BARGAINING WITH MECHANISMS: TWO-SIDED UNCERTAINTY

MARCIN PĘSKI

Abstract. Bargaining with mechanisms - general results for general incomplete information. We characterize the worst and the best payoffs with two-sided incomplete information and two types for each player.

1. Introduction

?.

2. Model

2.1. **Environment.** There are two players $i = A, B$. Players must decide on a social outcome $x \in X$ The space of outcomes X is convex and it contains outcome 0 that corresponds to punting on the decision.

Player i has a finite set of types T_i. We assume finiteness to simplify the level of technicality needed for this paper - the assumption can be easily lifted. Let ΔT_i be the space of beliefs about the types. Let $T = \times_i T_i$ be the space of type profiles, and let $\Delta T = \times_i \Delta T_i$ be the space of belief profiles. For each type t_i and outcome x, let $\tau(x, t_i)$ denote the payoff of type t_i from allocation x. We assume that the payoffs are convex.

An environment is tuple $(X, (T_i), \tau)$, a space of outcomes X, sets of types, and payoff mapping.

An (incentive compatible) allocation given beliefs $p \in \Delta T$ is a mapping $\chi : T \to X$ such that for each type $t_i, s_i \in T_i$,

$$\tau(\chi, t_i) := \int \tau(\chi(t_i, t_{-i}), t_i)\, dp_{-i}(t_{-i}) \geq \int \tau(\chi(s_i, t_{-i}), t_i)\, dp_{-i}(t_{-i}).$$

Date: May 31, 2023.

Let $U=\mathbb{R}^{\bigcup_i T_i}$ be the space of payoff vectors. Let

$$\mathcal{U}(p)=\left\{(\tau(\chi,t_i))_{i,t_i\in T_i}\in U:\chi \text{ is incentive compatible allocation}\right\}$$

be the set of payoff vectors in all incentive compatible allocations given prior p. Observe that correspondence $\mathcal{U}:\Delta T\rightrightarrows U$ is nonempty-valued ($0\in\mathcal{U}(p)$ for each p) and convex-valued. Additionally, assume that $\mathcal{U}$ is u.h.c. Correspondence $\mathcal{U}$ contains all information about the environment that is relevant in the rest of the paper.

From now on, we refer to correspondences $e:\Delta T\rightrightarrows\mathbb{R}^{\bigcup_i T_i}$ that are u.h.c., nonempty- and convex-valued as *Kakutani correspondences.* Let $\mathcal{E}_0$ be the space of Kakutani correspondences.

For any two Kakutani correspondences e,f, say that $e\subseteq f$ if $e(p)\subseteq f(p)$ for each p.

Let $d_{\Delta T},d_U$ be metrics on the space of belief profiles and payoff vectors. Define (asymmetric) distance function between correspondences as:

$$d_0(e,f)=\max_{(p,u)\in f}\min_{(q,v)\in e}d_{\Delta T}(p,q)+d_U(u,v).$$

A sequence of correspondences e_n *approximates* f, $e_n\rightarrowtail f$, if $d_0(e_n,f)\to 0$.

Let $d(e,f)=d_0(e,f)+d_0(f,e)$ be a metric on $\mathcal{E}$. Metric d is equivalent to the Hausdorff distance of correspondences e,f treated as subsets of $\Delta T\times U$. A sequence of correspondences e_n *converges to* f, $e_n\to f$, if $d(e_n,f)\to 0$.

2.2. **Mechanisms.** A (finite, compact) *game* is a tuple $g=((A_i),\xi)$, where A_i is a set of actions of player i and $\xi:\times_i A_i\to X$ is an outcome function. In the game, players simultaneously choose actions $a_i\in A_i$ from (finite, compact) set A_i and outcome $\xi(a_i,a_{-i})$ is implemented.

A game g and a belief profile p gives rise to a *game with incomplete information* $G(g,p)$. We assume that, prior to taking their actions, players observe a public randomization device. For each mechanism g and belief profile p, let $E_g(p)\subseteq\mathcal{U}(p)$ denote the set of payoff vectors in Bayesian equilibria of $G(g,p)$. Then, E_g is a Kakutani correspondence.

An (abstract) mechanism is defined as a Kakutani correspondence $e \subseteq \mathcal{U}$. Let $\mathcal{E} \subseteq \mathcal{E}_0$ denote the set of all abstract mechanisms.

For example, a correspondence of all incentive compatible allocations $\mathcal{U}$ is an abstract mechanism. We refer to $\mathcal{U}$ as the universal mechanism. For another example, if g is a game, then $E(g)$ is a mechanism. For each mechanism m, if $E(g) = m$, then we say that g fully implements m, and if $E(g) \subseteq m$, then we say that g partially implements m.

The above definition deviate from the standard terminology, in which mechanism is typically understood as a "game designed for a particular purpose". What we mean by a mechanism is more typically understood as a social choice correspondence. Our choice is driven by the goals of this paper, where we are less interested in implementation issues (where the distinction between a social choice correspondence and a game is of primary importance) and more in the sets of equilibrium payoffs obtained in particular games. Instead of relying on mouthful "equilibrium payoff correspondence", we replace it by a mechanism.

The question of which mechanisms can be implemented in games is important, but orthogonal to questions studied in this paper.

2.3. **Efficient mechanisms.** (Welfare) weights is a function $\Lambda : T_1 \cup T_2 \to [0,1]$ such that $\sum_{t_i \in T_i} \Lambda(t_i) = 1$ for each i.

Payoff vector u is Λ-optimal under beliefs p if

$$u \in \arg\max_{u \in \mathcal{U}(P)} \sum_{t_i} \Lambda(t_i)\, u(t_i).$$

Payoff vector is p-interim efficient if it is Λ-optimal for some weights Λ and beliefs p. Let $\mathcal{U}_{\text{eff}}(p) \subseteq \mathcal{U}(p)$ denote the set of all interim efficient vectors.

A mechanism m is *interim efficient* if $m \subseteq \mathcal{U}_{\text{eff}}$.

2.4. **Examples of mechanisms.** This section contains constructions of new mechanisms. Let m be a mechanism.

- For each $\delta < 1$, define "discounted" mechanism δm as a game obtained from mechanism m by multiplying all payoffs in m by δ.

- Let $\mu \in \Delta\mathcal{E}$be a probability distribution over mechanisms with a finite support. Define a randomized mechanism $R(\mu)$ so that for each p

$$R(\mu)(p) = \left\{ \int \xi(m)\, d\mu(m) \text{ where } \xi : \text{supp}\mu \to U \text{ is a selection } \xi(m) \in m(p) \right\}.$$

- Let $I_i m$ be a mechanism, where player i makes an announcement from a rich Borel space, following which mechanism m is implemented. We refer to $I_i m$ as the player i information revelation game.
- Given a set of mechanisms A, define a menu of mechanisms $MM_i(A)$ as a mechanism where player i chooses a mechanism from the set A of mechanisms and makes an announcement from a rich Borel space S; after the choice, the mechanism is implemented.
- Given mechanism c,define informed principal with continuation game $IP_i(c)$ as a mechanism, where player i designs a mechanism m; after that, players play $MM(-i, \{m, c\})$:

$$IP_i(c) = MM_i(\{MM_{-i}(\{m, c\}) : m \text{ is a mechanism}\})$$

2.5. **Random-proposer bargaining game.** Random-proposer bargaining is defined as the largest fixed point:

$$RB = IP\left(i, \delta \cdot R\left((RB_A)^{\beta}(RB_B)^{1-\beta}\right)\right),$$

where RB_i is the game after which proposer i is selected and $RB = R\left((RB_A)^{\beta}(RB_B)^{1-\beta}\right)$ is the randomized mechanism before the proposer is selected.

Show that the above is well-defined: the largest fixed point exists.

2.6. **Equilibrium selection.** Perfect Bayesian equilibrium.

Equilibrium selection: $MM(i, \{m, c\})$ game.

3. Equilibrium payoff bounds

3.1. **Derived mechanisms.** In this subsection, we present definitions of various abstract mechanisms derived from a given mechanism m.

For each player i and prior $p_i \in \Delta T_i$, define comparisons between player i payoffs in payoff vectors

- $u \leq_{p_i} u'$ (resp. $u <_{p_i} u'$) iff $u(t_i) \leq u'(t_i)$ for each t_i st. $p(t_i) > 0$ (resp. if, additionally, at least one of the inequalities is strict),
- not $u \leq_{p_i} u'$ (resp. not $u <_{p_i} u'$) iff there exist t_i st. $p(t_i) > 0$ and $u(t_i) > u'(t_i)$ (resp. $u(t_i) \geq u'(t_i)$).

In other words, $\leq_{p_i}$ is a standard vector comparison but applied only to payoffs associated with player i types that have strictly positive probability under p_i.

3.1.1. *Lower bounds.* For each mechanism m, define an *upper contour* mechanism $U_i m$:

$$U_i m(p) = \{u \in \mathcal{U}(p) : u' \leq_{p_i} u \text{ for some } q_i \in \Delta T_i \text{ and } u' \in m(q_i, p_{-i})\},$$
$$U_i^0 m(p) = \{u \in \mathcal{U}(p) : u' <_{p_i} u \text{ for some } q_i \in \Delta T_i \text{ and } u' \in m(q_i, p_{-i})\}.$$

$U_i m(p)$ (resp. $U_i^0 m(p)$) contains incentive compatible payoff vectors u that dominate (resp., strictly dominate), from the point of view of player i, some payoff vector u' that is an equilibrium of the mechanism m under some other beliefs about player i type. For future reference, we note without the proof that $U_i m \subseteq U_i n$ for any pair of mechanisms $m \subseteq n$ and that $U_i(U_i m) = U_i m$.

