Subgame Perfect
Plea Bargaining in Biform
Judicial Contests

Mordechai E. Schwarz
Subgame Perfect Plea Bargaining in Biform Judicial Contests

Mordechai E. Schwarz

The Open University of Israel and the Open University Research Institute for Policy, Political Economy and Society (OURI-PPES)

February 14, 2011

Abstract

The judicial process is modeled as a biform contest with risk-averse contestants. I provide a sufficient condition for a non-empty core and show that the effect of the severity of the charges on the core is ambiguous. I also show that the practice of plea bargaining actually applies Moulin's (1984) mechanism which non-cooperatively implements the Kalai-Smorodinsky (1975) bargain solution in subgame perfect equilibrium. The charge reduction rate is inversely related to the defendant's and positively related to the prosecutor's "fear of ruin" index, but increases with the defendant's stake, implying that the defendant's gain (and social cost) from the guilty plea deal increases with the severity of the crime.

Keywords: plea bargaining, subgame-perfect equilibrium, Kalai-Smorodinsky solution, implementation.

Jel Code: K41, C78, D86.
1. Introduction

Plea bargaining involves negotiations between a defendant and a prosecutor regarding the conditions under which the defendant enters a guilty plea. It is estimated that over 95% of the convictions in criminal cases in the United States result from a negotiated guilty plea, but the rate varies with the seriousness of the crime.

From a positivist point of view, as Carney and Fuller (1969) pointed out, "[I]t would seem reasonable to expect that the weaker the case against a defendant, the more entitled is he to a trial by jury, and, correspondingly, the more questionable is the propriety of the negotiated plea. In practice, however, it seems that the weaker the case (up to a point), the more likely is the prosecutor to press for a guilty plea, inasmuch as this was the most often mentioned reason given by prosecutors for trying a negotiated plea." Hence, the positivist puzzle of plea bargaining is what drives defendants to plead guilty when evidence is weak, and prosecutors to ask for lenient sentences in return, instead of declining the case.

Apparently, charge reduction may explain the defendants' tendency to plead guilty, so the question that remains is what drives a prosecutor to reduce charges in return for a guilty plea. If evidence is insufficient to prove the original charges but strong enough to prove the less serious charges, one would expect the prosecutor to indict for the reduced charges from the start, without making "concessions" to the defendant, which may be perceived by the public as weakness of the law enforcement system. On the other hand, if the evidence is too weak even for the lesser charge, what drives the defendant to plead guilty, even for reduced charges?

Moreover, one would expect that once a prosecutor agrees to reduce charges against a defendant in return for a guilty plea, he would also insist that this defendant receive the maximal sentence for this less serious offence if he pleads guilty. According to the Federal Sentencing Guidelines (FSG), a guilty plea is not recognized

---

2 Plea bargains constituted 98% of convictions for embezzlement, immigration and gambling, about 80% for kidnapping and assault but only 67% of those for murder (Reigannon, 2000).
3 Carney and Fuller (1969) reported that 67.8% of their sample entered pleas of guilty and 93.2% of these guilty pleas included charge reduction. Wright and Engen (2006, 2007) found that charge reduction occurs in roughly half of all felony cases that resulted in conviction, and has a large effect on sentencing. See also Wright and Miller (2002).
as relevant for sentencing (Ulmer et al., 2009). Nevertheless, Wright (2005), King et al. (2005) and Ulmer et al. (2009) found significant differences between sentences imposed on trial convicted defendants and those defendants who pleaded guilty for the same offense, even after controlling for legitimate factors, implying a trial penalty for insisting on the constitutional right to be judged. Namely, defendants who plead guilty enjoy charge reduction and additional sentence discounts.

Classical economic studies of law and economics (c.f. Becker, 1968) treated the judicial procedure as a lottery, assuming that conviction probability is exogenously determined by evidence. More recent studies shifted to the contest model, but the lottery model remained popular in plea bargaining analyses. The few exceptions either lack an explicit analysis of the judicial process as a contest, assume that litigants are risk-neutral and the prosecutor is sincerely motivated. These studies investigated, for example the effect of expected trial outcome on litigants' expenditures (c.f. Katz, 1988), or vice versa (c.f. Skaperdas and Gan, 1995), but I am unaware of any study which explicitly analyzes how all these variables are endogenously interrelated. To this end, a biform contest model is required.

In this article, I analyze the judicial procedure as a biform contest, namely a two-stage game. The first stage is a rent-seeking contest which results in a second stage bargaining game. I provide a sufficient condition for a non-empty core, show that the effect of the severity of the crime on agreement availability is ambiguous and then show that the common plea bargaining practice of charge reduction actually applies Moulin's (1984) mechanism which non-cooperatively implements the Kalai-Smorodinsky (1975) bargain solution in subgame-perfect equilibrium. The charge reduction rate is inversely related to the defendant's and positively related to the prosecutor's "fear of ruin" index. Charge reduction increases with the defendant's stake, implying that the defendant's gain (and social cost) from a guilty plea deal increases with the severity of the crime. It is also shown that risk-aversion is a necessary condition for guilty-plea deals.

The article proceeds as follows. Section 2 contains a brief review of related literature. Section 3 presents the benchmark one-shot contest model equilibrium. Section 4 presents the extended biform game model of judicial process with risk-averse rent-seeking contestants. Section 5 discusses the implications of risk-seeking and section 6 summarizes the article. All proofs are relegated to the appendix.

2. Related Literature

As mentioned above, classical economic studies of law and economics assumed that conviction probability is exogenously given, according to evidence strength, and although more recent studies shifted to the contest model, plea bargaining analyses continued to assume the lottery model. Also, economic analyses of plea bargaining usually assume risk-neutral litigants and a sincere prosecutor, or lack an explicit analysis of the judicial process as a contest.

Nevertheless, the controversy among philosophers and jurists regarding the moral and judicial aspects of plea bargaining indicates that scholars have the contest model in mind. Proponents of negotiated justice claim that the high costs associated with the production of evidence justifies plea bargaining which "saves the taxpayer money," especially when the defendant is not forced to plead guilty. The plea bargain, they argue, replicates the trial outside the courtroom and enables the parties to estimate their chances. Thus, the plea agreement reflects the expected verdict more efficiently. Clearly, this argumentation makes sense only if "evidence" is assumed to be a function of "efforts" exerted by the litigants, as assumed by the contest model.

