to reach agreement; how such agreements might unfold is the topic we address next.

## 3. Bargaining Followed by FOA

Define the bargaining FOA (BFOA) game as a two-person noncooperative game $\langle N, (S_T, S_P), (v_T, v_P)\rangle$ with player set $N = \{T, P\}$ and strategy sets $S_T = S_P = \mathbb{R}$. As usual, $T$'s strategy is a pay offer, $a$, and $P$'s strategy is a pay demand, $b$. The payoffs $v_T(a, b)$ and $v_P(a, b)$ are the players' utilities at the outcome of a bargaining problem $\langle C, d(a, b)\rangle$ in which the disagreement point is the FOA outcome, $d(a, b) = (d_T(a, b), d_P(a, b))$, where $d_P(a, b)$ is defined by (2), and

$$d_T(a, b) \equiv u_T(-a)F\left(\frac{a+b}{2}\right) + u_T(-b)\left[1 - F\left(\frac{a+b}{2}\right)\right].$$

The feasible set of the bargaining problem is

$$C = \{(Eu_T(-l), Eu_P(l)) \colon l \in \mathcal{L}\},$$

[5] Note that this result holds whether $P$'s risk aversion is greater than, equal to, or less than $T$'s.

where $\mathscr{L}$ is the set of all simple (finite) money lotteries, and $Eu_i(l)$ is player $i$'s expected utility for lottery $l \in \mathscr{L}$, for $i = T, P$. Thus, the BFOA game represents bargaining between $T$ and $P$, with any disagreement to be resolved by FOA.

We assume that $u_T$ and $u_P$ are twice continuously differentiable and strictly increasing. We also assume that the arbitrator's distribution $F$ is known to both players, is continuously differentiable, and has compact support; we denote the support of $F$ by $[D, U] \subseteq \mathbb{R}$, where $D$ is the maximum solution of $F(D) = 0$ and $U$ is the minimum solution of $F(U) = 1$. We also assume that $F$ is strictly increasing on $[D, U]$ and denote by $f$ the corresponding density function.

Of course, the BFOA game is not fully specified until we indicate how the outcome of the bargaining problem, $\langle C, d(a, b)\rangle$, is to be determined. First, analogous to Nash (1951), we make minimal assumptions about the resolution of bargaining and study the resulting Nash equilibria.

Assume that the bargaining outcome is always efficient and strictly dominates the FOA (disagreement) outcome whenever it is possible to do so. Assume further that the team $T$ and the player $P$ are risk-averse, so that an efficient outcome is a degenerate lottery, i.e., a sure pay level for $P$. Then the offers $a$ and $b$ determine $x(a, b)$, and the payoffs $v_T(a, b)$ and $v_P(a, b)$ of the BFOA game $\langle N, (S_T, S_P), (v_T, v_P)\rangle$ are

$$v_T(a, b) = u_T[-x(a, b)]$$
$$v_P(a, b) = u_P[x(a, b)].$$

In particular, the BFOA game is strictly competitive, and all Nash equilibria (if any) are equivalent in the sense that $T$'s payoffs at every equilibrium are equal, and so are $P$'s. Note that, if the FOA outcome $d(a, b)$ is itself efficient (which occurs whenever the choices $a$ and $b$ make the FOA lottery, based on $F$, degenerate), then $x(a, b)$ is simply defined by $u_P[x(a, b)] = d_P(a, b)$.

We now consider the BFOA game in the more typical case when the FOA outcome, $d(a, b)$, is inefficient. Assume that whenever $a < b$ and $D < (a + b)/2 < U$ the resolution of the bargaining problem $\langle C, d(a, b)\rangle$ is $x(a, b) = \bar{x}[d_T(a, b), d_P(a, b)]$, where the bargaining solution $\bar{x}(d_T, d_P)$ satisfies $O(d_T, d_P, \bar{x}) = 0$ for some real function $O$. Assume further that $O$ has continuous derivatives satisfying[6]

Condition C1. $\partial O/\partial d_T > 0$, $\partial O/\partial d_P < 0$ and $\partial O/\partial \bar{x} > 0$.

Provided all derivatives are well defined, we have $\partial x/\partial a = (\partial \bar{x}/\partial d_T)(\partial d_T/\partial a) + (\partial \bar{x}/\partial d_P)(\partial d_P/\partial a)$ and $\partial x/\partial b = (\partial \bar{x}/\partial d_T)(\partial d_T/\partial b) + (\partial \bar{x}/\partial d_P)(\partial d_P/\partial b)$, where $\partial \bar{x}/\partial d_T = -(\partial O/\partial d_T)/(\partial O/\partial \bar{x})$ and $\partial \bar{x}/\partial d_P = -(\partial O/\partial d_P)/(\partial O/\partial \bar{x})$. It therefore follows from Condition C1 that $\partial \bar{x}/\partial d_T < 0$ and $\partial \bar{x}/\partial d_P > 0$ and hence that $\partial x/\partial a$ and $\partial x/\partial b$ exist and are continuous whenever $a < b$ and $D < (a + b)/2 < U$.