One can interpret the elements of $U_i m(p)$ as payoffs that can be sustained in equilibrium of some arbitrary game, when player i has an option to reject the game and revert to mechanism m. If the rejection is exercised, player $-i$ updates beliefs about i's types to q_i and a continuation $u' \in m(q_i, p_{-i})$ is implemented. The equilibrium conditions ensure that $u' \leq u$. In fact, a subsequent Lemma shows that mechanism $U_i m$ is payoff equivalent to a menu of mechanisms where player i chooses between universal mechanism $\mathcal{U}$ and mechanism m:

$$U_i m = MM_i(\{m, \mathcal{U}\}). \tag{1}$$

In general, we have the following characterization of menus of mechanisms:

Lemma 1. *For each player i, each compact set of mechanisms A, each mechanism*

$$MM_i(A) = I_i\left(\bigcup_{m\in A} m\right) \cap \bigcap_{m\in A} U_i m \text{ for each compact set of mechanisms } A$$
$$IP_i(c) = I_i(U_{-i}c) \cap \bigcap_m U_i MM_{-i}(\{m,c\}) \text{ for each mechanism } c.$$

Proof. Suppose $u \in MM_i(A)$. Let S be the space of announcements. There exists $\alpha \in \Delta(A\times S)$ and measurable mappings $u, q : A\times S \to \bigcup_{m\in A} m$ such that $u(m,s) \in m(q(m,s))$ for each $m \in A$ and $s \in S$ and such that

$$u = \int u(m,s)\, d\alpha(m,s) \text{ and } u(t) \geq u(t|m,s) \text{ for all } m, s \text{ and } p_i\text{-almost all } t_i.$$

This implies that $u \in I_i(\bigcup_{m\in A} m)$ and that $u \in U_i m$ for each $m \in A$.

On the contrary, suppose that $u \in I_i(\bigcup_{m\in A} m)$ and that $u \in U_i m$ for each $m \in A$. The former implies that there exists signaling space S, strategy $\beta \in \Delta(S)$ and measurable mappings $u, q : S \to \bigcup_{m\in A} m$ such that $u(s) \in (\bigcup_{m\in A} m)(q(s))$ for each $s \in S$ and

$$u = \int u(s)\, d\beta(s) \text{ and } u(t) \geq u(t|s) \text{ for all } m, s \text{ and } p_i\text{-almost all } t_i.$$

The latter implies that for each m, there are $v^m \in m(q^m)$ such that $v^m \leq u$.

Find a measurable function $\mu : S \to A$ so that for $u(s) \in \mu(s)(q(s))$ β-a.s. Construct a strategy $\alpha \in \Delta(A\times S)$ so that for each measurable function $f : A\times S \to \mathbb{R}$, $\int f(\mu(s), s)\, d\beta(s) = \int f(m,s)\, d\alpha(m,s)$. Let $u^{\mu(s),s} = u(s)$ and $q^{\mu(s),s} = q(s)$ and for each (m,s) such that $m \neq \mu(s)$, let $u^{m,s} = v^m$ and $q^{m,s} = q^m$. For all $m \in A\backslash\mu(S)$, let Find a measurable selection and such that

$$u = \int u(m,s)\, d\alpha(m,s) \text{ and } u \geq u(m,s) \text{ for all } m, s.$$

The first claim implies that for each mechanism c,

$$\begin{aligned} IP_i(c) &= I_i\left(\bigcup_m MM_{-i}(\{m,c\})\right) \cap \bigcap_m U_i MM_{-i}(\{m,c\}) \\ &= I_i\left(U_{-i}c \cap \left[\bigcup_{m\in A} U_{-i}m \cap I_{-i}(m\cup c)\right]\right) \cap \bigcap_m U_i MM_{-i}(\{m,c\}) \end{aligned}$$

By taking $m = \mathcal{U}$, we obtain that $\bigcup_{m\in A} U_i m \cap I_i(m \cup c) = \mathcal{U}$. This demonstrates the last claim. $\square$

The characterization of a menu of mechanism with the universal mechanism (1) is a corollary to the Lemma (recall that the universal mechanism is closed with respect to revelation of information (i.e., $I_i(\mathcal{U}) = \mathcal{U}$).

For any two mechanisms m, n, we say that mechanism m is a (resp. strict) *i-lower bound* on n if $n \subseteq U_i m$ (resp., $n \subseteq U_i^0 m$). The lower bound relation is transitive: if m is i-lower bound on m', and m' is i-lower bound on m'', then m is i-lower bound on m''. Additionally if one of the first two relations is strict, the third relation is strict as well. We have the following observation.

Lemma 2. *If mechanism n is i-lower bound on m for some mechanism $m \in A$, then n is i-lower bound on $MM_i(A)$. It follows that for any mechanism n, if $I_{-i}(U_i n) \subseteq U_i n$, then n is i-lower bound on $IP(-i, n)$.*

Proof. The first claim follows directly from Lemma 1 and the monotonicity of U_i operator. It follows that n is a i-lower bound on $MM_i(\{m, n\})$ for any mechanism m. Applying Lemma 1, the first claim, and the monotonicity of I_{-i} operator, we obtain

$$\begin{aligned} IP(-i, n) &= MM_{-i}(\{MM_i\{m, n\} : m \text{ is a mechanism}\}) \\ &\subseteq I_{-i}\left(\bigcup_m MM_i\{m, n\}\right) \\ &\subseteq I_{-i}(U_i n) \subseteq U_i n. \end{aligned}$$

$\square$

3.1.2. *Upper bounds.* For each mechanism m, each $p \in \Delta T$, define

$$\begin{aligned} W_i m(p) &= \{u \in \mathcal{U}(p) : \text{not } u <_{q_i} u' \text{ for each } q_i \in \Delta T_i \text{ and } u' \in m(q_i, p_{-i})\}, \\ W_i^0 m(p) &= \{u \in \mathcal{U}(p) : \text{not } u \leq_{q_i} u' \text{ for each } q_i \in \Delta T_i \text{ and } u' \in m(q_i, p_{-i})\}. \end{aligned}$$

Here, not $u \leq_{p_i} u'$ means that there exist t_i st. $p(t_i) > 0$ and $u(t_i) > u'(t_i)$, with a similar definition of not $u <_{p_i} u'$.

Mechanism W_i contains payoff vectors that are not dominated (from the point of view of player i) by any payoff in mechanism m for some (possibly different) belief about types of player i. The difference between the two mechanisms in the complete information case is illustrated in Figure ???. The next result describes a relation between the two mechanisms. Illustrate on complete information example.

Lemma 3. *For each m and n,*

(1) $m \cap U_i n = \emptyset \Leftrightarrow n \subseteq W_i^0 m$,
(2) $m \cap U_i^0 n = \emptyset \Leftrightarrow n \subseteq W_i m$.

Proof. For claim (1), suppose that $m \cap U_i n = \emptyset$. Take $u \in n(p)$. Then, for each q_i, each $u' \geq_{q_i} u$, we have $u' \in U_i n(q_i, p_{-i})$, which implies that $u' \notin m(q_i, p_{-i})$. In other words, for any q_i and $u' \in m(q_i, p_{-i})$, we have not $u \leq_{q_i} u'$. It follows that $u \in W_i^0 m(p)$.

Next, suppose that $n \subseteq W_i^0 m$. Take any $u \in U_i n(p)$. Then, there exists q_i such that $u' \leq_{p_i} u$ and $u' \in n(q_i, p_i)$. Because of the latter, $u' \in W_i^0 m(q_i, p_{-i})$, which, together with the former, implies that $u \notin m(p)$.

The second claim follows from analogous arguments used for the strict inequality. □

If $m \cap U_i n = \emptyset$ (resp. $m \cap U_i^0 n = \emptyset$), we say that mechanism n is strict (resp. weakly) *i-upper bound* on m. The characterization of the definition from Lemma 3 illustrates that, in some sense, the upper bound is an opposite notion to the lower bound. As with the lower bound, the upper bound relation is transitive.

The definition is supposed to capture an idea that an upper bound is a preferable mechanism. This is formalized in the next result: we show that if n is a strict i-upper bound on m, then when allowed to choose between the two mechanisms, player i will choose mechanism n.

Lemma 4. *If mechanism n is a strict i-upper bound on mechanism m, then $MM_i(\{m,n\}) \subseteq I_i(n)$ and $I_i(n)$ is a $(-i)$-lower bound on $IP_{-i}(m)$.*

Proof. Suppose that n is a strict i-upper bound on mechanism. By Lemma 1, $MM_i(\{m,n\}) = U_i m \cap U_i n \cap I_i(m \cup n)$. Take any $u \in MM_i(\{m,n\}) \subseteq I_i(m \cup n)(p) \cap U_i n(p)$. Let $\alpha \in \Delta(\{m,n\} \times S)$ be the equilibrium strategy and $u^{x,s}, q_i^{x,s}$ payoffs and beliefs supporting u as an equilibrium of game $I_i(m \cup n)$. Suppose that $u \notin I_i(n)(p)$, which implies that $\alpha(m) > 0$. Then, $u(t) = u^{m,s}(t)$ for a α-positive probability signals s and $q_i^{m,s}$-all types t, which implies that $u \leq_{q_i^{m,s}} u^{m,s}$. We can choose s so that $\mathrm{supp} q_i^{m,s} \subseteq \mathrm{supp} p_i$.

Because $u \in U_i n(p)$, there exists $q_i' \in \Delta T_i$ and $u' \in n(q_i', p_{-i})$ such that $u' \leq_{p_i} u$. The latter and the choice of $q_i$implies that $u' \leq_{q_i^{m,s}} u$, and, by transitivity, $u' \leq_{q_i^{m,s}} u^{m,s}$. But this implies that $u' \notin W_i^0 n(q_i, p_{-i})$ and it contradicts the fact that n is a strict i-upper bound on mechanism. □

3.1.3. *Efficiency.* We defined mechanism m as interim efficient if, given beliefs p, $m(p)$ maximizes agents payoffs among all feasible and incentive compatible payoff vectors. In this section, we define a stronger notion of efficiency that incorporates payoffs of players after deviation and change of beliefs. Say that mechanism m is *robustly efficient* if $U_i^0 m \cap U_{-i}^0 m = \emptyset$. Intuitively, mechanism is robustly efficient if its payoffs cannot be dominated by payoff vector combined of payoffs obtained for each player after some revelation of information for this player.