The sincere prosecutor assumption is plausible within the lottery model framework because the exogenously fixed conviction probability neutralizes the prosecutor's incentives and motivation effect. The contest model assumes that "evidence," and thus winning probabilities, are determined endogenously according to

---

7 See, for example, Kipnis (1976), Huff, Rattner and Sagarin (1986), Scott and Stunz (1992), Schulhofer (1992), Gazal and Bar-Gill (2004) and Wright (2005).
8 Ulmer et al. (2009) report that 34% of the federal judges they interviewed agreed or strongly agreed that an "efficient case is an end in itself." No data were supplied about the prosecutors' attitude, but it is hard to believe that prosecutors who bear most of the burden of evidence production would be less generous to defendants for saving them time and effort.
efforts and resources invested by the contestants, implying that the sincere prosecutor assumption is implausible. Kipnis (1976) noted that prosecutors operate under public or administrative pressure to "produce" convictions, and are judged professionally by their conviction productivity. Miceli (1990) noted that the adversarial system itself naturally places the prosecutor and the defense attorney on opposite sides of an all-or-nothing contest. Consequently, each side has an incentive to present its most favorable case to the court, but the heavy caseload facing most prosecutors necessitates that the majority of cases be settled out of court through plea bargains. Therefore, to encourage such plea bargains, and to insure that the prosecutor is in a strong bargaining position, it is important that he establish a record of consistent success at trial, especially when the prosecutor's salary, prestige and promotion or reelection are presumably linked to his conviction record as a measure of his success. Reinganum (2000) noted that an elected prosecutor may maximize his chances of reelection by being "tough on crime" on one hand, and "saving the taxpayers' money" on the other. Plea bargaining serves both these apparently contradictory goals. For example, Wright (2005) found that as plea bargains became more common, the conviction rate increased while acquittals and dismissals declined sharply.9

These studies and the above mentioned empirical findings of charge reduction and sentencing discounts in return for "saving taxpayers' money" or "trial penalty" for wasting public resources by insisting on the constitutional right to a trial, reflect a principal-agent problem between the public (the principal) and the prosecutor (the agent). Principal-agent situations are characterized by an informational advantage of the agent over the principal. The prosecutor's informational advantage explains the almost automatic approval of plea arrangements by courts, as plea negotiations between the prosecution and the defense include fact bargaining and charge reduction, implying that the court and the public are not exposed to "bad facts." Guilty pleas are mutually beneficial to the prosecutor and the defendant, but not necessarily to society.

9 This finding also indicates that plea arrangements did not reflect expected verdicts but replaced inquiries and investigations as the main source of evidence. Wright (2005), Ulmer et al. (2009), Bowers (2007) and other researchers documented some commonly used techniques applied by prosecutors in order to extract guilty pleas from defendants, which cannot be reconciled with sincere truth seeking. These findings also undermine the "voluntary" and "knowing" basis of the defendant’s guilty plea. Similarly, assuming sincere judges and jurors ignores their incentive to save their own time and efforts, and to encourage litigants to end the process as quick as possible. The strategic bias of judges and jurors has been studied extensively in the literature. See, for example, Austin-Smith and Banks (1996), Feddersen and Pesendorfer (1998), Dekel and Piccione (2000) and Coughlan (2000).
The prosecution's ability to obtain judicial approval for socially detrimental plea agreements stems from the public's unawareness of these "bad facts." Hence, the deadweight loss associated with the future consequences of lenient sentences and short terms of imprisonment for ruthless criminals is underestimated by the public. Future victims have no faces, and their cries are presently silent.

3. The Model

3.1. Basic Settings

Consider two contestants, 1 (a defendant) and 2 (a prosecutor), competing in a judicial rent-seeking type contest. Denote the contestants' expenditures vector by \( \mathbf{x} = \{x_1, x_2\} \) and assume that the winning probability of agent \( i \) is given by a Tullock (1980) type Contest Success Function,

\[
p_i(\mathbf{x}) = \frac{dh_i(x_i)}{h_i(x_i) + dh_j(x_j)}, \quad i, j = 1, 2
\]

Where \( h_i(x_i) \) is agent \( i \)'s production function, which measures the effectiveness of agent \( i \)'s effort in producing evidence, and \( d \) is a positive parameter which measures agent \( i \)'s advantage (or disadvantage) over agent \( j \). In a judicial context, \( d \) reflects the standard of proof. Namely, if \( d_i > 1 \), the prosecutor bears a heavier standard of proof. For the sake of simplicity, assume \( h_i(x_i) = x_i \) and \( d = 1 \). This simplifying assumption does not affect the qualitative results of the model.\(^{10} \)

The judicial contest has two contingent results: acquittal (I) or conviction (II). In case of conviction, the defendant loses \( v_1 \) and the prosecutor gains \( v_2 \). Denote the

\(^{10}\) For axiomatization of Contest Success Functions, see Skaperdas (1996), Clark and Riis (1998), Epstein and Nitzan (2007) and Corhón and Dahm (2010).
initial wealth of the contestant $i$ by $A_i$ and his post contest net wealth by $w_i^k$, $i = 1, 2$, $k = I, II$. It follows that,

$$w_i^{I} = A_i - x_i, \quad w_i^{II} = A_i - x_i - v_i$$

(2)

$$w_2^{I} = A_2 - x_2, \quad w_2^{II} = A_2 - x_2 + v_2$$

### 3.2. Attitude toward Risk

Assuming risk aversion significantly complicates the analysis.\(^\dagger\) Therefore, the rent-seeking literature generally assumes risk-neutrality.\(^\dagger\) Nevertheless, in a bargaining, and particularly in a plea bargaining context, this assumption is not only extremely unrealistic, but misses an important behavioral factor (see Svejnar, 1986).

The most prevalent risk-aversion indices are Pratt's (1964) absolute risk-aversion index, defined as $R_i = -u''_i/u'_i$, Arrow's (1971) relative risk-aversion index, defined as $AR_i = R_i w_i$, and Aumann and Kurz's (1977) "fear of ruin" index, defined as $F_i \equiv u_i/u'_i$.\(^\dagger\) According to common definitions that prevail in the literature, an agent $i$ is defined as risk-neutral if $R_i = 0$, risk-averse if $R_i > 0$ and risk-seeker if $R_i < 0$. $F_i$ is a measure of agent $i$'s "fear of ruin" or "fear of disagreement," in Svejnar's (1986) term, which is more relevant in our context. Suppose that an agent faces a gamble in which he risks his entire wealth, $w_i$ in return for an additional small gain of $v_i$. As $v_i$ is small relative to $w_i$, the agent's loss probability, $q_i$, must be very small in order to make him gamble his entire wealth, and even smaller, as the agent's "fear of ruin" is larger. Thus, the "fear of ruin" is inversely related to "boldness." As $v_i$ is smaller, $q_i$ approaches zero. The "boldness" index is therefore, $\lim_{v_i \to 0} q_i/v_i$. To see that notice that the agent indifference condition is


\(^\dagger\) See Epstein and Nitzan (2007).

\(^\dagger\) On the interrelations between these indices of risk-aversion, see Foncel and Treich (2005).
Rearranging (3), assuming \( u_i(0) = 0 \) yields:

\[
\frac{q_i}{v_i} = \frac{u_i(w_i + v_i) - u_i(w_i)}{u_i(w_i + v_i)}.
\]

Taking the limit of (4) yields \( \lim_{v_i \to 0} q_i/v_i = u_i'/u_i'' = F_i \). Notice that even if \( R_i = 0 \), namely when an agent is risk-neutral according to common definitions, he still may have positive "fear of disagreement."

