Notes on Mechanism Design and Contract

(focusing on the case of two players, complete information, and non-durable trading opportunities)

Joel Watson

August 2010 version

1 Mechanism Design Problems: The Basic Notion

Mechanism design is the mathematical analysis of strategic settings from the point of view that:

- A group of players will interact by playing a game,
- Elements of the game are fixed,
- Other elements of the game can be designed (that is, chosen) in advance by either a third party or the players themselves.

One way to conceptualize the designed aspects of the game is that the entire game is fixed but includes a non-strategic player (one whose preferences are not modeled) whom we call the *external enforcer* or third party. In this interpretation, the design problem amounts to selecting the behavioral rule (strategy) that the external enforcer will use. Another way to conceptualize a mechanism-design problem is that there is a prespecified set of games from which we can select; typically, the games share some structure that is given by real-world constraints. These are very general and abstract ways of thinking about mechanism design. We naturally add an economic objective for the problem.

Typical Structure in the Literature

Much of the literature on mechanism-design theory focuses on a small number of specific settings with a great deal of structure. In most cases, the game to be played has three phases that we can describe as follows:

1. **Initial Phase**, where parameters are exogenously determined and the players obtain information;

2. **Communication Phase**; and

3. **Outcome Phase**, where productive activity occurs (yielding the productive/physical outcome).

The first phase has a fixed structure that is exogenously given. The parameters are generally payoff-relevant in that they affect the players' payoffs. The external enforcer's information is also specified. In many interesting applications, the external enforcer does not observe the parameters that arise in the first phase.

The third phase also has a fixed structure, which is a specification of feasible productive actions for the parties. In some applications, the external enforcer takes an action in the outcome phase, in addition to or instead of the regular players taking actions. The external enforcer's productive action in the outcome phase (if any) is called a *public action*, whereas the players' productive actions are called *individual actions*. In general, the productive activity can include sequential moves.

The second phase of the game is a communication protocol that can be designed. The communication protocol specifies how the players and the external enforcer can communicate. The third party may serve as a mediator in the communication phase, receiving messages from the players and sending messages back. Communication activity is assumed to be payoff irrelevant.

There are two aspects of design in this description. First, we have the communication protocol. For instance, the protocol could specify that player 1 send a message to player 2, who will then send a message to the external enforcer. As another example, the protocol could specify that the players simultaneously send messages to the external enforcer. The second aspect of design is to specify the strategy of the external enforcer. This involves describing how the external enforcer is to behave in the communication phase and in the third phase as well (if there is a public action in the third phase). The specification of these two design elements is called a *mechanism*. A standard *mechanism-design problem* is the analysis of mechanisms and rational behavior with specific economics objectives in mind.

Another required ingredient for the analysis of a mechanism-design problem is some notion of rational behavior for the players — that is, a prediction for how the players will behave. The appropriate solution concept depends on the application.

Notation for the Fundamentals

Here is some notation that describes the fundamental elements of mechanism-design problems, at a moderate level of detail. The mechanism-design problems described above have the following fixed elements:

- A set of players $1, 2, \ldots, n$,
- A set of *states* Θ representing payoff-relevant events that occur in the initial phase (before the strategic interaction to be studied),
- A description of the information that the players receive about the state (in probability terms),
- A set of feasible productive outcomes D which gives the possible combination of productive actions in the third phase, and
- A specification of the players' preferences, which are given by a utility function $\mu : D \times \Theta \to \mathbb{R}^n$.

Note that the payoffs are a function of only the state $\theta \in \Theta$ and the productive outcome $d \in D$.

Relative to these fundamentals, the elements that can be designed are:

- A protocol for communication among the players and the third party and
- A behavioral rule for the third party (who is committed to it), which includes how to communicate with the players and what public action to take.

Variations

There are many classes of mechanism-design problems that fit within the general model described above. These are distinguished by variations in the structure of the three phases. For example, different cases for the initial phase are:

- **Complete but unverifiable information** – Players jointly observe the state but the external enforcer does not observe it.
- **Incomplete information** – Players learn different aspects of the state (different signals) and thus do not generally have the same information.

In the latter case, there is also the issue of whether the players have a *common prior*, which means a shared understanding of what generates the state and each player's signal (in probability terms).

Different cases for the outcome phase are:

- **The public-action setting** – Players have no productive actions and the external enforcer selects d (a "Bayesian collective choice problem" in Myerson's terminology).
- **The pure individual-action setting** – Players have productive actions and there is no public action.
- **The (general) individual-action setting** – Players have productive actions and there is also a public action.