It is easy to see that a robustly efficient mechanism must be interim efficient. The following test based on the above intuition can be used to check robust efficiency:

Lemma 5. *Suppose that for all beliefs p and q, all $u \in m(p)$, all $u^i \in m(q_i, p_{-i})$ for each player i, there exist weights Λ such that u is Λ-optimal, and*

$$\sum \Lambda_i u_i \leq \sum \Lambda_i u_i^i.$$

Then, m is robustly efficient.

Proof. Take $u \in m(p)$ and, on the contrary, suppose that there exists $v \in U_i^0 m \cap U_{-i}^0 m$. Find q and, for each player i,$u^i \in m(q_i, p_{-i})$ such that $v_i > u_i^i$ for each i. But this leads to the contradiction with the hypothesis. □

The next result establishes a connection between robust efficiency and lower and upper bounds:

Lemma 6. *Suppose that m is robustly efficient. If m is a strict i-lower bound on n, then m is $(-i)$-upper bound on n.*

Proof. If $n \subseteq U_i^0 m$ and $U_i^0 m \cap U_{-i}^0 m = \emptyset$, then $n \cap U_{-i}^0 m = \emptyset$. $\square$

It is easy to see that any robustly efficient mechanism must be efficient. On the other hand, for a mechanism to be robustly efficient, it must

3.1.4. *Information revelation.* The statements of some results (specifically Lemmas 2 and 4) become much simpler if the underlying mechanisms m are closed wrt revelation of information. The next two results present some characterization.

Lemma 7. *Suppose that*

3.2. **General results.**

Theorem 1. *Suppose that $a(x), x \in [0,1]$ is a collection of mechanisms such that for each x,*

(1) $a(x)$ is robustly efficient,
(2) $a(x)$ is closed wrt. $(-i)$-revelation,
(3) $U_i a(x)$ is closed wrt $(-i)$-revelation,
(4) $a(tx)$ is a strict i-lower bound on $t'a(x)$ for any $t < t' \leq 1$,
(5) $a(1 - t'(1-x))$ is strict $(-i)$-upper bound on $ta(x)$ for any $t < t' \leq 1$,,
(6) for each $\alpha \in [0,1]$ and each $x, x' \in [0,1]$, $a(\alpha x + (1-\alpha)x')$ is i-lower bound on $\alpha U_i a(x) + (1-\alpha) U_i a(x')$,
(7) $a_i(0)$ is i-lower bound on RB.

Then, $a_i(\beta_i)$ is a lower bound on RB.

Proof. Let $x^* = \sup\{x : a(x) \text{ is a } i\text{-lower bound on } RB\}$. Because of assumption 7, the set is non-empty and $x^* \geq 0$ is well-defined. Suppose that $x^* \leq \beta_i$. We will show that $a\left(x^* + \frac{1}{2}(1-\delta)(\beta_i - x^*)\right)$ is a i-lower bound on RB. If $x^* < \beta_i$, we obtain a contradiction with the definition of x^*. If $x^* = \beta_i$, this shows that $a(\beta_i)$ is a i-lower bound on RB, which establishes our claim.

First, property 4 implies that that $a(x)$ is a strict i-lower bound on $a(y)$ for any $x < y$.

Next, the definition of x^* and the transitivity of the lower bound relation implies that $a(x)$ is a i-lower bound on RB for any $x < x^*$.

Let $x_0 = x^* - \frac{1}{2}(1-\delta)(\beta - x^*) < x^*$ and take any $y \in (x_0, x^*)$. By the above, $a(y)$ is a strict i-lower bound on RB, i.e., $RB \subseteq U_i^0 a(y)$. Because $a(y)$ is robustly efficient (assumption 1), Lemma 6 implies that $a(y)$ is $(-i)$-upper bound on by RB. Because $U_i \delta m = \delta U_i m$ for any mechanism m, $\delta a(y)$ is $(-i)$-upper bound on δRB. Assumption 5 implies that $a(1-\delta(1-x_0))$ is strict $(-i)$-upper bound by δRB (take $t' = \delta \frac{1-x_0}{1-y}$ and $t = \delta$). Assumption 2 and Lemma 4 imply that $a(1-\delta(1-x_0))$ is a i-lower bound on $RB_i = IP(i, \delta RB)$.

At the same time, mechanism $\delta a(x_0)$ is a i-lower bound on δRB, which, by transitivity and assumption 4, implies that $a(\delta x_0)$ is a i-lower bound on δRB. Assumption 3 and Lemma 2 imply that $a(\delta x_0)$ is a i-lower bound on $RB_{-i} = IP(-i, \delta RB)$.

It follows that

$$(1-\beta_i) U_i a(\delta x_0) + \beta_i U_i a(1-\delta(1-x_0))$$

is a i-lower bound on $RB = RB_i^{\beta_i} RB_{-i}^{1-\beta_i}$. Let

$$\begin{aligned}
x_1 &= (1-\beta_i)\delta x_0 + \beta_i(1-\delta(1-x_0)) \\
&= \delta x_0 + \beta_i(1-\delta) \\
&= x_0 + (1-\delta)(\beta_i - x_0) \\
&\geq x_0 + (1-\delta)(\beta_i - x^*) \\
&= x^* + \frac{1}{2}(1-\delta)(\beta_i - x^*) = x_2,
\end{aligned}$$

Because of assumption 6, we obtain that $a(x_1)$ is a i-lower bound on RB. Because $x_2 \leq x_1$, we obtain $a(x_2)$ is a i-lower bound on RB. □

4. Single good with transfers and two types

This section is devoted to an environment with a single good, transfers, and one-dimensional types.

Let $X = \{(q,t) : q \in [0,1], t \in \mathbb{R}\}$. Suppose that $T_i \subseteq \mathbb{R}$ for each player i. For each type $v_i \in T_i$, each outcome $(q,t) \in X$, define

$$\tau_i((q,t), v_i) = \begin{cases} qv_i - t & i = 1 \\ (1-q)v_i + t & i = 2 \end{cases}.$$

Peski (22) studied the bargaining in this environment with one sided uncertainty, , ie. where $|T_2| = 1$. The main result is that there are unique equilibrium payoffs and they equivalent to a property-rights mechanism. In the mechanism, each agent becomes an owner of the probability share of the good equal to its bargaining power and chooses a price at which they are willing to sell their share of the good to the other player.

There are two types $l_i < h_i$ of each player i. We assume that

$$l_1 < l_2 < h_1 < h_2.$$

The probability of high type is denoted with p_i.

4.1. **Property rights mechanism.** Under the property rights mechanism, player 1 owns x fraction of the good, and player 2 owns $1-x$. Each player decides a price at which they are willing to sell their fraction to the other player. The other player always accepts the offer if indifferent:

- type l_1 of player 1 offers to sell at price $\pi = l_2$ if $p_2 \le p_2^* = \frac{l_2 - l_1}{}$ and price $\pi = h_2$, which leads to payoffs

$$L_1^*(p_2|x) = \begin{cases} xl_2 & p \le p_2^* \\ x(l_1 + p_2(h_2 - l_1)) & p \ge p_2^* \end{cases},$$

- type h_1 of player 1 sells at price $\pi = h_2$, which leads to payoffs

$$H_1^*(p_2|x) = x(h_1 + p_2(h_2 - l_1)),$$

- type l_2 of player 1 sells at price $\pi = h_1$, which leads to payoffs

$$L_1^*(p_2|x) = (1-x)(l_2 + p_2(h_2 - l_1)),$$

- type h_2 never sells. The payoff depends on the selling price of type l_1. Denote $H_1^*(p_2|x) = (1-x)h_2$. Then, the payoff of type h_2 is equal to

$$\begin{cases} H_1^*(p_2|x) + x(1-p_1)(h_2-l_1) & p \leq p_2^* \\ H_1^*(p_2|x) & p \geq p_2^* \end{cases}.$$

When $p \leq p_2^*$, type h_2 benefits from the low selling price of type l_1.

We refer to $\theta_i^*(p_{-i}|x)$ as the x-property rights payoff of type θ_i. This terminology is a bit abusive in the case of payoffs of type h_2, where the payoffs above contain additional component when $p_2 \leq p_2^*$.

4.2. **Player 1 payoff bounds.** Under mechanism $a^1(x)$, player 1 receives the payoffs from the property right mechanism where she owns x fraction of the good (i.e., payoffs $\theta_1^*(.|x)$ for $\theta = L, H$). Player 2 payoffs are described in Figure 1). The description is divided into three cases A, B,and C. The payoffs on the boundaries are convex combination of continuous limits of payoffs in the neighboring areas. In each case, the payoffs *beyond* the payoffs from the property rights mechanism are colored. In each case, a small illustration shows the allocation of the good as a function of realized types, with the shaded area corresponding to the probability that the good goes to player 1.

4.2.1. *Interim efficiency.* In each of the three cases, the payoffs are interim efficient:

- Zone A: if $p_2 \geq p_2^*$, and $p_1 \leq p_1^*(x) = \frac{x}{1-x+x\frac{h_2-h_1}{h_2-l_1}}$, the mechanism behaves as the property right mechanism, with a modification that, when type l_1 faces type l_2, type l_1 offers a sale of γ units of the good at price l_1 (out of their share of x units). The modification does not affect payoffs (or incentive conditions) of player 1. The modification increases the payoffs of type l_2 by $\gamma(1-p_1)(l_2-l_1)$. The amount γ is chosen so to make type h_2 indifferent between revealing her type truthfully and mimicking type l_2:

$$(1-x)h_2 = (1-x)((1-p_1)h_2 + p_1h_1) + \gamma(1-p_1)(h_2-l_1),$$

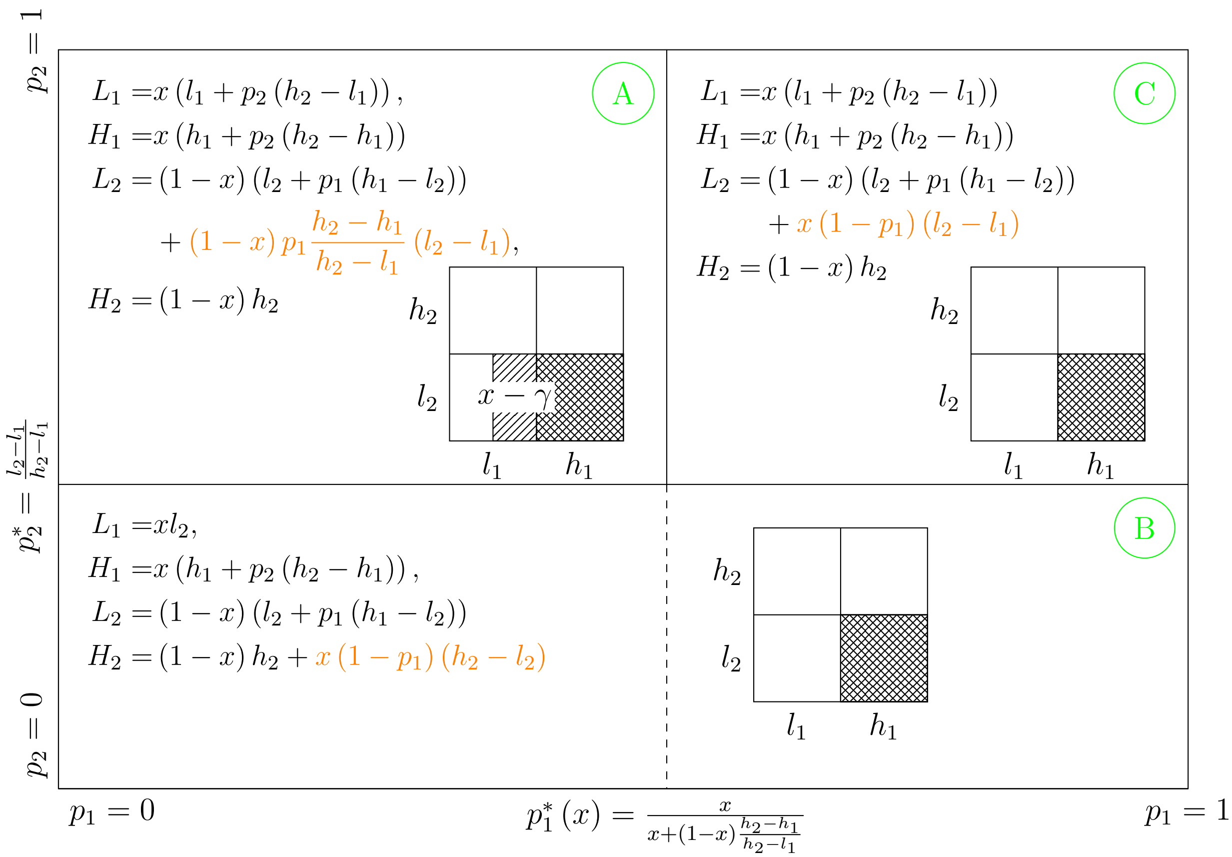

FIGURE 1. Mechanism a^1.

which implies that $\gamma = (1-x)\frac{p_1}{1-p_1}\frac{h_2-h_1}{h_2-l_1}$. (Notice that when $p_1 = p_1^*(x)$, γ becomes equal to x and type l_1 sells her entire share.) In order to verify interim efficiency, it is enough to notice that the payoffs solve a welfare maximization problem with properly chosen weights.[1]

- Zones B and C: if $p_1 \geq p_1^*(x)$ or $p_2 \leq p_2^*$, the allocation is ex post efficient, which implies interim efficiency of the payoffs.

It follows that $a^1(x)$ is interim efficient.

[1]More precisely, let $\Lambda_2 = p_2 - (1-p_2)\frac{l_2-l_1}{h_2-l_2}$. Consider the optimization problem:

$$\begin{aligned} &\max\, (1-\Lambda_2)\, L_2 + \Lambda_2 H_2 \\ &\quad \text{st. } (H_1, L_1, H_2, L_2) \in U(p_1, p_2),\, L_1 \geq L_1^*(p_2),\, H_1 \geq H_1^*(p_2). \end{aligned}$$

We can check that with such weights, any $\gamma \leq x$ is optimal. ???

4.2.2. *Implementation.* Mechanism $a^1(x)$ can be approximated by a sequence of finite games $a^\varepsilon(x)$ for small $\varepsilon \to 0$. Player 1 gets x (probability) share of the good and player 2 gets $1-x$ share. The players can trade their shares in the following game:

(1) Player 1 announces the price $\pi \in \{l_2, h_2\}$ per unit at which she is willing to sell her share of the good.
(2) If $\pi = h_2$, than, in order,
   (a) player 1 declares share $\gamma \leq x, \gamma = 0, \varepsilon, 2\varepsilon, ...$ that she is willing to sell at price $l_1 + \varepsilon$ per unit if player 2 agrees to sell his share at price $h_1 - \varepsilon$,
   (b) player 2 announces whether he agrees to sell his share at price $h_1 - \varepsilon$ per unit,
   (c) if player 2 agrees in the previous step, player 1 decides whether to buy player 2's share at price $h_2 - \varepsilon$ or sell γ units of her share at price $l_1 + \varepsilon$ per unit.
(3) The following allocations are implemented:
   (a) if $\pi = l_2$ or $\pi = h_2$ and player 2 does not agree to sell his share, player 2 gets the good with probability 1 and transfers $x\pi$ to player 1,
   (b) if $\pi = h_2$, player 2 agrees sell his share, and player 1 decides to sell γ units, player 2 gets the good with probability $1 - x + \gamma$ and transfers $\gamma(l_1 + \varepsilon)$ to player 1,
   (c) if $\pi = h_2$, player 2 agrees sell his share, and player 1 decides to buy player 2's share, player 2 gets the good with probability 0 and receives transfer $(1-x)(h_1 - \varepsilon)$ from player 1.

4.2.3. *Other properties of Theorem 1.* Leave for later.

(1) $a(x)$ is self 2-undominated: If it is 2-undominated by mechanism $U_1 m$.
(2) $a(x)$ is closed wrt. 2-revelation,
(3) $U_1 a(x)$ is closed wrt 2-revelation,
(4) $a(tx)$ is a 1-lower bound on $ta(x)$ for any $t \leq 1$,
(5) $a(1 - t(1-x))$ is strictly 2-larger than $t'a(x)$ for any $t' > t$,
(6) for each $\alpha \in [0,1]$ and each $x, x' \in [0,1]$, $a(\alpha x + (1-\alpha)x')$ is 1-lower bound on $\alpha U_1 a(x) + (1-\alpha) U_1 a(x')$,

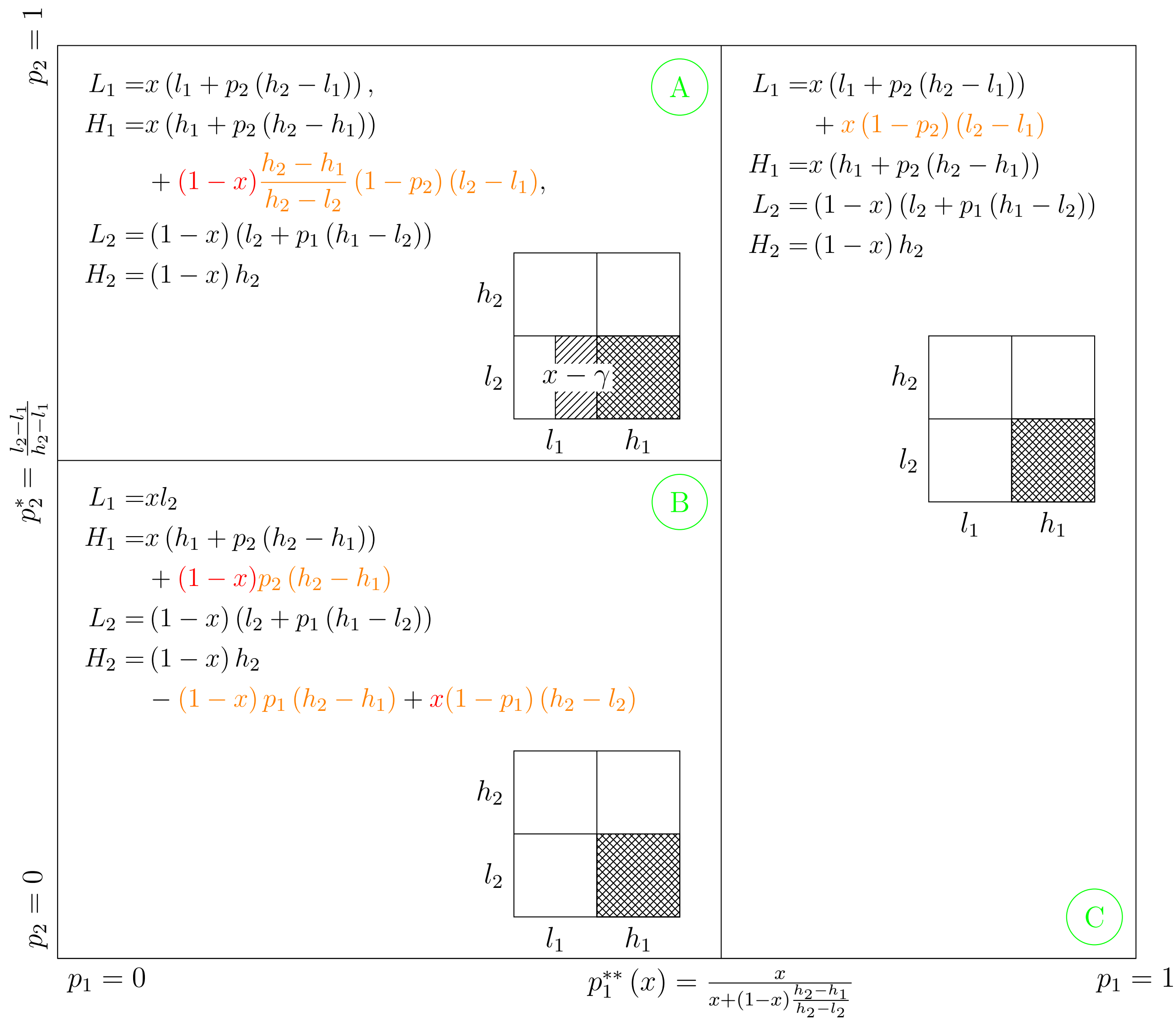

FIGURE 2. Mechanism a^2 - version II.