**3.3. The One-Shot Judicial Contest Competitive Equilibrium**

As a benchmark, let us begin by analyzing the judicial contest as a one-shot common rent-seeking contest. Assume that agents' preferences are represented by von-Neumann-Morgenstern utility functions \( u_i(w_i) \), \( u_i' > 0, u_i'' < 0 \). The agents' target functions are

\[
Eu_i(w_i) = p_1(x_i) u_i(w_i') + p_2(x_i) u_i(w_i''), \quad i = 1, 2
\]

First order conditions for interior solution is

\[
p_i''(x^*) \Delta u_i - Eu_i' = 0 \quad \forall i
\]
Where asterisks denote equilibrium values, \( p_i' = \frac{\partial p_i}{\partial x_i} \), \( \Delta u_i = u_i(w^i) - u_i(w'^i) \) and \( Eu'_i = p_i^* \left( x^* \right) u'_i(w^i) + \left( 1 - p_i^* \left( x^* \right) \right) u'_i(w'^i) \). With risk-neutral contestants, (6) is reduced to \( p_i'(x^*) v_i = 1 \) \( \forall i \), implying

\[
\begin{pmatrix}
\hat{x} \\
\hat{p}
\end{pmatrix} = \begin{pmatrix}
v_1 v_2 \\
\frac{v_1 v_2}{(v_1 + v_2)^2} \end{pmatrix}, \quad \begin{pmatrix}
v_1 \\
\frac{v_2}{v_1 + v_2}
\end{pmatrix},
\]

Where \( \hat{x} \) and \( \hat{p} \) denote equilibrium vectors under a risk-neutrality assumption.

4. Biform Games and Biform Contests

Brandenburger and Stuart (2007) described a biform \( n \)-player game (or hybrid game) as a two-stage game. The first stage is non-cooperative and is designed to describe the strategic moves of the players. However, the consequences of these moves are not payoffs. Instead, each profile of strategic choices at the first stage leads to a second stage cooperative game. This gives the competitive environment created by the choices that the players made in the first stage. Formally, a biform game is a collection \( (S, V, a) \), where \( S = S^1 \times S^2 \times \ldots \times S^n \) is a profile of strategies, \( V : A \times 2^n \rightarrow \mathbb{R} \) is a cooperative game coalitional function and \( a = \left\{ \alpha_i \right\}_{i=1}^n \) is a vector of "confidence indices" for each player \( \alpha_i \in [0,1] \). Roughly speaking, the confidence indices evaluate every contingent outcome in the core according to player \( i \)'s preferences.14

Similarly, a biform \( n \)-player contest is a two-stage game. The first stage is a rent-seeking contest designed to determine the bargaining power of the contestants. The

14 For a more accurate definition and explanation of confidence index, see Brandenburger and Stuart (2007) or Stuart (2005).
contest's outcome is not prize allocation, but a vector of equilibrium expenditures and winning probabilities that characterizes each contestant's bargaining power. Formally, a biform contest is a collection \((x, p, u, b)\) where \(x = \{x_i\}_{i=1}^n\) is the contestants rent-seeking expenditures vector, \(p\) is the contest success function, \(u\) is a vector of contestants' utilities and \(b\) is a bargaining problem.\(^{15}\)

The benchmark analysis in the previous section modeled the judicial contest as a one shot zero-sum non-cooperative game, in which the expected equilibrium net wealth of the defendant and the prosecutor are 

\[ w_1^* = A_1 - p_1^* v_1 - x_1^* \]

and 

\[ w_2^* = A_2 + p_2^* v_2 - x_2^* \]

respectively. Assuming risk-aversion and applying Jensen's inequality implies that 

\[ u_i(w_i^*) \geq E u_i(w_i) \]

Hence, there is \( w_i < w_i^* \), satisfying 

\[ u_i(w_i) = E u_i(w_i) \] (see Figure 1). \( w_i \) is the certainty equivalent wealth of agent \( i \).

A plea bargain aims at avoiding contest confrontation through a compromise that assigns the defendant a charge reduction of \( f(v_i) \) in return for pleading guilty.

Skaperdas and Gan (1995) analyzed a compromise in a contest problem, emphasizing that ”An agreed division of the prize, though, does not imply the absence of effort on

the contestants' part. Each contestant would like to establish a good bargaining position and the best way to do that is to have a credible fallback position in case negotiations fail." Skaperdas and Gan's study is thus an early analysis of a biform contest, where in the first stage contestants exert efforts to establish bargaining power and a credible negotiation threat in case of fallback, but also to be well equipped for the conflict, if necessary. If an agreement is achieved, the defendant's and the prosecutor's net wealth are \( w_1 = A_1 - p_2f(v_1) - x_1 \) and \( w_2 = A_2 + p_2f(v_2) - x_2 \), respectively. A compromise agreement is achievable if it is mutually beneficial, namely, if it ensures each contestant \( u_i(w_i) \geq u_i(w_i) \).^16

4.1. The Core of a Plea Bargaining Biform Contest

For simplicity, assume \( f(v_i) = \beta v_i \). A subgame perfect equilibrium is calculated by backward induction. Hence, assume that \( \beta \in [0,1] \) is the second stage bargaining solution and is common knowledge. Thus, the defendant's target function is

\[
\pi_1(x) = u_1(A_1 - p_2\beta v_1 - x_1),
\]

and the prosecutor's target function is

\[
\pi_2(x) = u_2(A_2 + p_2\beta v_2 - x_2),
\]

---

^16 Skaperdas and Gan (1995) assumed that the sharing rule, \( f(p_2) \), refers to the winning probability of agent 2. Here I assume that the sharing rule refers to \( v_i \). This formulation has a straightforward interpretation. \( p_2v_1 \) is the penalty discount for the defendant which reflects its expected value, and \( f(v_i) \) is the rate of charge reduction.
where $p_2$ is the prosecutor's winning probability. Clearly, the defendant seeks to minimize $p_2$ and the prosecutor seeks to maximize it. Solving the two optimization problems simultaneously yields:

$$
\hat{x} = \beta \hat{x} \Rightarrow \hat{p} = \hat{p}
$$

Where $\hat{x}$ and $\hat{p}$ are defined above in (7). Namely, as pointed out by Skaperdas and Gan (1995), bargaining position optimization implies that contestants' expenditures are proportional to risk-neutral agents' expenditures, where $\beta$ is the proportionality coefficient. Denote the expected payoff of the defendant and the prosecutor by $c_1 = p_2 v_1 + x_1$ and $c_2 = p_2 v_2 - x_2$, respectively, and it is easily obtained from (10):

$$
\tilde{c}_1 = \beta \tilde{c}_1, \quad \tilde{c}_2 = \beta \tilde{c}_2.
$$

The derivation of (11) was based on the assumption that in the first stage, the sharing rule, $\beta$, is exogenous and common knowledge. In real world plea bargaining negotiations, however, $\beta$ lies in the center of the dispute between the parties. Therefore, we first have to define the upper and lower limits of $\beta$. In other words, we now turn to explore the core of the plea bargain game.

A sharing rule $\beta$ (namely, a deal), is privately rational if it ensures each agent at least his certainty equivalent utility. A deal is rational if it is privately rational for all agents. The core, or the bargaining set (denoted by $B$), is the set of all rational deals.

Formally, define an agents' gain function by

$$
g_i(\beta) = \pi_i(\beta, \hat{x}) - Eu_i(w_i, \hat{x}^*).
$$
It can easily be verified that \( \frac{\partial g_1}{\partial \beta} < 0, \frac{\partial g_2}{\partial \beta} > 0, \frac{\partial^2 g_1}{\partial \beta^2} < 0 \) and \( \frac{\partial g_2}{\partial \beta} > 0 \), implying that the \( g_1 \) curve is downward sloping and the \( g_2 \) curve is upward sloping (see Figure 2). The core is defined as

\[
B(\beta) = \left\{ \beta \big| g_i(\beta) \geq 0, \forall i \right\}.
\]

In panel \( a \) of Figure 2, the core is represented by the segment \( AB \). The other two panels of Figure 2 show examples of empty cores.