Furthermore, one may classify mechanisms in terms of the structure they impose on the communication phase:

- **Public messages** – Each party sends a message that everyone observes.
- **Private messages** – Players send private messages to the external enforcer and/or one another.
- **Correlation device/mediation** – Players send private messages to the external enforcer, who then sends conditional private messages back to the players.

Outcomes as State-Contingent Payoffs

For some mechanism-design problems, a useful way to represent a productive outcome is in terms of its payoff implications as a function of the state. For example, consider any public-action setting. Since d is chosen by the external enforcer, the mechanism dictates how it is to be conditioned on the messages sent earlier. This productive outcome implies a payoff vector $\mu(d, \theta)$ in state θ, a payoff vector $\mu(d, \theta')$ in state θ', and so on. Thus, d defines a *payoff outcome* function $w : \Theta \to \mathbb{R}^n$ that is given by

$$w(\theta) \equiv \mu(d, \theta) \tag{1}$$

for every $\theta \in \Theta$. We will often call such a function an "outcome" and we let W be the set of outcome functions:

$$W \equiv \{w : \Theta \to \mathbb{R}^n \mid \text{there exists } d \text{ such that (1) holds for all } \theta \in \Theta\}.$$

We can then think of the mechanism as specifying w as a function of messages, rather than specifying d as a function of messages. To avoid confusion, remember that w itself is a function of the state. Thus, the mechanism selects a particular w given whatever messages were sent, and then this w yields a payoff vector $w(\theta)$ in state θ. So, to belabor this a bit (but it's worth doing), different messages lead the mechanism to select different w functions (representing different d's). The messages leading to any particular w can arise in any of the states; if the realized state is θ then the payoff vector is $w(\theta)$.

Representing productive outcomes by using payoff outcome functions can also be appropriate when there are individual productive actions, as long as the mechanism-design problem has complete information and the messages are public. Here are the details of the construction. Let d^{i} denote the individual productive actions of the players and let d^{p} denote the public action, so we can write $d = (d^{\mathrm{i}}, d^{\mathrm{p}})$. It may be that d^{i} or d^{p} is null, meaning that either the players or the external enforcer have no productive choice. Also, these could describe strategies in a sequential-move setting. From the public messages sent in the communication phase, the mechanism prescribes a public action d^{p}. Because the players have common knowledge about the state and the messages (remember that we are assuming complete information and public messages), the implied public action d^{p} is also commonly known and so the productive interaction in the outcome phase is a complete-information game. Note that this game may differ depending on the messages sent earlier, because d^{p} is conditioned on the messages.

We next must apply a theory of behavior (a solution concept) for the activity in the outcome phase. Given that it is a game of complete information, the appropriate equilibrium concept might be Nash equilibrium or subgame-perfect equilibrium.[1] In the event that there are multiple equilibria, our theory of behavior would have to make, or allow for, a selection among them.

To summarize, with any particular messages sent in the communication phase, the mechanism specifies a public action d^{p} for the outcome phase. We can define a corresponding payoff outcome $w : \Theta \to \mathbb{R}^n$ as follows. For each state θ, d^{p} implies a well-defined complete-information game. We have an equilibrium concept for this game and we can select arbitrarily among equilibria. Letting $d^{\mathrm{i}}(\theta)$ denote the selected equilibrium actions of the players in state θ, and writing $d(\theta) = (d^{\mathrm{i}}(\theta), d^{\mathrm{p}})$, we can then define $w(\theta) \equiv \mu(d(\theta), \theta)$. Letting W denote the set of outcome functions that can be constructed in this way, the mechanism can be viewed as mapping messages to functions in W.

If there is incomplete information from the initial phase and/or messages are not public, then the players enter the outcome phase generally with asymmetric information. In such a setting, it may not be helpful, or even possible, to collapse the events of the outcome phase into a payoff-outcome representation. This is because there will be multiple ways of reaching the outcome phase that are indistinguishable to a given player and rational play in the outcome phase is thus intertwined with rational play earlier. In other words, one cannot fix an equilibrium in the continuation from the start of the outcome phase without also examining behavior in the communication phase.

For the rest of these notes, we focus on settings in which outcomes can be expressed in terms of state-contingent payoffs. That is, we shall make the following assumption:

Assumption 1: The mechanism-design problem has only public actions in the outcome phase and/or has complete information and public messages.

Many mechanism-design problems studied in the literature satisfy this assumption. An important exception are problems in which the external enforcer serves as a correlation device in the communication phase, as with the construction of correlated equilibria.