(7) $a(0)$ is 1-lower bound on RB.

4.3. **Player 2 payoff bounds.** The payoffs in mechanism $a(x)$ are described below. The payoffs on the boundaries are convex combination of continuous limits of payoffs in the neighboring areas. The shaded area describes the probability that player 1 gets the good conditionally on the realization of the

In each of the three cases, the payoffs are interim efficient:

- Zone A: is similar to Zone A in mechanism a^1, but with a modification that, when type l_1 faces type l_2, type l_1 offers a sale of γ units of the good at price l_2. The largest γ that does not threaten the truthtelling incentives of type h_2 is $(1-x)\frac{p_1}{1-p_1}\frac{h_2-h_1}{h_2-l_2}$. (Notice that when $p_1 = p_1^{**}(x)$, γ becomes equal to x and type l_1 sells her entire share.) In order to verify interim efficiency, it is enough to notice that the payoffs solve a welfare maximization problem with properly chosen weights.[2]
- Zones B and C: if $p_1 \geq p_1^{**}(x)$ or $p_2 \leq p_2^*$, the allocation is ex post efficient, which implies interim efficiency of the payoffs.

We have the following mechanism that virtually implements the above payoffs. Take $\varepsilon > 0$.

(1) First, player 1 declares the amount $\gamma \leq x$ of share of the good that she is willing to sell at price $l_1 + \varepsilon$ to type l_2 of player 2. She must choose $\gamma \in \{0, \varepsilon, 2\varepsilon, ...\}$.
(2) Upon observing γ, player 1 declares his type. Next, player 2 reveals her type.
(3) The following allocations are implemented:
    (a) if types are (l_1, l_2), player 2 gets the good with probability $1 - x + \gamma$ and transfers $\gamma(l_1 + \varepsilon)$ to player 1,
    (b) if types are (l_1, h_2), player 2 gets the good with probability 1 and transfers xh_2 to player 1,
    (c) if types are (h_1, l_2), player 2 gets the good with probability 0 and receives transfer $(1-x)h_1$ from player 1,
    (d) if types are (h_1, h_2), player 2 gets the good with probability 1 and transfers xh_2 to player 1.

[2] More precisely, let $\Lambda_2 = p_2 - (1-p_2)\frac{l_2-l_1}{h_2-l_2}$. The payoffs solve

$$\begin{aligned} &\max\,(1-\Lambda_2)\,L_2 + \Lambda_2 H_2 \\ &\text{st.}\,(H_1, L_1, H_2, L_2) \in U(p_1, p_2)\,, L_1 \geq L_1^*(p_2)\,, H_1 \geq H_1^*(p_2)\,. \end{aligned}$$

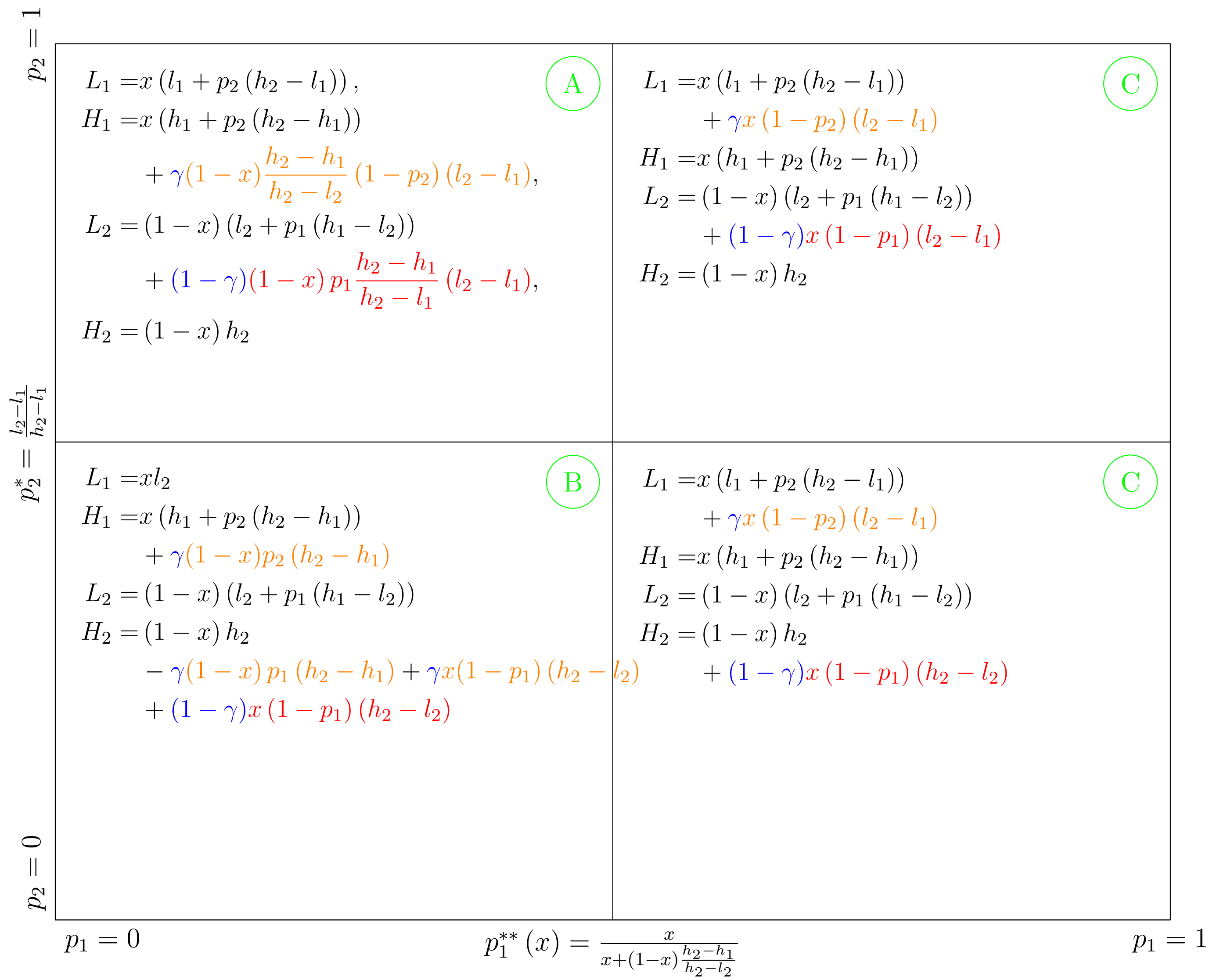

FIGURE 3. Mechanism a^2 - version II.

4.4. **Equilibria.** The payoffs in mechanism $a(x)$ are described below. The payoffs on the boundaries are convex combination of continuous limits of payoffs in the neighboring areas. The shaded area describes the probability that player 1 gets the good conditionally on the realization of the

In each of the three cases, the payoffs are interim efficient:

- Zone A: is similar to Zone A in mechanism a^1, but with a modification that, when type l_1 faces type l_2, type l_1 offers a sale of γ units of the good at price l_2. The largest γ that does not threaten the truthtelling incentives of type h_2 is $(1-x)\frac{p_1}{1-p_1}\frac{h_2-h_1}{h_2-l_2}$. (Notice that when $p_1 = p_1^{**}(x)$, γ becomes equal to x and type l_1 sells her entire share.) In order to verify interim efficiency, it is enough to notice that the payoffs solve a welfare maximization problem with properly chosen weights.[3]
- Zones B and C: if $p_1 \geq p_1^{**}(x)$ or $p_2 \leq p_2^*$, the allocation is ex post efficient, which implies interim efficiency of the payoffs.

We have the following mechanism that virtually implements the above payoffs. Take $\varepsilon > 0$.

(1) First, player 1 declares the amount $\gamma \leq x$ of share of the good that she is willing to sell at price $l_1 + \varepsilon$ to type l_2 of player 2. She must choose $\gamma \in \{0, \varepsilon, 2\varepsilon, ...\}$.
(2) Upon observing γ, player 1 declares his type. Next, player 2 reveals her type.
(3) The following allocations are implemented:
   (a) if types are (l_1, l_2), player 2 gets the good with probability $1 - x + \gamma$ and transfers $\gamma(l_1 + \varepsilon)$ to player 1,
   (b) if types are (l_1, h_2), player 2 gets the good with probability 1 and transfers xh_2 to player 1,
   (c) if types are (h_1, l_2), player 2 gets the good with probability 0 and receives transfer $(1-x)h_1$ from player 1,
   (d) if types are (h_1, h_2), player 2 gets the good with probability 1 and transfers xh_2 to player 1.

4.5. **Old equilibria.** In this subsection, we show that payoff bounds $a^i(\beta_i)$ are tight for large discount factors: For each i, there are equilibria of the bargaining game with payoffs close to $a^i(\beta_i)$.