We can now establish some important properties of the BFOA game.[7]

Proposition 1. *If Condition* C1 *holds in the BFOA Game, then*

(P1-1) *For any $a$ and $b$, $x_T(b) \equiv \min_{a'} x(a', b)$ and $x_P(a) \equiv \max_{b'}(a, b')$ exist and are continuous.*

(P1-2) $\max_b x_T(b)$ *and* $\min_a x_P(a)$ *both exist.*

*Moreover, if $(a^*, b^*)$ is a Nash equilibrium of the BFOA game, then*

(P1-3) $D \le \max\{a^*, 2D - a^*\} \le x(a^*, b^*) \le \min\{b^*, 2U - b^*\} \le U$ *and* $a^* < x(a^*, b^*) < b^*$.

(P1-4) $a^* < b^*$ *and* $D < (a^* + b^*)/2 < U$.

(P1-5) $(\partial x/\partial a)(a^*, b^*) = (\partial x/\partial b)(a^*, b^*) = 0$.

It follows from (P1-3) and (P1-4) that, at any Nash equilibrium $(a^*, b^*)$, $D < x(a^*, b^*) < U$, $D - (U - D) < a^* < U$, and $D < b^* < U + (U - D)$. Note that $a^* < D$ is possible, as is $b^* > U$.

Proposition 1 demonstrates many properties that a Nash equilibrium must have in the BFOA game, but it does not establish that such equilibria exist. In general, existence is an issue; many games, even with well behaved, continuous payoffs, have no pure-strategy Nash equilibrium.[8] We now introduce Condition C2, a mild "partial" concavity condition that guarantees that each player has a unique best response to every strategy of the opponent.

Condition C2. For $a < b$ and $D < (a + b)/2 < U$, if $\partial x/\partial a = 0$, then $\partial^2 x/\partial a^2 > 0$ and if $\partial x/\partial b = 0$ then $\partial^2 x/\partial b^2 < 0$.

The next proposition shows that Conditions C1 and C2 together are sufficient for the existence of a unique Nash equilibrium, which is a saddle point because the game is strictly competitive.

Proposition 2. *If Conditions* C1 *and* C2 *both hold, then the BFOA game has a unique Nash equilibrium.*

If $F$ represents a uniform distribution, so that $f(c) = 1/(U - D)$ for $c \in [D, U]$, then we can give sufficient conditions for C2 in terms of the Arrow–Pratt measure of risk aversion for player $i$, defined by $R_i(z) \equiv -u_i''(z)/u_i'(z)$.

[6] Our intention is to discuss a broad class of bargaining solutions without regard to explicit functional forms. But note that we demonstrate below that the Nash bargaining solution satisfies all of our assumptions, including Condition C1. In this condition, the opposite directions of the inequalities for $T$ and $P$ imply that each player gains from improvement in the disagreement outcome, a requirement that reflects the strictly competitive nature of the game.

[7] Proofs of propositions are in the appendix.

[8] For example, the strictly competitive game with payoff functions $(a - b)^2$ and $-(a - b)^2$ has no pure-strategy Nash equilibrium.

Proposition 3. *Suppose that Condition* C1 *holds and that F represents a uniform distribution. Then Condition* C2 *is also satisfied provided that*

$$\text{C2-1:}\quad \frac{\partial^2 O}{\partial (d_T)^2}=\frac{\partial^2 O}{\partial (d_P)^2}=\frac{\partial^2 O}{\partial d_T \partial d_P}=0, \quad \textit{and}$$

$$\text{C2-2:}\quad R_T(z)<\frac{1}{U-D}\quad \textit{for } z<-D \quad \textit{and}$$

$$R_P(z)<\frac{1}{U-D}\quad \textit{for } z<U.$$

Below, we show that Condition C2-1 holds for the Nash bargaining solution.[9]

Note that neither a uniform arbitrator's distribution nor the two conditions of Proposition 3 are necessary for the existence or uniqueness of Nash equilibrium. For instance, the next example has a unique Nash equilibrium even though all three of these conditions are violated.