Under Assumption 1, we can formally describe a mechanism-design problem by the fundamentals n, Θ, and the set W of feasible (payoff) outcome functions.[2]

[1] Typically equilibrium is assumed. It would be interesting to examine a weaker concept such as rationalizabililty, but this would lead to some additional and complicated modeling issues.

[2] Remember that, in the case of public actions (no individual productive actions), we can equivalently express outcomes as the set D.

A mechanism defines what we call a *game form*. This is an extensive-form diagram describing interaction in the communication phase, with elements of W at the terminal nodes.[3] In a given state θ, at each terminal node we have a specific payoff vector $w(\theta)$ that the players' commonly know will result. This implies a well-defined game in which the players interact in the communication phase.

Implementation

With these fundamentals, our key concern is the relation between states and outcomes. For example, in the public-action setting, we may have objectives of the form "if the state is θ then the public action should be d, if the state is θ' then the public action should be d'," and so on. More generally, our objectives could be stated using payoff outcomes thus: "If the state is θ then the outcome should be w, if the state is θ' then the outcome should be w'," and so on, for appropriately selected outcome functions $w, w', \ldots \in W$.

A *(state-contingent) value function* is a function $v : \Theta \to \mathbb{R}^2$ that gives the players' expected payoff vector conditional on the state. We can use such a value function to describe specific objectives for a mechanism (we want to find a mechanism such that $v(\theta)$ is the payoff vector in state θ, for all $\theta \in \Theta$) or to describe the outcome under any given mechanism. Our objective for a given problem may be to determine whether a particular outcome function v can be achieved, or we might want to maximize welfare subject to the outcome function being achievable.

For a mechanism-design problem, the notion of a value function being "achievable" is captured by the following definition: Take as given a notion of rational behavior *RB* (Nash equilibrium, rationalizability, perfect Bayesian equilibrium, etc.). A mechanism is said to *implement* a particular value function v if in each state θ, there is behavior consistent with *RB* that leads to the payoff vector $v(\theta)$. A value function is *implementable* if there exists a mechanism that implements it.

A version of implementation — called "strong" or "full" implementation — requires not just that there exists a specification of behavior that is consistent with *RB* and achieves v, but that *every* specification of behavior that is consistent with *RB* leads exactly to this value function. One argument for strong implementation is that, given *RB*, the outcome is ensured to be the desired v. Different strands of the literature (and different groups of theorists) have adopted different implementation conditions. Folks who work on social-choice topics (e.g. whether to build a bridge) tend to call their work "implementation theory" and tend to require strong

[3] In the special case of public actions (where $d = d^{\mathrm{p}}$), it is equivalent to put a public action d at each terminal node.

implementation. Folks who study contracts (which allow the contracting parties to specify the mechanism) typically call their work "mechanism design" and tend to insist less on strong implementation. In my view, the weak notion of implementation is quite suitable for the typical contractual settings, because the players who design the mechanisms that they will play can be expected to coordinate their future behavior when contracting.

[Discuss examples of mechanism-design problems: auctions, screening and signaling in principal-agent relationships, social choice, contracting with external enforcement, market design.]

2 A Class of Problems with Complete Information, Public Actions, and Transferable Utility

The remainder of this document focuses on the following class of mechanism-design problems:

Assumption 2: The mechanism-design problem has two players, complete information, and transferable utility.

The assumption of transferable utility means we can think of utility in monetary terms and that the outcome d includes a transfer vector $t = (t_1, t_2)$ such that

$$t \in \mathbb{R}^2_0 \equiv \{(\tau, -\tau) \mid \tau \in \mathbb{R}\}.$$

This represents a monetary transfer between the players, so that t_1 is the amount player 1 receives and t_2 is the amount the player 2 receives. A negative value implies that the player gives this amount to the other.

In this section, I also limit attention to settings with public actions. That is, we'll assume that the players have no individual actions in the outcome phase, and so we can write $d = d^{\mathrm{p}}$. Here are the defining characteristics of this class of problems:

- $n = 2$;
- There is complete information, meaning that the state is commonly observed by the two players;
- The state is *unverifiable*, meaning that the third party who selects the public action does not observe θ;
- The outcome phase consists of a public action (no individual actions), which includes a monetary transfer;
- The players' payoffs are linear in money; and
- The solution concept (assumption about behavior in the communication phase) is Nash equilibrium.