---

[3]More precisely, let $\Lambda_2 = p_2 - (1-p_2)\frac{l_2-l_1}{h_2-l_2}$. The payoffs solve

$$\begin{aligned} &\max (1-\Lambda_2) L_2 + \Lambda_2 H_2 \\ &\text{st.} (H_1, L_1, H_2, L_2) \in U(p_1, p_2), L_1 \geq L_1^*(p_2), H_1 \geq H_1^*(p_2). \end{aligned}$$

We construct such equilibria in the following way. The equilibrium behavior alternates between two states, one for each player i. In state i, if player j makes an offer, he or she offers $b^{i,j}$. If player j offers $b \neq b^{i,j}$ and the offer is rejected, the behavior continues in state j. The expected payoff in mode i is equal to

$$b^i = \beta_i b^{i,i} + (1 - \beta_i) b^{i,-i}.$$

More precisely, we construct mechanisms $b^{i,i}$ (which we define below) and $b^{i,-i} = a^i(\delta\beta_i)$ for each i such that

- $b^{i,i}$ (and, as a result b^i) is $(1-\delta)$-close to $a^i(\beta_i)$,
- $b^{i,i} \subseteq MM_{-i}(\delta b^i, b^{i,i})$,
- $b^{i,-i} \subseteq U_{-i}MM_i(\delta b^{-i}, b)$ for each mechanism b: for each p, there is $q_{-i} \in \Delta T_{-i}$ and $u \in MM_i(\delta b^{-i}, b)(q_{-i}, p_i)$ so that $u \leq_{q_{-i}} b^{i,-i}(p)$,
- $b^{i,-i} \subseteq I_{-i}(U_i \delta b^i)$,

Mechanism $b^{i,i}$ is defined so that that

$$b_i^{i,i}(p)$$

Let $x = b^i \cup b^{-i}$. Then, the last two claims imply that

$$\begin{aligned} b^{i,-i} &\subseteq I_{-i}\left(U_i \delta b^i\right) \cap \bigcup_{\text{mechanism } b} U_{-i}MM_i\left(\delta b^{-i}, b\right) \\ &\subseteq I_{-i}(U_i \delta x) \cap \bigcup_{\text{mechanism } b} U_{-i}MM_i(\delta x, b) \\ &= IP_{-i}(\delta x). \end{aligned}$$

The first claim implies that

$$\begin{aligned} b^{i,i} &\subseteq MM_{-i}\left(\delta RB, b^{i,i}\right) \\ &\subseteq \end{aligned}$$

By induction, if (Note that operators U_i and I_i are monotonic and $x \subseteq U_i x$ and $x \subseteq I_i x$.) Then, the above claims imply that $b^i \in RB$ is $(1-\delta)$-close to $a^i(\beta_i)$ and

$$b^{i,j} \subseteq RB_j = IP_j(\delta RB) = MM_i\left(\left\{MM_{-i}\left(\left\{\delta b^i, b\right\}\right) : m \text{ is a mechanism}\right\}\right)$$

5. Equilibrium

5.1. **General results.** In this subsection, we provide a general sufficient conditions for an equilibrium in the game RB.

Suppose that a^1, a^2 are mechanisms such that

(1) Incentives to accept (IA): Let

$$a(p) = \beta a^1 + (1-\beta) a^2.$$

Then, for each $u \in a^i(p)$, there exists $u' \in \delta a$ such that $u_{-i} \geq u'_{-i}$.

(2) Interim efficiency (IE): for each $p = (p_i, p_{-i})$, each $q = (q_i, q_{-i})$, each $e \in IC(q_i, p_{-i})$, each $u \in a^i(p)$, and each $u' \in \delta a(q)$, if $e_i(t_i) \geq u(t_i)$ with a strict inequality for some t_i, then there is t_{-i} such that such that $e_{-i}(t_{-i}) < u'_{-i}(t_{-i})$.

To explain, suppose that mechanisms a^i are single-valued. The first property implies that agent $-i$ prefers to accept mechanism a^irather than wait for the next period. The second property says that payoff vector $(a^i_i(p), \delta a_{-i}(q))$ is (q_i, p_{-i})-interim efficient.

Lemma 8. *Suppose that a^1, a^2 are two mechanisms satisfying the above properties. Then, $a(p)$ are equilibrium payoffs in the bargaining game with prior p ($a(p) \subseteq RB(p)$).*

Proof. Fix prior p and payoffs $u = \beta u^1 + (1-\beta) u^2$, where $u^i \in a^i(p)$. We construct an equilibrium of the bargaining game with such payoffs. On the equilibrium path, if player i is proposer, mechanism a^i is offered and it is accepted. In the mechanism, when player i is proposer, the expected payoffs are given by u^i. The expected payoffs prior to the choice of the proposer are equal to u.

We will show that a^i is accepted by $-i$ when offered. In fact, suppose that the beliefs p_i remain unchanged, and the beliefs p_{-i} after accept or reject decision remain unchanged as well. If the proposal is accepted, equilibrium with payoffs u is implemented. If the mechanism is rejected, the continuation equilibrium payoff is $u' \in \delta a(p)$ such that $u_{-i} \geq u'_{-i}$. (We can find such u' due to the property 1.). Hence, accepting is a best response for player $-i$.

Next, we show that player i does not benefit from offering any other mechanism $M \neq a^i$. Define $m = GM\left(-i, \{M, \delta RB\}\right)$ as the Accept-Reject game played by player $-i$.

Suppose first that there exists q_i and $e \in m\left(q_i, p_{-i}\right)$ such that $e_i\left(t\right) \leq u_i^i\left(t\right)$ for each $t \in T_i$. Then, upon offering M by player i, update beliefs to q_i, and continue with payoffs e. Such an outcome does not lead to profitable deviation.

Next, suppose that for each q_i, each $e \in m\left(q_i, p_{-i}\right)$, there is $t \in T_i$ such that

$$e_i\left(t_i\right) > u_i^i\left(t\right).$$

Because m is Kakutani, there is $\varepsilon > 0$ such that the above inequality above can be replaced by $e_i\left(t\right) \geq u_i^i\left(t\right) + \varepsilon$. Let $D_\varepsilon = \left\{x \in \mathbb{R}^{T_i} : \max_{t \in T_i} x\left(t\right) \geq \varepsilon\right\}$. Let $f : D_\varepsilon \to \Delta T_i$ be a continuous function given by $f\left(t|x\right) = \frac{\max(0, x(t))}{\sum \max(0, x(s))}$. Construct correspondences:

- $m^1 : \Delta T_i \to D_\varepsilon$ so that $m^1\left(q_i\right) = \left\{e_i - u_i^i : e \in m\left(q_i, p_{-i}\right)\right\} \subseteq \mathcal{U}_i$. This is a correspondence of equilibrium payoffs normalized by subtracting $u_i^i\left(t\right)$. Because equilibrium payoffs m are Kakutani, m^1 is Kakutani as well.
- $m^2 = f \circ m^1$ so that $m^2\left(q_i\right) = \left\{f\left(x\right) : x \in m^1\left(q_i\right)\right\}$.

As a composition of a Kakutani correspondence and a continuous function, m^2 has a fixed point q_i (??? - we can prove it by, say, finding a continuous function that approximates m^2 (using Lemma 9 from Peski (21)) and then show the convergence of fixed points ???). There exists $e \in m\left(q_i, p_{-i}\right)$ such that $q_i\left(t\right) > 0$ implies $e_i\left(t\right) \geq u_i^i\left(t\right) + \varepsilon$. Let q_{-i} and $u' \in \delta a\left(q_i, q_{-i}\right)$ be the beliefs and payoffs following the decision to reject the mechanism M in equilibrium e^*. Because of property 2, there exists $t_{-i} \in T_{-i}$ such that $e_{-i}\left(t_{-i}\right) < \delta u'_{-i}\left(t_{-i}\right)$. Hence, $e_{-i}\left(t_{-i}\right)$ cannot be the equilibrium payoff of type t_{-i}. The contradiction ends the proof. □

5.2. **Two types.** We verify that the pair of mechanisms $a^1 = RMRC_{l_1}\left(1 - \delta\left(1 - \beta\right)\right)$ and $a^2 = RMRC_{l_1}\left(\delta\beta\right)$ satisfy the conditions of Lemma 8. For any b, $RMRC_{l_1}\left(b\right)$ payoffs are, for $X = L, H$,

$$X_1 = b f_1^X\left(p_2\right),$$
$$X_2 = \left(1 - b\right) f_2^X\left(p_1\right) + b \mathbf{1}_{p_2 \leq p^*} g^X\left(p_1\right)$$

for some functions f_i^X, g^X. We start with incentives to accept. Because

$$\beta\left(1-\delta\left(1-\beta\right)\right)+\left(1-\beta\right)\left(\beta\delta\right)=\beta, \tag{2}$$

$$\left(1-\beta\right)\left(1-\delta\beta\right)+\beta\left(1-\beta\right)\delta=1-\beta \tag{3}$$

we immediately obtain that, for $X=L,H$,

$$\begin{aligned}a_1^2\left(X|p\right)=\delta\beta f_1^X\left(p_2\right)&=\delta\left(\beta\left(1-\delta\left(1-\beta\right)\right)+\left(1-\beta\right)\left(\beta\delta\right)\right)f_1^X\left(p_2\right)\\&=\delta\left(\beta a_1^1\left(X|p\right)+\left(1-\beta\right)a_1^2\left(X|p\right)\right)=\delta a_1\left(X|p\right).\end{aligned}$$

Additionally, because (3) and

$$1-\delta\left(1-\beta\right)\geq\left(\beta\left(1-\delta\left(1-\beta\right)\right)+\left(1-\beta\right)\delta\beta\right)$$

we have

$$\begin{aligned}a_2^1\left(X|p\right)=&\delta\left(1-\beta\right)f_2^X\left(p_1\right)+\left(1-\delta\left(1-\beta\right)\right)\mathbf{1}_{p_2\geq p^*}g^X\left(p_1\right)\\\geq&\delta\left[\left(1-\beta\right)\left(1-\delta\beta\right)+\beta\left(1-\beta\right)\delta\right]f_2^X\left(p_1\right)\\&+\delta\left[\beta\left(1-\delta\left(1-\beta\right)\right)+\left(1-\beta\right)\delta\beta\right]\mathbf{1}_{p_2\geq p^*}g^X\left(p_1\right)\\=&\beta\delta\left[\left(1-\beta\right)\delta f_2^X\left(p_1\right)+\left(1-\delta\left(1-\beta\right)\right)\mathbf{1}_{p_2\geq p^*}g^X\left(p_1\right)\right]\\&+\left(1-\beta\right)\delta\left[\left(1-\delta\beta\right)f_2^X\left(p_1\right)+\delta\beta\mathbf{1}_{p_2\geq p^*}g^X\left(p_1\right)\right]\\=&\delta\left(\beta a_2^1\left(X|p\right)+\left(1-\beta\right)a_2^2\left(X|p\right)\right)=\delta a_2\left(X|p\right).\end{aligned}$$

This shows that the property 1 holds.