Example 2. Suppose that the arbitrator's distribution is

$$F(c)=\left(\frac{c-10}{4}\right)^2 \quad \text{for } 10\le c\le 14.$$

Thus, the support of the distribution is $[10,14]$ and larger values have higher probabilities. Suppose further that the bargaining outcome equals the average certainty equivalent of the disagreement outcome values for both players, which is equivalent to

$$O(\bar{x},d_T,d_P)\equiv \tfrac{1}{2}u_T^{-1}(d_T)-\tfrac{1}{2}u_P^{-1}(d_P)+\bar{x}.$$

Finally, assume that $u_T(z)=z$ and $u_P(z)=\sqrt{z}$ as in Example 1, so that $\bar{x}=\frac{1}{2}(-d_T)+\frac{1}{2}(d_P)^2$. The unique solution of $\partial x/\partial a=\partial x/\partial b=0$ that also satisfies $a<b$ and $D<(a+b)/2<U$ is $(a^*,b^*)=(10.69,12.116)$, where $x(a^*,b^*)=11.938$. This is also the unique Nash equilibrium.[10]

## 4. An Illustration: The Nash Bargaining Solution

To sharpen some of the results obtained above, we now assume that, for any $a$ and $b$ with $a<b$, the bargaining problem $\langle C,d(a,b)\rangle$ is resolved using the Nash bargaining solution (Nash 1950).[11] Thus, the outcome is the player's pay level that maximizes the product of the improvements in the parties' utilities relative to their expected utilities under FOA. For the team, this improvement is $u_T(-\bar{x})-d_T\ge 0$; for the player, it is $u_P(\bar{x})-d_P\ge 0$. The Nash solution is therefore defined by

$$\bar{x}(d_T,d_P)=\arg\max_y\{[u_T(-y)-d_T][u_P(y)-d_P]:\; u_T(-y)\ge d_T,\, u_P(y)\ge d_P\}. \quad (10)$$

The second derivative of the maximand in (10),

$$-2u_T'(-\bar{x})u_P'(\bar{x})+u_T''(-\bar{x})[u_P(\bar{x})-d_P]+u_P''(\bar{x})[u_T(-\bar{x})-d_T],$$

is negative for concave $u_T$ and $u_P$, provided that $u_T(-\bar{x})\ge d_T$ and $u_P(\bar{x})\ge d_P$. It follows that, if $d(a,b)$ is inefficient, i.e., if $a<b$ and $D<(a+b)/2<U$, then $\bar{x}(d_T,d_P)$ is characterized by the first-order condition

$$O(\bar{x},d_T,d_P)\equiv u_T'(-\bar{x})[u_P(\bar{x})-d_P]-u_P'(\bar{x})[u_T(-\bar{x})-d_T]=0. \quad (11)$$

Now Condition C1 holds because $\partial O/\partial d_T=u_P'(\bar{x})>0$, $\partial O/\partial d_P=-u_T'(-\bar{x})<0$, and $\partial O/\partial \bar{x}=-u_T''(-\bar{x})[u_P(\bar{x})-d_P]-u_P''(\bar{x})[u_T(-\bar{x})-d_T]+2u_T'(-\bar{x})u_P'(\bar{x})>0$, because $u_T$ and $u_P$ are strictly increasing as well as concave. Moreover, Condition C2-1 of Proposition 3 is also true.

Denote the solution of (10) by $x(a,b)=\bar{x}[d_T(a,b),d_P(a,b)]$. Incorporating the conditions $\partial x/\partial a=\partial x/\partial b=0$ produces the following system of equations for unknowns $x$, $a$, and $b$:

$$u_T'(-x)\left[u_P(x)-u_P(a)F\left(\frac{a+b}{2}\right)-u_P(b)\left[1-F\left(\frac{a+b}{2}\right)\right]\right]-u_P'(x)\left[u_T(-x)-u_T(-a)F\left(\frac{a+b}{2}\right)-u_T(-b)\left[1-F\left(\frac{a+b}{2}\right)\right]\right]=0;$$

$$u_T'(-x)\left[-u_P'(a)F\left(\frac{a+b}{2}\right)-\frac{1}{2}u_P(a)f\left(\frac{a+b}{2}\right)+\frac{1}{2}u_P(b)f\left(\frac{a+b}{2}\right)\right]-u_P'(x)\left[u_T'(-a)F\left(\frac{a+b}{2}\right)-\frac{1}{2}u_T(-a)f\left(\frac{a+b}{2}\right)+\frac{1}{2}u_T(-b)f\left(\frac{a+b}{2}\right)\right]=0;$$

$$u_T'(-x)\left[-\frac{1}{2}u_P(a)f\left(\frac{a+b}{2}\right)-u_P'(b)\left[1-F\left(\frac{a+b}{2}\right)\right]+\frac{1}{2}u_P(b)f\left(\frac{a+b}{2}\right)\right]$$

[9] In fact, it also holds for all nonsymmetric (weighted) Nash solutions.