**Figure 2**

Define \( \underline{\beta} = \inf \{ \beta \in B(\beta) \} \) (point \( A \) in Figure 2), and \( \overline{\beta} = \sup \{ \beta \in B(\beta) \} \) (point \( B \) in Figure 2).

**Proposition 1**

a. \( \beta = \frac{c_i^*}{\tilde{c}_2}, \quad \overline{\beta} = \frac{c_i^*}{\tilde{c}_1} \).

b. \( B(\beta) \neq \emptyset \iff \left( v_2 - v_1 \right) \frac{\partial p_2}{\partial F} \bigg|_{F=0} \leq \frac{\partial X}{\partial F} \bigg|_{F=0} \).
Proof: See appendix.

Proposition 1 implies that the core may be empty even when both agents are characterized with concave utility functions. The model is analytically solvable only if 

\[ u_i(w_i) = -e^{-w_i} \]

Figure 3 presents combinations of \( v_i \) and \( v_2 \) for a non-empty core simulated for this utility function specification.\(^{17}\)

---

4.2. Comparative Statics of the Core

Although a non-empty core may contain an infinite number of rational deals, our attention is focused on Pareto-efficient deals. Denote by \( P(B) \) the collection of all Pareto efficient deals in \( B \). The above mentioned characteristics of \( g_i \) imply that

\[ \frac{\partial g_2}{\partial g_1} < 0 \quad \text{and} \quad \frac{\partial^2 g_2}{\partial g_1^2} > 0 \]

Namely, the efficiency frontier curve, \( P(B) \), is downward sloping on the utilities plane \( (u_1, u_2) \), and the bargaining set, \( B(\beta) \), is convex and compact, as demonstrated in Figure 4. (\( d \) in Figure 4 denotes the disagreement point).

---

\(^{17}\) Notice that for this utility function \( R_j = 1 \) and \( F_i = -1 \).
In a plea bargaining context, disagreement means going to trial, hence \( d = (E_{u_1}, E_{u_2}) \). Namely, both \( B \) and \( d \) are endogenous, implying, as many authors have indicated, that comparative statics of the core in a biform contest framework is complicated.\(^\text{18}\) For instance, an increase in the prosecutor's first-stage efforts shifts the disagreement point from \( d_1 \) to \( d_2 \) and, consequently, the bargain set from \( B_1 \) to \( B_2 \) (see Figure 5).

For example, consider the effect of "fear of disagreement" on the core. Differentiating \( \overline{\beta} \) and \( \underline{\beta} \) with respect to \( F_i(A_i) \) and manipulating algebraically yields:

\(^{18}\) For example, see McDonald and Solow (1981), Alexander (1992), Anbarci, Skaperdas and Syropoulos (2002), Bayindir-Upmann and Gerber (2003) and Skaperdas (2006).
Corollary 1

\[ (a) \quad \frac{\partial \beta}{\partial F(A_i)} \geq 0 \iff \frac{\eta_{i,F_i}}{\eta_{i,F_i}} \geq \frac{x_i^*}{p_i v_i} \]

\[ (b) \quad \frac{\partial \beta}{\partial F(A_j)} \geq 0 \iff \frac{\eta_{j,F_j}}{\eta_{j,F_j}} \leq \frac{x_j^*}{p_j v_j} \]

where \( \eta_{ab} \) denotes the elasticity of \( a \) with respect to \( b \). It requires some tedious algebra to show that \( \text{sgn} \left( \frac{\partial \beta}{\partial v_i} \right) \), and \( \text{sgn} \left( \frac{\partial \beta}{\partial v_j} \right) \) are indeterminate. Assuming that \( v_i \) is positively correlated with the severity of the charge leads to the following corollary.

Corollary 2

The effect of the severity of the charges on agreement availability is ambiguous.

4.3. The Confidence Index

The main drawback of the core is that, on one hand, it may be empty, and on the other, it may contain an infinite number of solutions. In the latter case, a confidence index is useful. Consider two different cores that result from two different strategic profiles, as demonstrated in Figure 6. Denote the upper and lower limits of these cores by \( [\beta^U, \beta^L] \) and \( [\beta^0, \beta^L] \), respectively. The confidence index is used to evaluate the weighted average utility of each player for a given core. Roughly speaking, the confidence index indicates how well player \( i \) anticipates doing in the resulting
cooperative games. For example, in Figure 6, player \( j \) will choose strategy \( a \) if
\[
\alpha' u_i (\beta^l) + (1 - \alpha') u_i (\beta^u) > \alpha' u_i (\beta'^l) + (1 - \alpha') u_i (\beta'^u).
\]

In a plea bargaining context, a confidence index is less useful. First, the confidence level is defined in terms of the critical level for preferring strategy \( a \) over strategy \( b \) (see Figure 6), but, from (14) and Corollary 2, we know that the effect of the "fear of ruin" index on \( \beta \) and \( \beta' \) is ambiguous. The confidence index is probably correlated with the "fear of ruin" index, and this correlation should be taken into account. Secondly, equation (10) implies that equilibrium expenditures and \( \beta \) are interrelated. Namely, expenditures depend on the anticipated specific value of \( \beta \), while the boundaries of the \( \beta \) values depend on first stage equilibrium expenditures. The analysis of these interrelations requires a unique and precise bargain solution.

Nash (1950) suggested a unique bargain solution which satisfies four basic axioms: Invariance to affine transformation (IAT), Efficiency or Pareto Optimality (PO), Symmetry (S) and Independence of Irrelevant Alternatives (IIA). The IIA axiom was criticized and replaced with the monotony axiom (M) by Kalai and Smorodinsky (1975), who suggested an alternative unique bargain solution that satisfies IAT, PO, S and M.
4.4. The Kalai-Smorodinsky Plea Bargaining Solution

Define \( m_1 = g_1|_{\beta=\hat{\beta}} \) and \( m_2 = g_2|_{\beta=\hat{\beta}} \) as the maximum feasible utilities for agents 1 and 2, respectively. \( \mathbf{m} = (m_1, m_2) \) is known as the utopia point. The Kalai-Smorodinsky (henceforth KS) plea bargaining solution is the \( \beta \) which solves:

\[
g_i\left(\beta(v_i)\right) \frac{m_i}{g_i} = g_j\left(\beta(v_j)\right) \frac{m_j}{g_j} \quad \forall i, j
\]

Namely, the KS solution equalizes the proportional gains of every agent, relative to his utopian gain.\(^{19}\) In Figure 7, the KS solution is the intersection point of the \( \mathbf{dm} \) line and the efficiency frontier.

Consider the following Plea Bargaining-Protocol (PBP):

**Round 0:** The two parties invest \( x_1 \) and \( x_2 \) in order to establish a bargaining position.

\(^{19}\) Alternately, the KS solution is defined by \( ks_i = \max_{\beta \in [\hat{\beta}, \bar{\beta}]} \min_i \left( g_i(\beta)/m_i \right) \).
**Round 1:** Suppose that \( p_i > p_j \). Contestant \( i \) makes a proposal of \( \beta_i \) to \( j \). If \( j \) accepts the proposal, it is implemented by the court.\(^{20}\) If \( j \) rejects the proposal, proceed to Round 2.