We verify the property 2. For each i, and each $r\in\Delta T$ the payoffs $a_j^i\left(r\right)$ depend only on r_{-i}. Hence,

$$\left(a_i^i\left(p\right),\delta a_{-i}\left(q\right)\right)=\left(a_i^i\left(q_i,p_{-i}\right),\delta a_{-i}\left(q_i,p_{-i}\right)\right).$$

Moreover, due to (3), if $p_2\geq p_2^*$,

$$\left(a_i^i\left(q_i,p_{-i}\right),\delta a_{-i}\left(q_i,p_{-i}\right)\right)=a^i\left(p_i,q_{-i}\right),$$

which is interim-efficient. If $p_2 \leq p_2^*$,

$$\left(\delta a_1\left(p_1, q_2\right), a_2^2\left(p_1, q_2\right)\right) = \left(a_1^2\left(p_1, q_2\right), a_2^2\left(p_1, q_2\right)\right)$$

is interim efficient. Finally, $\left(a_1^1\left(p\right), \delta a_2\left(q\right)\right)$ solves the interim optimization for $\Lambda_1 = q_1$ and any $\Lambda_2 \leq p_2$, including $\Lambda_2 = 0$,

- if $p_2 \leq p_2^*$ and $p_1 \leq p_1^*\left(b\right)$, the payoffs are equal to
$$= \left(a_1^1\left(q_1, p_2\right), a_2^1\left(q_1, p_2\right)\right),$$
which are interim efficient for $\Lambda_1 = p_1$ and $\Lambda_2 = p_2 - \left(1 - p_2\right)\frac{l_2 - l_1}{h_1 - l_1}$,
- if $p_2 \leq p_2^*$ and $p_1 \geq p_1^*\left(b\right)$, the payoffs solve the interim optimization for $\Lambda_1 = q_1$ and any $\Lambda_2 \geq p_2$, including $\Lambda_2 = 1$.

Due to the above calculations, for each $q, p \in \Delta T$,

$$\left(\delta a_1\left(q\right), a_2^2\left(p\right)\right) = \left(\delta a_1\left(p_1, q_2\right), a_2^2\left(p_1, q_2\right)\right) = \left(a_1^2\left(p_1, q_2\right), a_2^2\left(p_1, q_2\right)\right),$$

where the first inequality is due to the fact that payoffs of player i in the $RMRC_{l_1}$ mechanism do not depend on p_i beliefs. The last vector of payoffs is (p_1, q_1)-interim efficient due to the general characterization of efficiency in the $RMRC_{l_1}$ mechanism.

Next, notice that

$$\left(a_1^1\left(p\right), \delta a_2\left(q\right)\right) = \left(a_1^1\left(q_1, p_2\right), \delta a_2\left(q_1, p_2\right)\right).$$

Consider the following cases:

- if $p_2 \leq p_2^*$, the above payoffs solve the interim optimization for $\Lambda_1 = q_1$ and any $\Lambda_2 \leq p_2$, including $\Lambda_2 = 0$,
- if $p_2 \leq p_2^*$ and $p_1 \leq p_1^*\left(b\right)$, the payoffs are equal to
$$= \left(a_1^1\left(q_1, p_2\right), a_2^1\left(q_1, p_2\right)\right),$$
which are interim efficient for $\Lambda_1 = p_1$ and $\Lambda_2 = p_2 - \left(1 - p_2\right)\frac{l_2 - l_1}{h_1 - l_1}$,
- if $p_2 \leq p_2^*$ and $p_1 \geq p_1^*\left(b\right)$, the payoffs solve the interim optimization for $\Lambda_1 = q_1$ and any $\Lambda_2 \geq p_2$, including $\Lambda_2 = 1$.

In all cases, the payoffs are interim efficient.

6. Conclusions

Appendix A. Interim efficiency

The goal of this section is to characterize optimal payoff vectors in the trading mechanism with two types. Denote

$$\Delta_i = h_i - l_i \text{ for each } i \text{ and } R = l_2 - l_1.$$

For each belief $p = (p_1, p_2)$, denote $p_i(h) = p_i = 1 - p_i(l)$.

An allocation is defined as probabilities $(q_i^{x_i y_{-i}})_{i,x_i,y_i}$ (where q_i^{xy} is interpreted as a conditional probability that player i gets the good conditionally on player i type being x and player $-i$ type being y) and transfers $t_i^l = t_i$ and $t_i^h = t_i + \Delta t_i$ such that two condition hold

- the feasibility condition

$$q_i^{xy} + q_{-i}^{yx} = 1 \text{ for each } x, y, \text{ and} \tag{4}$$

- ex ante budget balance:$\sum_i t_i + p_i \Delta t_i = 0$.

Let$q_i^x = \frac{1}{p_i(x)} \left(q_i^{xl} p_{-i} + q_i^{xh} p_{-i} \right)$.

The allocation is incentive compatible, if, $q_i^l \leq q_i^h$ and $\Delta t_i \in \left[q_i^l \Delta_i, q_i^h \Delta_i \right]$. We say that $IC_i(h)$ constraint is binding if $\Delta t_i = q_i^l \Delta_i$ and $IC_i(l)$ constraint is binding if $\Delta t_i = q_i^h \Delta_i$. If $\Delta q_i = 0$ then both constraints for player i are biding in any incentive compatible mechanism.

Theorem 2. *A payoff vector (L_i, H_i) is Λ-optimal under beliefs p if and only if it is the payoff vector in incentive compatible allocation (q_i, t_i) such that the following conditions are satisfied:*

(1) Optimal allocation: $q_2^{lh} = 1 - q_1^{lh} = 0$, $q_2^{hl} = 1 - q_1^{hl} = 1$,

$$q_2^{ll} = 1 - q_1^{ll} = \begin{cases} (1 - p_i)(1 - p_{-i}) & \Lambda_2 > p_2 - (1 - p_2) \frac{l_2 - l_1}{h_2 - l_2} \\ 0 & \Lambda_2 < p_2 - (1 - p_2) \frac{l_2 - l_1}{h_2 - l_2} \end{cases},$$

and

$$q_2^{hh} = 1 - q_1^{hh} = \begin{cases} p_1 p_2 & R + \max\left(\frac{1}{p_2}\Lambda_2, 1\right)\Delta_2 > \frac{1}{p_1}\Lambda_1\Delta_1, \\ 0 & R + \max\left(\frac{1}{p_2}\Lambda_2, 1\right)\Delta_2 < \frac{1}{p_1}\Lambda_1\Delta_1. \end{cases},$$

(2) *Incentive constraints: For each player* i,

   (a) *if* $\Lambda_i > p_i$, *then* $IC_i(l)$ *constraint is binding,*

   (b) *if* $\Lambda_i < p_i$, *then* $IC_i(h)$ *constraint is binding,*

A.1. **Application: Robust efficiency of mechanisms a^i.** TBA

A.2. **Proof.** Denote the expected payoffs of each type as

$$\begin{aligned} L_i =& q_i^l l_i - t_i, \\ H_i =& q_i^h h_i - t_i - l_i\left(q_i^h - q_i^l\right) - \alpha_i \Delta_i \Delta q_i \\ =& L_i + \Delta_i q_i^h - \alpha_i \Delta_i \Delta q_i, \end{aligned}$$

where we denote $\Delta q_i = q_i^h - q_i^l$.

The budget balance implies that the expected welfare must be equal to the expected utility from allocations:

$$\sum_i L_i + \sum p_i\left(H_i - L_i\right) = \sum\left(1 - p_i\right) q_i^l l_i + \sum p_i q_i^h h_i.$$

After substitutions and some algebra, we get

$$\begin{aligned} \sum_i L_i &= \sum_i \left[\left(1 - p_i\right) q_i^l l_i + p_i q_i^h h_i - p_i \Delta_i q_i^h + p_i \alpha_i \Delta_i \Delta q_i\right] \\ &= \sum_i \left[q_i^l l_i + p_i \Delta q_i\left(l_i + \alpha_i \Delta_i\right)\right]. \end{aligned} \tag{5}$$

Consider the welfare maximization problems with weights $\Lambda_i \in [0, 1]$:

$$\max \sum_i \left(1 - \Lambda_i\right) L_i + \Lambda_i H_i \text{ st. feasibility and IC constraints.}$$

Using the formula (5), the objective function can be rewritten as

$$\begin{aligned}
&\sum_i (1-\Lambda_i) L_i + \Lambda_i H_i \\
=&\sum_i \left[q_i^l l_i + p_i \Delta q_i (l_i + \alpha_i \Delta_i)\right] + \sum_i \Lambda_i \left(\Delta_i q_i^l + (1-\alpha_i)\Delta_i \Delta q_i\right) \\
=&\sum_i [l_i + \Lambda_i \Delta_i] q_i^l + \sum_i (p_i h_i + (1-\alpha_i)(\Lambda_i - p_i)\Delta_i)\Delta q_i .
\end{aligned}$$

Thus, if $\Delta q_i > 0$, then $\alpha_i = 0$ if $\Lambda_i > p_i$ and $\alpha_i = 1$ if $\Lambda_i < p$. Conversely, if $\Delta q_i = 0$, the value α_i does not matter. Further, the above is equal to

$$\begin{aligned}
=&\sum_i [l_i + \Lambda_i \Delta_i] q_i^l + \sum_i (p_i l_i + \Lambda_i \Delta_i - \alpha_i (\Lambda_i - p_i)\Delta_i)\Delta q_i \\
=&\sum_i [(1-p_i) l_i + \alpha_i (\Lambda_i - p_i)\Delta_i] q_i^l + \sum_i (p_i (l_i + \alpha_i \Delta_i) + (1-\alpha_i)\Lambda_i \Delta_i) q_i^h \\
=&\sum_i \left[l_i + \frac{1}{1-p_i}\alpha_i (\Lambda_i - p_i)\Delta_i\right]\left(q_i^{ll} + q_i^{lh}\right) + \sum_i \left(l_i + \alpha_i \Delta_i + \frac{1}{p_i}(1-\alpha_i)\Lambda_i \Delta_i\right)\left(q_i^{hl} + q_i^{hh}\right).
\end{aligned}$$