[10] To see that $(a^*,b^*)$ is a Nash equilibrium, note that $a^*$ is a best response for $T$ to $b^*$, because $a^*$ is the unique solution of $(\partial x/\partial a)(a,b^*)=0$ that satisfies $a<b$ and $D<(a+b)/2<U$. Moreover, the endpoints $2D-b^*$ and $2U-b^*$ are worse for $T$ than $a^*$ because $x(a^*,b^*)<\min\{x(2D-b^*,b^*),x(2U-b^*,b^*)\}=\min\{b^*,2U-b^*\}=b^*$. Similarly it can be verified that $b^*$ is a best response for $P$ to $a^*$.

[11] The Nash bargaining solution can of course be derived from axioms (Nash 1950, 1953), but it has also been justified on noncooperative, strategic grounds (Binmore et al. 1986, Rubinstein et al. 1992).

$$-u_P'(x)\left[-\frac{1}{2}u_T(-a)f\left(\frac{a+b}{2}\right)+u_T'(-b)\left[1-F\left(\frac{a+b}{2}\right)\right]\right.$$
$$\left.+\frac{1}{2}u_T(-b)f\left(\frac{a+b}{2}\right)\right]=0. \tag{12}$$

Now assume that the team $T$ is risk-neutral and write $u_T(z)=z$. Solving the first pair and then the last pair of equations of (12) for $F((a+b)/2)$ yields

$$F\left(\frac{a+b}{2}\right)=\frac{-u_P(x)+u_P(b)+u_P'(x)(b-x)}{-u_P(a)+u_P(b)+u_P'(x)(b-a)}$$
$$=\frac{1}{2}f\left(\frac{a+b}{2}\right)\frac{-u_P(a)+u_P(b)+u_P'(x)(b-a)}{u_P'(a)+u_P'(x)}$$
$$F\left(\frac{a+b}{2}\right)=\frac{u_P'(b)+u_P'(x)}{u_P'(a)+2u_P'(x)+u_P'(b)}.$$

To verify this solution, assume that the player $P$ is also risk-neutral, and write $u_P(z)=z$. Then it is easy to check that $F((a+b)/2)=\frac{1}{2}$, $\frac{1}{2}=(b-x)/(b-a)$, and $\frac{1}{2}=\frac{1}{2}f((a+b)/2)(b-a)$. Define $m\equiv F^{-1}(\frac{1}{2})$. It is straightforward to show that, when the players are risk-neutral, the outcome of the BFOA game is $a^*=m-1/(2f(m))$, $b^*=m+1/(2f(m))$, and $x=(a^*+b^*)/2=m$.

Returning now to the case in which $T$ is risk-neutral but $P$ is not, we find that equilibrium offers are not centered on the median, $m$, of the distribution $F$. This is the content of the following corollary to Proposition 1:

Corollary 1. *If the player $P$ is risk-averse and team $T$ is risk-neutral, then at any Nash equilibrium $(a^*, b^*)$, $(a^*+b^*)/2<m$.*

Proof. Risk aversion implies that $u_P$ is increasing and concave, so that $u_P'(b^*)<u_P'(a^*)$ because $a^*<b^*$. It follows that

$$F\left(\frac{a^*+b^*}{2}\right)=\frac{u_P'(b^*)+u_P'[x(a^*,b^*)]}{u_P'(a^*)+2u_P'[x(a^*,b^*)]+u_P'(b^*)}<\frac{1}{2},$$

which implies that $(a^*+b^*)/2<m$. □

We close with a specific example that uses the Nash bargaining solution, a risk-neutral team, and a risk-averse player.

Example 3. Suppose that player $P$ has utility $u_P(z)=0.01z-e^{-0.2z}$. Note that $u_P'(z)=0.01+0.2e^{-0.2z}$ and that $P$ is risk-averse because $u_P''<0$. Suppose further that the arbitrator's distribution is uniform on $[D,U]=[10,14]$ so that $m=12$ and $F(c)=(c-10)/4$ for $10\le c\le 14$. Note that all hypotheses of Proposition 3 hold because $-u_P''(z)/u_P'(z)=0.04/(0.01e^{0.2z}+0.2)<1/(14-10)$ for any $z<14$. Then a pure-strategy Nash equilibrium of the BFOA game is a solution of the following system of equations:

$$\frac{(a+b)/2-10}{4}=[(0.01+0.2e^{-0.2b})+(0.01+0.2e^{-0.2x})]\cdot[(0.01+0.2e^{-0.2a})+2(0.01+0.2e^{-0.2x})+(0.01+0.2e^{-0.2b})]^{-1};$$
$$\frac{(a+b)/2-10}{4}=[-(0.01x-e^{-0.2x})+(0.01b-e^{-0.2b})+(0.01+0.2e^{-0.2x})(b-x)]\cdot[-(0.01a-e^{-0.2a})+(0.01b-e^{-0.2b})+(0.01+0.2e^{-0.2x})(b-a)]^{-1};$$
$$\frac{(a+b)/2-10}{4}=\frac{1}{8}[-(0.01a-e^{-0.2a})+(0.01b-e^{-0.2b})+(0.01+0.2e^{-0.2x})(b-a)]\cdot[(0.01+0.2e^{-0.2a})+(0.01+0.2e^{-0.2x})]^{-1}. \tag{13}$$

The unique solution of (13) was determined numerically to be $a^*=9.7352$, $b^*=13.7357$, and $x(a^*,b^*)=11.869$. Denoting the certainty equivalent of $d_i(a^*,b^*)$ by $s_i(a^*,b^*)$, we have $D<(a^*+b^*)/2=11.735<U$, $s_T(a^*,b^*)>12$ and $s_P(a^*,b^*)=11.739$.

Note that in Example 3 the average offer, the settlement, the median (and mean) of the arbitrator's distribution, and the FOA certainty equivalents satisfy

$$\frac{a^*+b^*}{2}<s_P(a^*,b^*)<x(a^*,b^*)<m<s_T(a^*,b^*).$$

Thus, bargaining benefits the risk-averse player relative to FOA, in that the player receives $x(a^*,b^*)$ from bargaining but only (the equivalent of) $s_P(a^*,b^*)$ from FOA. But a risk-averse player achieves a smaller settlement than a risk-neutral player, who receives $m$, the median of the arbitrator's distribution.

## 5. Conclusions

The objective of this article has been to assess the extent to which adjoining FOA to a bargaining process tends to induce bargaining agreement. Experience in Major League Baseball and elsewhere suggests that this institutional arrangement often, but not always, leads to agreement. We have attempted to link this observation to theory.

Our primary argument is that risk aversion explains why parties come to an agreement, rather than submit to FOA. If at least one side is risk-averse, then there are agreements that both sides find preferable to FOA. Our results hold despite our assumption that both sides have the same beliefs about what the arbitrator considers fair; these beliefs may favor one side or the other, but we require that both sides agree on them.

The observation that the bargainers are unable to "divide the dollar" in up to 20% of cases may reflect disagreements about the arbitrator's distribution; an alternative explanation, one that we think unlikely, is that both sides are risk-neutral, which would imply that there is no settlement that both prefer to FOA.

The evidence we uncovered about the effect of FOA on prior bargaining is generally consistent with experience in baseball. We argued that players are most likely to be risk-averse, and in MLB players win about 60% of FOA decisions. As has been found elsewhere, FOA induces the more risk-averse party to be less demanding; if so, this party wins more often, but achieves less favorable settlements, than if it were risk-neutral.

When first proposed, final-offer arbitration was believed by many to be capable of inducing closure in bargaining—to draw the parties so close together that settlement would be automatic, or at least an easy step. But it is now generally accepted that, in FOA and related procedures, parties with equal information do not rationally diverge widely—but do not usually converge either. (See Armstrong 2004 and Brams et al. 1991.) However, newer arbitration procedures, such as Zeng's AFOA (Zeng 2003), appear to induce convergence reliably; they may be even more likely than FOA to lead to agreements in prior bargaining.

### Acknowledgments
The second author was supported by the Social Sciences and Humanities Research Council of Canada.

## Appendix

### Proof of Proposition 1
(P1-1) follows because $x(a, b)$ is continuous, bounded from below for any fixed $b$, and bounded from above for any fixed $a$. For example, if $b$ is fixed, then $a \geq b$ implies $x(a, b) = (a+b)/2 \geq b$, while $a \leq 2D - b$ implies $x(a, b) = b$; the situation is similar if $a$ is fixed.