**Round 2:** Contestant \( j \) makes a counter-proposal, \( \beta_j \). If \( i \) accepts, the court implements the deal with probability \( p_i \) and the status quo with probability \( 1 - p_i \). If \( i \) rejects the counter-proposal, the status-quo point is implemented with certainty.\(^{21}\) (See Figure 8).

---

**Figure 8**

![Diagram of Round 1 and Round 2 decision-making process](image)

This protocol is a modification, for a competitive environment, of the static mechanism suggested by Moulin (1984). Notice that after Round 0, contestants' expenditures are sunk-costs, implying that this mechanism pegs the bargaining set and sets \( d_i = \hat{p}_1 u_i \left( w_i \left( I \right) \right) + \hat{p}_2 u_i \left( w_i \left( II \right) \right) \). Hence, for convenience, let us normalize \( d = 0 \). Of course, legally \( \beta \in (0,1) \), thus \( m_1 = \pi_{1, \beta=0} \), and \( m_2 = \pi_{2, \beta=1} \), implying that the KS plea bargaining solution is \( \beta \in (0,1) \), which solves:

---

\(^{20}\) Legally, the court is not obliged to approve any plea agreement submitted by the litigants. However, since plea negotiations include fact bargaining, as explained above, the court is usually not exposed to "bad facts." Hence, the assumption that a "reasonable" plea agreement submitted by both sides will always be approved by the court is plausible.

\(^{21}\) The rationale of Round 2 stems from the assumption of "almost" automatic approval of guilty plea agreements in Round 1, as explained in note 20. Namely, as the bargaining process continues, flowing into Round 2, the public and the court are exposed to more "bad facts." Therefore, in Round 2 the approval probability is only \( p_i \).
Proposition 2

The Plea Bargaining Protocol (PBP) non-cooperatively implements the Kalai Smorodinsky (1975) bargaining solution in subgame-perfect equilibrium.

Proof: See appendix.

Intuitively, this mechanism solves the prosecutor’s credibility problem (Franzoni, 1999), since in case of disagreement the outcome is imposed by the court. Assuming

\[ u_i(w_i) = -e^{-w_i} \]

yields \( ks = \hat{p}_2 \). Figure 9 presents a simulation of \( ks \) for this function.
4.5. **Comparative Statics of the Sharing Rule**

Non-cooperative subgame-perfect implementation of a bargaining solution, \( f(v_i) \), implies that \( \hat{x} \) and \( \hat{p} \) are given by (10), thus the PBP mechanism pegs the bargaining set, simplifies comparative static analysis and reduces ambiguity.

By common wisdom, an increase in \( F_1 \) increases the defendant's tendency to accept worse deals, and an increase in \( F_2 \) increases the prosecutor's generosity toward the defendant. This intuition is true. The Taylor expansion of (16) is

\[
1 - \frac{u'(A)}{u_1(A)} \beta \hat{c}_1 = 1 + \frac{u'(A)}{u_2(A)} (\beta - 1) \hat{c}_2.
\]

Solving (17) for \( \beta \) and rearranging yields:

\[
ks = \frac{\hat{c}_2 F_1}{\hat{c}_1 F_2 + \hat{c}_2 F_1}
\]

Differentiating (18) yields:

\[
\frac{\partial ks}{\partial F_1} = \frac{\hat{c}_1 \hat{c}_2 F_1}{(\hat{c}_1 F_2 + \hat{c}_2 F_1)^2} > 0
\]

\[
\frac{\partial ks}{\partial F_2} = -\frac{\hat{c}_1 \hat{c}_2 F_1}{(\hat{c}_1 F_2 + \hat{c}_2 F_1)^2} < 0
\]
And,

\[
\frac{\partial k_s}{\partial v_1} = \frac{-v_2^2(4v_1 + v_2)F_1F_2}{4\left(v_1F_2^2(2v_1 + v_2) + v_2^2F_1^2\right)} < 0
\]

\[\text{(20)}\]

\[
\frac{\partial k_s}{\partial v_2} = \frac{v_1v_2(4v_1 + v_2)F_1F_2}{4\left(v_1F_2^2(2v_1 + v_2) + v_2^2F_1^2\right)} > 0
\]

Define the defendant's "discount function" by \(G_1 = \hat{p}_2v_1(1 - ks)\). This function measures the expected trial penalty imposed on a defendant who insists on his constitutional right to a court trial. Similarly, define the prosecutor's "concession function" by \(G_2 = \hat{p}_2v_2(1 - ks)\). By plugging \(\hat{c}_1, \hat{c}_2, \hat{p}_2\) and \(ks\) into \(G_1\) and \(G_2\), we obtain:

\[
G_1 = \frac{v_2^2F_2(2v_1 + v_2)}{(v_1 + v_2)[F_2v_1(2v_1 + v_2) + F_1v_2^2]}
\]

\[\text{(21)}\]

\[
G_2 = -\frac{v_1v_2^2F_2(2v_1 + v_2)}{(v_1 + v_2)[F_2v_1(2v_1 + v_2) + F_1v_2^2]}
\]

It can be easily verified that

\[
\frac{\partial G_1}{\partial F_1} < 0, \quad \frac{\partial G_1}{\partial F_2} > 0
\]

\[\text{(22)}\]

\[
\frac{\partial G_1}{\partial v_1} > 0, \quad \frac{\partial G_1}{\partial v_2} \geq 0 \iff \frac{F_1}{F_2} \leq \frac{2v_1^2(2v_1 + v_2)^2}{v_2^2(v_2^2 + 2v_1(v_2 + v_1))}
\]
However,

\[
\frac{\partial G_2}{\partial v_2} \geq 0 \iff 0 \leq \frac{F_2}{F_1} \leq \frac{v_2^3}{6v_1v_2(2v_1 + v_2) + v_2^3 + 8v_1^3}
\]

(23)

\[
\frac{\partial G_2}{\partial v_1} \leq 0 \iff \frac{F_2}{F_1} \leq \frac{v_2^2(v_2^2 + 4v_2v_1 + 2v_1^2)}{v_1^2(2v_1 + v_2)^2}
\]

Figure 10 presents our simulations for \( G_1 \) and \( G_2 \), respectively, assuming \( u_i = -e^{-u_i} \).

Figure 10 is compatible with Ulmer and Bradley (2006), who indicated that "In fact, under an organizational efficiency explanation, by far the most common in the sentencing literature, one might expect greater trial penalties among the relatively less serious cases here like robbery, because these might be seen as wasting the courts' time with unsuccessful trials when stakes (potential sentence severity) are lower compared to more severe cases. We find the opposite pattern – trial penalties increase with offense severity, at least for incarceration."
Define the total social cost of plea bargaining by $G = G_1 + G_2$. It follows that:

\[
G = \frac{v_1 v_2 (v_1 - v_2) F_2 (2v_1 + v_2)}{(v_1 + v_2) F_2 v_1 (2v_1 + v_2) + F_1 v_2^2}.
\]

**Corollary 3**

\[
G \gtrless 0 \iff v_i \gtrless v_j,
\]

(25)

\[
\frac{\partial G}{\partial F_i} \gtrless 0 \iff v_i \gtrless v_j
\]

Nevertheless, $\text{sgn} \left( \frac{\partial G}{\partial v_1} \right)$ and $\text{sgn} \left( \frac{\partial G}{\partial v_2} \right)$ are indeterminate. Figure 11 presents our simulations for $G$. (The left panel is calibrated by $F_1 = F_2 = 1$ and the right panel is calibrated by $v_1 = 100$ and $v_2 = 50$).