Recalling the feasibility conditions, the above is equal to

$$\begin{aligned}
=&q_2^{ll}\left(R + \frac{1}{1-p_2}\alpha_2 (\Lambda_2 - p_2)\Delta_2 - \frac{1}{1-p_1}\alpha_1 (\Lambda_1 - p_1)\Delta_1\right) \\
&+ q_2^{lh}\left(R + \frac{1}{1-p_2}\alpha_2 (\Lambda_2 - p_2)\Delta_2 - \alpha_1 \Delta_1 - \frac{1}{p_1}(1-\alpha_1)\Lambda_1 \Delta_1\right) \\
&+ q_2^{hl}\left(R + \alpha_2 \Delta_2 + \frac{1}{p_2}(1-\alpha_2)\Lambda_2 \Delta_2 - \frac{1}{1-p_1}\alpha_1 (\Lambda_1 - p_1)\Delta_1\right) \\
&+ q_2^{hh}\left(R + \alpha_2 \Delta_2 + \frac{1}{p_2}(1-\alpha_2)\Lambda_2 \Delta_2 - \alpha_1 \Delta_1 - \frac{1}{p_1}(1-\alpha_1)\Lambda_1 \Delta_1\right) \\
&+ l_1 + \Lambda_1 \Delta_1 .
\end{aligned}$$

Consider the following cases:

- $\Lambda_i > p_i$ for each i. Then, $\alpha_i = 0$ for each i (which means that the low types are indifferent between telling the truth and reporting the high type), and the objective becomes

$$\begin{aligned}
=&q_2^{ll}R\\
&+q_2^{lh}\left(R-\frac{1}{p_1}\Lambda_1\Delta_1\right)\\
&+q_2^{hl}\left(R+\frac{1}{p_2}\Lambda_2\Delta_2\right)\\
&+q_2^{hh}\left(R+\frac{1}{p_2}\Lambda_2\Delta_2-\frac{1}{p_1}\Lambda_1\Delta_1\right)\\
&+l_1+\Lambda_1\Delta_1.
\end{aligned}$$

It follows that $q_2^{lh}=0$, $q_2^{ll}=(1-p_1)(1-p_2)$, and $q_2^{hl}=p_2(1-p_1)$.

- The value of q_2^{hh} depends on the sign of $R+\frac{1}{p_2}\Lambda_2\Delta_2-\frac{1}{p_1}\Lambda_1\Delta_1$, which is always positive if $R+\Delta_2>\frac{1}{p_1}\Delta_1$, but it can be negative for some values of Λ_1 otherwise. If the sign is negative, then $\Delta q_2=0$, whcih means that α_2 could be whatever.

- $\Lambda_1=p_1$ and $\Lambda_2>p_2$. Then, $\alpha_2=0$ and α_1 is anything. The objective becomes

$$\begin{aligned}
=&q_2^{ll}(R)\\
&+q_2^{lh}(R-\Delta_1)\\
&+q_2^{hl}\left(R+\frac{1}{p_2}\Lambda_2\Delta_2\right)\\
&+q_2^{hh}\left(R+\frac{1}{p_2}\Lambda_2\Delta_2-\Delta_1\right)\\
&+l_1+\Lambda_1\Delta_1.
\end{aligned}$$

which means $q_2^{lh}=0$, $q_2^{ll}=(1-p_1)(1-p_2)$, $q_2^{hl}=p_2(1-p_1)$ and $q_2^{hh}=p_1p_2$.

- $\Lambda_1 < p_1$ and $\Lambda_2 > p_2$. Then, $\alpha_2 = 0$ and $\alpha_1 = 1$. The objective becomes

$$
\begin{aligned}
=& q_2^{ll}\left(R + \frac{1}{1-p_1}\left(p_1 - \Lambda_1\right)\Delta_1\right) \\
&+ q_2^{lh}\left(R - \Delta_1\right) \\
&+ q_2^{hl}\left(R + \frac{1}{p_2}\Lambda_2\Delta_2 + \frac{1}{1-p_1}\left(p_1 - \Lambda_1\right)\Delta_1\right) \\
&+ q_2^{hh}\left(R + \frac{1}{p_2}\Lambda_2\Delta_2 - \Delta_1\right) \\
&+ l_1 + \Lambda_1\Delta_1.
\end{aligned}
$$

  which means $q_2^{lh} = 0$, $q_2^{ll} = (1-p_1)(1-p_2)$, $q_2^{hl} = p_2(1-p_1)$ and $q_2^{hh} = p_1 p_2$.
- $\Lambda_1 > p_1$ and $\Lambda_2 = p_2$. Then, $\alpha_1 = 0$ and α_2 is whatever. The objective becomes

$$
\begin{aligned}
=& q_2^{ll}\left(R\right) \\
&+ q_2^{lh}\left(R - \frac{1}{p_1}\Lambda_1\Delta_1\right) \\
&+ q_2^{hl}\left(R + \Delta_2\right) \\
&+ q_2^{hh}\left(R + \Delta_2 - \frac{1}{p_1}\Lambda_1\Delta_1\right) \\
&+ l_1 + \Lambda_1\Delta_1.
\end{aligned}
$$

  It follows that $q_2^{lh} = 0$, $q_2^{ll} = (1-p_1)(1-p_2)$, $q_2^{hl} = p_2(1-p_1)$.
  - The value of q_2^{hh} depends on the sign of $R + \Delta_2 - \frac{1}{p_1}\Lambda_1\Delta_1$, which is always positive if $R + \Delta_2 > \frac{1}{p_1}\Delta_1$, but it can be negative for some values of Λ_1 otherwise. If the sign is negative, then $\Delta q_2 = 0$, which means that α_2 could be any.

- $\Lambda_i = p_i$ for each i: The objective becomes

$$\begin{aligned}&= q_2^{ll}\left(R\right)\\&\quad + q_2^{lh}\left(R-\Delta_1\right)\\&\quad + q_2^{hl}\left(R+\Delta_2\right)\\&\quad + q_2^{hh}\left(R+\Delta_2-\Delta_1\right)\\&\quad + l_1 + \Lambda_1\Delta_1.\end{aligned}$$

which means $q_2^{lh} = 0$, $q_2^{ll} = (1-p_1)(1-p_2)$, $q_2^{hl} = p_2(1-p_1)$ and $q_2^{hh} = p_1 p_2$.

- $\Lambda_1 < p_1$ and $\Lambda_2 = p_2$ for each i: Then, $\alpha_1 = 1$. The objective becomes

$$\begin{aligned}&= q_2^{ll}\left(R + \frac{1}{1-p_1}\left(p_1-\Lambda_1\right)\Delta_1\right)\\&\quad + q_2^{lh}\left(R-\Delta_1\right)\\&\quad + q_2^{hl}\left(R+\Delta_2+\frac{1}{1-p_1}\left(p_1-\Lambda_1\right)\Delta_1\right)\\&\quad + q_2^{hh}\left(R+\Delta_2-\Delta_1\right)\\&\quad + l_1 + \Lambda_1\Delta_1.\end{aligned}$$

which means $q_2^{lh} = 0$, $q_2^{ll} = (1-p_1)(1-p_2)$, $q_2^{hl} = p_2(1-p_1)$ and $q_2^{hh} = p_1 p_2$.

- $\Lambda_1 > p_1$ and $\Lambda_2 < p_2$ for each i: Then, $\alpha_1 = 0, \alpha_2 = 1$. The objective becomes

$$\begin{aligned}&= q_2^{ll}\left(R - \frac{1}{1-p_2}\left(p_2-\Lambda_2\right)\Delta_2\right)\\&\quad + q_2^{lh}\left(R - \frac{1}{1-p_2}\left(p_2-\Lambda_2\right)\Delta_2 - \frac{1}{p_1}\Lambda_1\Delta_1\right)\\&\quad + q_2^{hl}\left(R+\Delta_2\right)\\&\quad + q_2^{hh}\left(R+\Delta_2-\frac{1}{p_1}\Lambda_1\Delta_1\right)\\&\quad + l_1 + \Lambda_1\Delta_1.\end{aligned}$$

which means $q_2^{lh} = 0$, $q_2^{hl} = p_1(1-p_2)$ and $q_2^{hh} = p_1 p_2$.

- The value of q_2^{ll} depends on the sign of $R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2$, which is always positive if we are in the efficient case and it can be negative for some choice of Λ_2 otherwise.
- The value of q_2^{hh} depends on the sign of $R + \Delta_2 - \frac{1}{p_1}\Lambda_1\Delta_1$, which is always positive if $R + \Delta_2 > \frac{1}{p_1}\Delta_1$, but it can be negative for some values of Λ_1 otherwise. If the sign is negative, then $\Delta q_2 = 0$, whcih means that α_2 could be whatever (?).

- $\Lambda_1 = p_i$ and $\Lambda_2 < p_2$, which means$\alpha_2 = 1$. The objective becomes

$$
\begin{aligned}
=&q_2^{ll}\left(R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2\right)\\
&+ q_2^{lh}\left(R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2 - \Delta_1\right)\\
&+ q_2^{hl}(R + \Delta_2)\\
&+ q_2^{hh}(R + \Delta_2 - \Delta_1)\\
&+ l_1 + \Lambda_1\Delta_1.
\end{aligned}
$$

which means $q_2^{lh} = 0$, $q_2^{hl} = p_2(1 - p_1)$ and $q_2^{hh} = p_1 p_2$.

- The value of q_2^{ll} depends on the sign of $R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2$, which is always positive if we are in the efficient case and it can be negative for some choice of Λ_2 otherwise.

- $\Lambda_i < p_i$, which means$\alpha_1 = \alpha_2 = 1$. The objective becomes

$$
\begin{aligned}
=&q_2^{ll}\left(R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2 + \frac{1}{1-p_1}(p_1 - \Lambda_1)\Delta_1\right)\\
&+ q_2^{lh}\left(l_2 - h_1 - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2\right)\\
&+ q_2^{hl}\left(h_2 - l_1 + \frac{1}{1-p_1}(p_1 - \Lambda_1)\Delta_1\right)\\
&+ q_2^{hh}(h_2 - h_1)\\
&+ l_1 + \Lambda_1\Delta_1.
\end{aligned}
$$

In maximum, the payoff assumptions (**??**) and the constraints imply that $q_2^{hh} = p_1 p_2$ and $q_2^{hl} = p_2(1-p_1)$ and $q_2^{lh} = 0$.

– The value of q_2^{ll} depends on the sign of $R - \frac{1}{1-p_2}(p_2 - \Lambda_2)\Delta_2,$ which is always positive if we are in the efficient case and it can be negative for some choice of Λ_2 otherwise.