To prove (P1-2), note that if $b \leq D$ then for any $a \leq b$, $x(a, b) = b$ so $x_T(b) = b$. If $b < U$ then $x_T(b) \leq b \leq 2U - b$. If $b \geq U$ and $2U - b \leq a \leq b$ then $x(a, b) = a$ and, in particular, $x(2U - b, b) = 2U - b$, so again $x_T(b) \leq 2U - b$. Therefore, for any $b$ it is true that $x_T(b) \leq \min\{b, 2U - b\}$. By continuity of $x_T(b)$, $\max_b \min_a x(a, b) = \max_b x_T(b)$ exists. Similarly, if $a \geq U$ then for any $b \geq a$, $x(a, b) = a$ and so $x_P(a) = a$. If $a > D$ then $x_P(a) \geq a \geq 2D - a$. If $a \leq D$ and $a \leq b \leq 2D - a$ then $x(a, b) = b$ and, in particular, $x(a, 2D - a) = 2D - a$, so again $x_P(a) \geq 2D - a$. Therefore, for any $a$, $x_P(a) \geq \max\{a, 2D - a\}$. Thus, by continuity of $x_P(a)$, $\min_a \max_b x(a, b) = \min_a x_P(a)$ exists.

Now we prove (P1-3), (P1-4), and (P1-5). For any max-minimizer $(a_0, b_0)$, $b_0 \geq D$ and $a_0 \leq \min\{b_0, 2U - b_0\}$ and, for any min-maximizer $(a^0, b^0)$, $a^0 \leq U$ and $b^0 \geq \max\{a^0, 2D - a^0\}$. In particular, these properties hold for any Nash equilibrium $(a^*, b^*)$. To establish that the inequalities are strict, suppose that $a$ and $b$ satisfy $a < b$ and $D < (a+b)/2 < U$. Taking derivatives, we find

$$\frac{\partial d_T}{\partial a} = \frac{1}{2} f\left(\frac{a+b}{2}\right)[u_T(-a) - u_T(-b)] - u_T'(-a) F\left(\frac{a+b}{2}\right)$$

$$\frac{\partial d_P}{\partial a} = -\frac{1}{2} f\left(\frac{a+b}{2}\right)[u_P(b) - u_P(a)] + u_P'(a) F\left(\frac{a+b}{2}\right)$$

$$\frac{\partial d_T}{\partial b} = \frac{1}{2} f\left(\frac{a+b}{2}\right)[u_T(-a) - u_T(-b)] - u_T'(-b)\left[1 - F\left(\frac{a+b}{2}\right)\right]$$

$$\frac{\partial d_P}{\partial b} = -\frac{1}{2} f\left(\frac{a+b}{2}\right)[u_P(b) - u_P(a)] + u_P'(b)\left[1 - F\left(\frac{a+b}{2}\right)\right].$$

Recall that $\partial\bar{x}/\partial d_T < 0$ and $\partial\bar{x}/\partial d_P > 0$ by Condition C1. Therefore, $\partial x/\partial a|_{a=(2D-b)^+} = ((\partial\bar{x}/\partial d_T)(\partial d_T/\partial a) + (\partial\bar{x}/\partial d_P)(\partial d_P/\partial a))|_{a=(2D-b)^+} < 0$ because $\partial d_T/\partial a|_{a=(2D-b)^+} > 0$ and $\partial d_P/\partial a|_{a=(2D-b)^+} < 0$. If $b \leq U$ then $\partial x/\partial a|_{a=b^-} > 0$ because $\partial d_T/\partial a|_{a=b^-} < 0$ and $\partial d_P/\partial a|_{a=b^-} > 0$. Thus, by continuity of $x(a, b)$, $b > D$ implies $x_T(b) < b$ and $D < b \leq U$ implies $x_T(b) < 2U - b$. Moreover, if $b > D$ and $\bar{a} \in \arg\min_{a'} x(a', b)$ then $2D - b < \bar{a} < b$. Likewise, if $\bar{a} < 2U - b$, and in particular if $D < b \leq U$, then $(\partial x/\partial a)(\bar{a}, b) = 0$ by continuity of $\partial x/\partial a$.

A similar argument shows that $\partial x/\partial b|_{b=(2U-a)^-} = ((\partial\bar{x}/\partial d_T)(\partial d_T/\partial b) + (\partial\bar{x}/\partial d_P)(\partial d_P/\partial b))|_{b=(2U-a)^-} < 0$. It follows that, if $a \geq D$, then $\partial x/\partial b|_{b=a^+} > 0$ because $\partial d_T/\partial b|_{b=a^+} < 0$ and $\partial d_P/\partial b|_{b=a^+} > 0$. Again by continuity of $x(a, b)$, if $a < U$ then $x_P(a) > a$, so if $D \leq a < U$ then $x_P(a) > 2D - a$. Moreover, if $a < U$ and $\bar{b} \in \arg\max_{b'} x(a, b')$ then $a < \bar{b} < 2U - a$. Analogously, if $\bar{b} > 2D - a$, and in particular if $D \leq a < U$, then $(\partial x/\partial b)(a, \bar{b}) = 0$ by continuity of $\partial x/\partial b$. (P1-3) now follows because $x(a^*, b^*) = x_T(b^*) = x_P(a^*)$. Note that $(D, D)$ and $(U, U)$ cannot be Nash equilibria, which implies (P1-4) and (P1-5), because $a^*$ is a min-maximizer and $b^*$ is a max-minimizer.