*Figure 11*
4.6. The Plea Bargaining Effect on Total Expenditures

Generally, the effect of risk-aversion on contestants' expenditures is ambiguous.\textsuperscript{22} Konrad and Schlesinger (1997) show that \( x_i^* \geq \hat{x}_i \iff Eu_i'(\hat{x}) \leq \frac{\Delta u_i(\hat{x})}{v_i} \). In terms of Figure 12, \( x_i^* \geq \hat{x}_i \iff \tan(\alpha) \leq \tan(\gamma) \).

However, from (10) we obtain \( x^* \geq \hat{x} \iff x^* \geq \beta \hat{x} \). Hence, the Konrad and Schlesinger condition is insufficient in a competitive environment.

**Corollary 4**

*The effect of guilty plea deals on total expenditures is ambiguous.*

Define the total expenditures function by \( X = x_1 + x_2 \), and denote the plea bargaining effect on expenditures by \( \Delta X = \bar{X} - X^* \). (Recall that \( \bar{X} = ks\bar{X} \)). The

upper row of Figure 13 contains combinations of $v_1$ and $v_2$, satisfying $\Delta X \geq 0$ for various values of $F_1$ and $F_2$, while the diagrams in the lower row of Figure 13 contain combinations of $F_1$ and $F_2$, satisfying $\Delta X \geq 0$ for various values of $v_1$ and $v_2$ (simulated for $u_i(w_i) = -e^{-w_i}$).

*Figure 13*

The ambiguity of plea bargaining effect on expenditures undermines the common "public resources savings" justification for the plea bargaining practice. Wright (2005) found that "a lighter workload for judges over the years produced fewer guilty pleas and more acquittals, but lighter workloads for prosecutors over time led to just the opposite result: more guilty pleas and fewer acquittals." Wright noted that this fact is "counterintuitive" and indicated that "When judges had more time to devote to criminal cases, the guilty plea rate tended to fall and the acquittal rate tended to rise: the judicial workload fell along with the guilty plea rate from the 1950's through the

---

23 For example, Justice Burger indicated: "If every criminal charge were subjected to a full-scale trial, the States and the Federal Government would need to multiply by many times the number of judges and court facilities." *Santobello v. New-York* 404 U.S. 257, 261 (1971).
1970's. The gentle rise in the judicial criminal caseload since 1981 corresponded with the large increase in the acquittal rate" (see also Ulmer et al., 2009).

5. Risk-Seeking Agents

The analysis in the previous section used the "fear of ruin" index as the risk-aversion measure. As noted above, $F_i$ may be positive even when $R_i = 0$, because this function measures "all or nothing" risk aversion, but not limited loss risk aversion. In particular, the $F_i$ function does not distinguish between agents with concave or convex utility functions (risk-aversers and risk-seekers by common definitions). Hence, for the analysis of risk-seeking effect on the core, we use Pratt's (1964) absolute risk-aversion index, defined as $R_i = -u''(u'_i)$.

Proposition 3

$B(\beta) \neq \emptyset \iff R_i \geq 0, \forall i$.

Proof: See appendix.

According to Kahneman and Tversky's (1979) "prospect theory," individuals classify outcomes relative to a certain reference point (for instance, the status quo). Outcomes that yield a higher utility level than the reference point are classified as gains while outcomes which yield lower utility are classified as losses. The expected utility of the individual is given by a value function such as $U(\bar{w}) = \sum_{l=1}^{L} \phi(p_l)u(w_l)$, where $w_l$ is a contingent outcome and $\phi(p_l)$ is the probability weighting function.

24 For case studies of the effect of the abolishment of plea bargaining, see, for example, Rubinstein and White (1979), Call, England and Talarico (1983) and Weininger (1987).
which is a non-linear transformation of the outcome's objective probability $p_i$ that overestimates small probabilities and underestimates large probabilities.

The value function does not measure wealth but changes in wealth, namely gains and losses. Hence, the reference point is normalized to the origin, the reference point is normalized to the origin, and the value function is S-shaped around it (see Figure 14), implying that individuals are not risk-avers but loss-avers, as the negative impact of losses is greater than the positive impact of identical (in absolute values) gains. In other words, according to prospect theory, when outcomes are perceived as losses, individuals tend to behave as risk-seekers.25

![Figure 14](image_url)

A defendant who wins in a trial is left with net wealth of $A_i - x_i$, and a defendant who loses is left with $A_i - x_i - v_i$. Namely, the two contingencies are net losses, implying that according to prospect theory, defendants are expected to behave as risk-seekers and reject any plea deal proposal. Proposition 3 also predicts that if agents are risk-seekers, the core is empty. Apparently, both predictions are incompatible with the above mentioned data about frequent plea bargaining deals.

One may argue that in terms of Figure 1, given the high rate of convictions, the defendants' reference point is $w_1^\mu$. However, this interpretation is incompatible with the bulk of evidence on defendants' behavior. On one hand, some researchers argue that defendants are subject to innocence bias, which causes them to over-estimate their winning probabilities and reject any guilty plea offers (c.f. Tor, Gazal-Ayal and

25 Kahneman and Tversky based their theory on experimental evidence, but their methodology evoked a polemic. See, for example, Gigerenzer (1991) and Kahneman and Tversky’s (1996) reply. However, this polemic is beyond the scope of this article.
Innocence bias implies that in Figure 1, the defendants' reference point is to the right of \( w_i^1 \). On the other hand, reports of the Innocence-Project organization revealed cases of innocent defendants falsely convicted after pleading guilty,\(^{26}\) implying that when the innocence bias is not binding, the defendants' reference point is, indeed, \( w_i^{1''} \).

Motti Michaeli\(^{27}\) suggested that our results can be reconciled with Prospect Theory because through charge reduction, the expected utility and the utility from expected wealth no longer lie on the same plumb. (In Figure 15, point \( A \) represents higher utility than point \( B \)). Nevertheless, the prevalence of guilty plea deals indicates that this explanation should be empirically tested and the Prospect Theory should be theoretically reconsidered and empirically retested.

\[ \text{Figure 15} \]

6. Summary and Discussion

In this article, I analyzed the judicial process as a biform contest, provided conditions for a non-empty core, and showed that the common plea bargain practice applies Moulin's (1984) mechanism which non-cooperatively implements the Kalai-Smorodinsky (1975) bargain solution in subgame-perfect equilibrium.

The PBP mechanism pegs the bargaining set and thus enables comparative static analysis of a compromise in a competitive environment. As expected, \( ks \) is positively

\(^{26}\) These reports are available at http://www.innocenceproject.org/.

\(^{27}\) A Ph.D. student at the Center for the Study of Rationality (Hebrew University of Jerusalem), under the supervision of Robert J. Aumann.
correlated with the defendant's "fear of disagreement", and negatively correlated with
the prosecutor's "fear of disagreement". However, $k_s$ is negatively correlated with the
defendant's stake and positively correlated with prosecutor's stake. If stakes
monotonically increase with crime severity, the defendant's gain and social costs
exhibit the same pattern. Contrary to the common justification for the plea bargaining
practice, the effect of guilty pleas on expenditures is ambiguous.