### Proof of Proposition 2
Condition C2 implies that, for any $b > D$, there exists a unique $\bar{a}(b) \in \arg\min_{a'} x(a', b)$ such that $\bar{a}(b)$ is a continuous function of $b$ and, if $(\partial x/\partial a)[\bar{a}(b), b] = 0$, and in particular if $D < b \leq U$, then $(dx_T/db)(b) = (\partial x/\partial b)[\bar{a}(b), b]$. Similarly, for any $a < U$ there exists a unique $\bar{b}(a) \in \arg\max_{b'} x(a, b')$ such that $\bar{b}(a)$ is a continuous function of $a$ and, if $(\partial x/\partial b)[a, \bar{b}(a)] = 0$, and in particular if $D \leq a < U$, then $(dx_P/da)(a) = (\partial x/\partial a)[a, \bar{b}(a)]$.

Proposition 1 guarantees the existence of a max-minimizer $(a_0, b_0)$ and a min-maximizer $(a^0, b^0)$. Note that $b_0 > D$ by the continuity of $x(a, b)$ and the facts that $dx_T/db|_{b=D^+} = (\partial x/\partial b)[\bar{a}(b), b]|_{b=D^+} > 0$, because $(\partial x/\partial a) \cdot [\bar{a}(b), b] = 0$, $2D - b < \bar{a}(b) < b$ and $\partial x/\partial b|_{b=a^+} > 0$. Suppose that $x_T(b_0) = 2U - b_0$, so that $\bar{a}(b_0) = 2U - b_0$. For every $b < b_0$, $x_T(b) \leq 2U - b_0 < 2U - b$, which implies that $\bar{a}(b) < 2U - b$. Then $dx_T/db|_{b=b_0^-} = (\partial x/\partial b)[\bar{a}(b), b]|_{b=b_0^-} = (\partial x/\partial b) \cdot [\bar{a}(b), b]|_{b=(2U-\bar{a}(b_0))^-} < 0$, which contradicts the fact that $x_T(b)$ is maximal at $b_0$. Therefore, $x_T(b_0) < 2U - b_0$ and $\bar{a}(b_0) < 2U - b_0$. Moreover, the continuity of $\partial x/\partial a$ and $\partial x/\partial b$ implies that $\partial x/\partial a = \partial x/\partial b = 0$ at $(a_0, b_0)$. It follows that $(a_0, b_0)$ is the unique max-minimizer, because $(d^2 x_T/db^2)(b_0) < 0$.

By analogous arguments, $a^0 < U$, $x_P(a^0) > 2D - a^0$ and $\bar{b}(a^0) > 2D - a^0$. Moreover, $\partial x/\partial a = \partial x/\partial b = 0$ at $(a^0, b^0)$ by continuity of $\partial x/\partial a$ and $\partial x/\partial b$, and $(a^0, b^0)$ must be the unique min-maximizer because $(d^2x_P/da^2)(a^0) > 0$. It follows that there is a unique pair $(a_0, b_0) = (a^0, b^0)$ for which $\partial x/\partial a = \partial x/\partial b = 0$. Thus, $\max_b \min_a x(a,b) = \min_a \max_b x(a,b)$, which implies that $(a_0, b_0) = (a^0, b^0)$ is a Nash equilibrium.

**Proof of Proposition 3**

Suppose that $a$ and $b$ satisfy $a < b$ and $D < (a+b)/2 < U$. Then we have the following second derivatives:

$$\frac{\partial^2 d_T}{\partial a^2} = \frac{1}{U-D}\left[-u'_T(-a) + u''_T(-a)\left(\frac{a+b}{2} - D\right)\right]$$
$$\frac{\partial^2 d_P}{\partial a^2} = \frac{1}{U-D}\left[u'_P(a) + u''_P(a)\left(\frac{a+b}{2} - D\right)\right]$$
$$\frac{\partial^2 d_T}{\partial b^2} = \frac{1}{U-D}\left[u'_T(-b) + u''_T(-b)\left(U - \frac{a+b}{2}\right)\right]$$
$$\frac{\partial^2 d_P}{\partial b^2} = \frac{1}{U-D}\left[-u'_P(b) + u''_P(b)\left(U - \frac{a+b}{2}\right)\right].$$

It follows that $\partial^2 d_T/\partial a^2 < 0$ and $\partial^2 d_P/\partial b^2 < 0$. Moreover, under Condition C2-2, $\partial^2 d_P/\partial a^2 > (1/(U-D))[u'_P(a) + u''_P(a)(U-D)] > 0$ because $a < U$, and $\partial^2 d_T/\partial b^2 > (1/(U-D)) \cdot [u'_T(-b) + u''_T(-b)(U-D)] > 0$ because $b > D$.