Although the "fear of disagreement" effect on the core is ambiguous, risk-
aversion is a necessary condition for a non-empty core. Apparently, this theoretical
result and the common rate of trials that result in guilty pleas are incompatible with
Kahneman and Tversky's (1979) Prospect Theory. Although Michaeli's suggestion for
reconciling my results with Prospect Theory is nice, I hold that further research is
required as this explanation should be empirically tested, and Prospect Theory should
be both theoretically reconsidered and empirically retested.

The simplified analysis above reflects the common wisdom about criminal justice
in countries where negotiated justice prevails: "criminal codes do not matter much." More
important than the codes' provisions is the prosecutor's application of the
criminal statutes. Wright and Engen (2006) examined charge reductions in North
Carolina and found that charge reductions were common, "occurring in roughly half
of all felony cases that resulted in convictions, and that the prosecutor's decision to
reduce criminal charges has a large effect on average sentence severity." But they also
found that "these effects do not apply equally, however, to all crimes. When a group
of related crimes offer deeper charging options (that is, the number of charges that
might apply to a given set of facts), the prosecution and defense agree more often to
reduce the charges. Large distances between the sentences that attach to available
charging options make it less likely that the prosecution and defense will agree on a
particular charge reduction. Thus, plea bargaining is not an entirely free-market
exercise that allows the parties to negotiate a customized outcome. Even in a world of
charge-driven sentencing where prosecutorial discretion is a dominant feature, the
substantive criminal law matters."\footnote{See also Wright and Engen (2007).}

Following the above analysis, I suggest a reverse causality interpretation of
Wright and Engen's findings. It is not the number of substitute charging options that
cause charge reduction in guilty plea deals, but the law enforcement system's interest in encouraging these judicial deals. Without deciding whether innocence bias matters and to what extent, it seems plausible to assume that both over-optimism and the mental price of a false guilty plea increases with the offence severity, implying that charge reduction weakens the innocence bias effect and encourages even innocent defendants to plead guilty. 29 I conjecture that as the severity of the charges increases, deeper charge reduction is required by the KS solution, leading the law enforcement system to press legislators to supply more charging options for certain felony crimes. 30

The positivist view of the judicial procedure as a rent-seeking contest and the state prosecutor as a rent-seeker is, indeed, troubling. Normatively, justice should be equal for the poor and the rich. State prosecutors should not behave as contractors of convictions but seek the truth honestly and prosecute sincerely. Conviction should be based on reliable evidence beyond doubt and on due procedure. However, "It is not the mouse that is the thief but the hole" (Babylonian Talmud, tractate Gittin 45a). Namely, the built-in conflict of interests of the prosecutor's position, due to its principal-agent characteristics creates an incentive for reduced efforts. The plea bargain practice is the "hole" that enables the prosecutor to escape his duties as a public agent.

7. Appendix

Proof of Proposition 1

If no deal is achieved, the agents turn to judicial competition in the court. Their expected utilities are

---

29 Of course, the defendant’s innocence is his private information; hence, the bargaining outcome is inefficient as guilty defendants are also expected to enjoy charge reductions.

30 Bower (2007) found that in petty cases, charge reduction is relatively rare and the plea bargain is centered mainly on the sentence.
\[ Eu_1(w_1) = p_1 u_1 (A_1 - x_1^*) + p_2 u_1 (A_1 - m_1) \]

(A1)

\[ Eu_2(w_2) = p_1 u_2 (A_2 - x_2^*) + p_2 u_2 (A_2 + m_2) \]

Where \( m_1 = (v_1 + x_1^*) \) and \( m_2 = (v_2 - x_2^*) \). On the other hand, if the two agents agree on a deal, \( \beta(v_i) \), their expected utilities are

\[ \pi_1 (\beta, x) = u_i (A_1 - \beta \hat{e}_1) \]

(A2)

\[ \pi_2 (\beta, x) = u_2 (A_2 + \beta \hat{e}_2) \]

where \( \hat{e}_1 = \hat{p}_2 v_1 + \hat{x}_1 \), \( \hat{e}_2 = \hat{p}_2 v_2 - \hat{x}_2 \), and \( \hat{x} \) and \( \hat{p} \) are defined in (7). Taking the Taylor expansion of (A1) yields:

\[ Eu_1(w_1) = p_1 \left[ u_1 (A_1) - x_1^* u_1'(A_1) \right] + p_2 \left[ u_1 (A_1) - m_1^* u_1'(A_1) \right] + \mathcal{R}_1 \]

(A3)

\[ Eu_2(w_2) = p_1 \left[ u_2 (A_2) - x_2^* u_2'(A_2) \right] + p_2 \left[ u_2 (A_2) + m_2^* u_2'(A_2) \right] + \mathcal{R}_2 \]

Where \( \mathcal{R}_i \) is the remainder of the Taylor sequence. Similarly, the Taylor expansion of (A2) yields:

\[ \pi_1 (\beta, x) = u_i (A_1) - \beta \hat{e}_1 u_1'(A_1) + \mathcal{R}_1 \]

(A4)

\[ \pi_2 (\beta, x) = u_2 (A_2) + \beta \hat{e}_2 u_2'(A_2) + \mathcal{R}_2 \]
The gain of agent $i$ from a deal is defined by $g_i = \pi_i(\beta, x) - Eu_i(w_i)$. The core is non-empty if $g_i \geq 0 \ \forall i$. Thus, combining (A3) and (A4), dividing by $u_i'(A_i)$ and rearranging yields the following conditions for a non-empty core:

\[
\begin{align*}
\bar{g}_i &= (1 - p^*_1 - p^*_2) F_1(A_i) - \beta \hat{c}_i + p^*_1 \hat{x}_i + p^*_2 m_i \geq 0 \\
\bar{g}_2 &= (1 - p^*_1 - p^*_2) F_2(A_2) - \beta \hat{c}_2 + p^*_1 \hat{x}_2 + p^*_2 m_2 \geq 0
\end{align*}
\]

where $\bar{g}_i = \frac{g_i}{u_i'(A_i)}$. Recall that by definition $1 - p^*_1 - p^*_2 = 0$, and it follows that

\[
\beta = \frac{c^*_2}{\hat{c}_2}, \quad \bar{\beta} = \frac{c^*_1}{\hat{c}_1} \quad \text{QED(a)}.
\]

By inserting $c^*_i$ and $\hat{c}_i$ into (A6), we easily obtain

\[
\beta = \frac{p^*_2 v_2 - x^*_2}{p^*_2 v_1}, \quad \bar{\beta} = \frac{p^*_2 v_1 + x^*_2}{p^*_2 v_1 (2 - \hat{p}_2)}.
\]

Define $\Delta \beta = \bar{\beta} - \beta$. Non-emptiness of the core implies $\Delta \beta \geq 0$. Hence, by (A7),

\[
\Delta \beta \geq 0 \iff \frac{p^*_2 v_2 - \hat{x}_2}{p^*_2 v_1 + \hat{x}_1} \geq \frac{p^*_2 v_2 - x^*_2}{p^*_2 v_1 + x^*_1}.
\]
The left hand side of the inequality in (A8) is constant. In other words, (A8) implies that the "fear of ruin" affects \( c_1^* \) (the denominator on the right side of (A8)) more than \( c_2^* \) (the numerator). Suppose that the "fear of ruin" increases for both agents at the same rate, namely, \( \partial F_1 = \partial F_2 = \partial F \). It follows that

(A9) \[ \left( \frac{\partial p_2}{\partial F} v_2 - \frac{\partial x_2}{\partial F} \right)_{F=0} \leq \left( \frac{\partial p_2}{\partial F} v_1 + \frac{\partial x_1}{\partial F} \right)_{F=0} \]

Rearranging (A9) and inserting \( \frac{\partial X}{\partial F} = \frac{\partial x_1}{\partial F} + \frac{\partial x_2}{\partial F} \) yields:

(A10) \[ (v_2 - v_1) \frac{\partial p_2}{\partial F} \bigg|_{F=0} \leq \frac{\partial X}{\partial F} \bigg|_{F=0} \text{ QED(b).} \]

**Proof of Proposition 2:**

A subgame perfect equilibrium is computed by backward induction. Suppose without loss of generality that \( p_1 > p_2 \), so in Round 1, agent 1 makes an offer, \( \beta \), to agent 2.

By rejecting agent 1's proposal and proposing \( \beta \), agent 2's expected utility is

(A11) \[ Eu_2 \geq \hat{p}_1 u_2(\beta) + \left(1 - \hat{p}_1\right) u_2(d_2) \]

Going back to Round 1, the best response of agent 1 is to offer \( \beta \), satisfying

(A12) \[ u_2(\beta) = \hat{p}_1 u_2(\beta) + \left(1 - \hat{p}_1\right) u_2(d_2) \]
By rearranging, (A12) can be rewritten as

\[
\frac{u_2(\beta) - u_2(d_2)}{u_2(\beta) - u_2(d_2)} = \hat{p}_1
\]

(A13)

Now suppose that \( p_2 > p_1 \), thus in Round 1 agent 2 makes an offer to agent 1. Similar argumentation leads to the conclusion that agent 2's best response strategy is to propose \( \beta \) satisfying

\[
u_1(\beta) = \hat{p}_1 u_1(\beta) + (1 - \hat{p}_1) u_1(d_1),
\]

(A14)

which by rearranging terms yields

\[
\frac{u_1(\beta) - u_1(d_1)}{u_1(\beta) - u_1(d_1)} = \hat{p}_1.
\]

(A15)

Combining (A14) and (A15) yields

\[
\frac{u_1(\beta) - u_1(d_1)}{u_1(\beta) - u_1(d_1)} = \frac{u_2(\beta) - u_2(d_2)}{u_2(\beta) - u_2(d_2)}.
\]

(A16)

By definition, the \( \beta^*(v_i) \), which solves (A16), is the Kalai-Smorodinsky bargaining solution. Going back to the first stage we obtain the subgame perfect equilibrium expenditures from (10). \textit{QED.}
Proof of Proposition 3:

If no deal is achieved and the agents turn to judicial competition in the court, their expected utilities are given by (A1), where \( m_1 = \left( v_1 + x_1^* \right) \) and \( m_2 = \left( v_2 - x_2^* \right) \). On the other hand, if the two agents agree on a deal, \( \beta(v_i) \), their expected utilities are given by (A2), where \( \hat{c}_1 = \hat{p}_1 v_1 + \hat{x}_1 \) and \( \hat{c}_2 = \hat{p}_2 v_2 - \hat{x}_2 \). Taking the Taylor expansion of (A1) and rearranging yields:

\[
Eu_i(w_i) = u_i - p_i \left[ x_i^* u'_i - \frac{1}{2} x_i^{*2} u''_i \right] - p_i \left[ m'_i u'_i - \frac{1}{2} m''_i u''_i \right] + R_i
\]

(A17)

\[
Eu_2(w_2) = u_2 - p_1 \left[ x_2^* u'_2 - \frac{1}{2} x_2^{*2} u''_2 \right] + p_2 \left[ m_2 + \frac{1}{2} m_2^{*2} u''_2 \right] + R_2
\]

Where \( u_i, u'_i \) and \( u''_i \) are evaluated at \( A_i \). Similarly, the Taylor expansion of (A2) yields:

\[
\pi_i(\beta, x) = u_i - \beta \hat{c}_i u'_i + \frac{(\beta \hat{c}_i)^2}{2} u''_i + R_i
\]

(A18)

\[
\pi_2(\beta, x) = u_2 + \beta \hat{c}_2 u'_2 + \frac{(\beta \hat{c}_2)^2}{2} u''_2 + R_2
\]

The gain of agent \( i \) from a deal is defined by \( g_i = \pi_i(\beta, x) - Eu_i(w_i) \). The core is non-empty if \( g_i \geq 0 \ \forall i \). Thus, subtracting (A18) from (A17), dividing by \( u'_i(A_i) \) and rearranging yields the following conditions:
\[
\bar{g}_1 = -\frac{1}{2} \left( \beta \hat{c}_1 \right)^2 R_1 - \beta \hat{c}_1 + \frac{1}{2} \left[ v_1 \left( 2x_1^* - v_1 \right) R_1 + 2v_1 \right] p_1^* + m_1^* \left( 1 + \frac{1}{2} m_1^* R_1 \right) \\
\bar{g}_2 = -\frac{1}{2} \left( \beta \hat{c}_2 \right)^2 R_2 + \beta \hat{c}_2 + \frac{1}{2} \left[ v_2 \left( 2x_2^* - v_2 \right) R_2 + 2v_2 \right] p_1^* - m_2^* \left( 1 - \frac{1}{2} m_2^* R_2 \right)
\]

(A19)

where \( \bar{g}_i = \frac{g_i}{u_i'(A_i)} \). The roots of (A19) are:

\[
\bar{\beta}_{1,2} = \frac{-1 \pm \Delta_{1}}{R_1 \hat{c}_1}, \quad \beta_{1,2} = \frac{1 \pm \Delta_{2}}{R_2 \hat{c}_2}
\]

(A20)

where

\[
\Delta_{1} = \sqrt{1 + \left[ v_1 \left( 2x_1^* - v_1 \right) p_1^* + m_1^{**} \right] R_1^2 + 2 \left( p_1^* v_1 + m_1^* \right) R_1}
\]

(A21)

\[
\Delta_{2} = \sqrt{1 + \left[ v_2 \left( 2x_2^* - v_2 \right) p_1^* + m_2^{**} \right] R_2^2 + 2 \left( p_1^* v_1 - m_2^* \right) R_2}
\]

The relevant solutions are, of course, non-negative values of \( \beta \), where

\( \bar{\beta} = \max \left( \bar{\beta}_1, \bar{\beta}_2 \right) \) and \( \underline{\beta} = \min \left( \bar{\beta}_1, \bar{\beta}_2 \right) \). As already mentioned above, the core is non-empty if and only if \( \Delta \beta = \bar{\beta} - \underline{\beta} \geq 0 \). Namely, the core is non-empty if and only if

\[
\bar{\beta} = \frac{-1 + \Delta_{1}}{R_1 \hat{c}_1} \geq \frac{1 - \Delta_{2}}{R_2 \hat{c}_2} = \beta
\]

(A22)
Clearly, (A22) can hold for non-negative values of $\bar{\gamma}$ and $\bar{\beta}$ if and only if $R_i > 0 \ \forall i$. \textit{QED.}

References