Note that $\partial x/\partial a = -((\partial O/\partial d_T)(\partial d_T/\partial a) + (\partial O/\partial d_P) \cdot (\partial d_P/\partial a))/(\partial O/\partial \bar{x})$ and assume that $\partial x/\partial a = 0$. Then Conditions C1 and C2-1 imply that $\partial^2 x/\partial a^2 = -((\partial O/\partial d_T) \cdot (\partial^2 d_T/\partial a^2) + (\partial O/\partial d_P)(\partial^2 d_P/\partial a^2))/(\partial O/\partial \bar{x}) > 0$. By an analogous argument, the assumption that $\partial x/\partial b = 0$ combined with Conditions C1 and C2-1 leads to the conclusion that $\partial^2 x/\partial b^2 < 0$.

## References

Armstrong, M. S. 2004. A comparison of arbitration procedures for risk averse disputants. *Decision Sci.* **35** 639–664.

Binmore, K., A. Rubinstein, A. Wolinsky. 1986. The Nash bargaining solution in economic modelling. *RAND J. Econom.* **17** 176–188.

Brams, S. J., D. M. Kilgour, S. Merrill III. 1991. Arbitration procedures. H. P. Young, ed. *Negotiation Analysis*. University of Michigan Press, Ann Arbor, MI, 47–65.

Chun, Y., W. Thompson. 1982. Nash solution and uncertain disagreement points. *Games Econom. Behav.* **2** 213–223.

Commission on Health and Safety and Workers' Compensation. 1999. Literature review: Final offer arbitration, http://www.dir.ca.gov/chswc/basebalarbffinal.htm.

Crawford, V. P. 1982. Compulsory arbitration, arbitral risk and negotiated settlements: A case study of bargaining under imperfect information. *Rev. Econom. Stud.* **49**(1) 69–82.

Curry, A. F., P. Pecorino. 1993. The use of FOA as a screening device. *J. Conflict Resolution* **37** 655–669.

Farmer, A., P. Pecorino. 1998. Bargaining with informative offers: An analysis of final offer arbitration. *J. Legal Stud.* **27** 415–432.

Gardner, R. 1995. *Games for Business and Economics*. John Wiley & Sons, New York.

Greenstein, E., I. Weissman, Y. Gerchak. 2004. Estimating arbitrator's hidden judgement in final offer arbitration. *Group Decision and Negotiation* **13**(3) 291–298.

Kilgour, D. M. 1994. Game-theoretic properties of final-offer arbitration. *Group Decision and Negotiation* **3** 285–301.

Luce, R. D., H. Raiffa. 1957. *Games and Decisions*. John Wiley & Sons, New York.

Major League Baseball Players Association. 2006. 2007–2011 basic collective agreement, http://mlbplayers.mlb.com/pa/pdf/cba_english.pdf.

Manzini, P., M. Mariotti. 2001. Perfect equilibria in a model of bargaining and arbitration. *Games Econom. Behav.* **37** 170–195.

Miller, P. 2005. It's arbitration time, http://thesportseconomist.com/2005/01/its-arbitration-time.htm.

Myerson, R. B. 1984. Two-person bargaining problems with incomplete information. *Econometrica* **52**(2) 461–488.

Myerson, R. B. 1991. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, MA.

Nash, J. F. 1950. The bargaining problem. *Econometrica* **18** 155–162.

Nash, J. F. 1951. Non-cooperative games. *Ann. Math.* **54** 286–295.

Nash, J. F. 1953. Two-person cooperative games. *Econometrica* **21** 128–140.

Rubinstein, A., Z. Safra, W. Thomson. 1992. On the interpretation of the Nash bargaining solution and its extension to non-expected utility preferences. *Econometrica* **60** 1171–1186.

Stevens, E. M. 1966. Is compulsory arbitration compatible with bargaining? *Indust. Relations* **5** 38–52.

Wittman, D. 1986. Final offer arbitration. *Management Sci.* **32**(12) 1551–1561.

Zeng, D.-Z. 2003. An amendment to final offer arbitration. *Math. Soc. Sci.* **46** 9–19.