The fourth item means that the public action can be written as $d = (a, t)$, where $t = (t_1, t_2)$ is the monetary/utility transfer between the players. We tend to then call a the "productive action" rather than using this term for the entire d. Let A be the set of productive actions. The assumption about transferable utility means that there is a function $u\colon A \to \mathbb{R}^2$ such that, for every $d = (a, t)$ and $\theta \in \Theta$, the payoffs can be written

$$\mu(d, \theta) = u(a, \theta) + t.$$

With Nash equilibrium as our solution concept, we do not have to look at general extensive form communication schemes but rather can focus on normal forms. A normal-form mechanism is simply a set of messages for each player, M_1 and M_2, and a mapping f from message profiles $M = M_1 \times M_2$ to outcomes. This is also known as a *message game*. The implementation problem for this class of environments is often referred to as "Nash implementation."

> Mechanism (M, f) is said to **(Nash) implement** value function v if $f\colon M \to W$ and, for each state θ, there is a Nash equilibrium of the message game that leads to the payoff vector $v(\theta)$. Value function v is said to be *implementable* if there is a mechanism that implements it.

Note that this is a weak implementation concept.

One of the benefits of mechanism-design theory comes from an important result known as the *revelation principle* which, in many settings, justifies limiting attention to the following subset of mechanisms:

> A **direct-revelation mechanism** is a mechanism in which each player is supposed to report the state. That is, each player's message space is the set of states ($M_i = \Theta$ for $i = 1, 2$).

Result 1: [Revelation Principle] If a value function v is implementable than it can be implemented by a direct-revelation mechanism in which the players report truthfully in equilibrium.

From this point we can concentrate on direct-revelation mechanisms, each which is described by a mapping f from $\Theta \times \Theta$ to W.

Exercise 1: Prove Result 1.

The transferable-utility assumption puts some structure on the implementable set:

Result 2: The set W is closed under constant transfers. That is, for any $w \in W$ and any $t \in \mathbb{R}^2_0$, $w + t$ is also in W.

Exercise 2: Prove Result 2.

3 An Example

Here is a numerical example of a public-choice problem (whether to install a WiFi system) with transfers:

- Two citizen players, 1 and 2.
- Two states: $\Theta = \{\mathrm{L}, \mathrm{H}\}$.
- Public action: a transfer τ from player 2 to player 1 and selection of whether to install ($a = 1$) or not ($a = 0$). Thus, a public action is written $d = (a, t)$ where $t = (t_1, t_2) = (\tau, -\tau)$.

- Payoffs: $\mu(d, \theta) \equiv u(a, \theta) + t$. Suppose that $u(1, \mathrm{H}) = (8, 0)$, $u(1, \mathrm{L}) = (0, 4)$, $u(0, \mathrm{H}) = (0, 0)$, and $u(0, \mathrm{L}) = (0, 0)$.

The story here is that a WiFi system could be installed for players 1 and 2 to use. Its benefit depends on the state, which indicates some other technical infrastructure for equipment that can utilize the WiFi system. Player 2 owns equipment that uses an older technology, whereas player 1 owns equipment that uses a newer technology. In state L, the technical infrastructure is suited to the older technology, where only player 2 would benefit from the WiFi system. State H represents an upgrade of the technical infrastructure, under which player 2's equipment is obsolete but player 1 would benefit a great deal from the WiFi system.

Thus, in this example, the players get zero if the WiFi system is not installed, regardless of the state. On the other hand, if installation were to occur then player 1 would get a benefit of 8 in the H state and player 2 would get a benefit of 4 in the L state. Note that public action $d = (a, t)$ implies the payoff outcome w, where $w(\mathrm{H}) = u(a, \mathrm{H}) + t$ and $w(\mathrm{L}) = u(a, \mathrm{L}) + t$. For instance, for $d = (1, \vec{0})$ we have $w(\mathrm{H}) = (8, 0)$ and $w(\mathrm{L}) = (0, 4)$.

Let us analyze the example with economic objectives of —*ex post efficiency* and distribution in mind. Ex post efficiency means that the outcome in each state is efficient conditioned on this state being reached. This is not always desirable from the *ex ante* perspective (you will see this later), but let us take it is the goal for now. Beyond ex post efficiency, we should also find the widest possible range of implementable value functions. I defer a fuller discussion of the economic objectives until Section 5, where the scope of analysis will be expanded to include ex ante considerations.

First consider what happens if the mechanism (game form) is trivial in that the public action is independent of messages (equivalently, the players have no opportunity to send messages). Note that a constant public action of $d = (1, \vec{0})$ implements the outcome associated with d. That is,

- A constant $d = (1, t)$ implements v given by $v(\mathrm{H}) = (8 + t_1, t_2)$ and $v(\mathrm{L}) = (t_1, t_2)$, and
- A constant $d = (0, t)$ implements v given by $v(\mathrm{H}) = v(\mathrm{L}) = (t_1, t_2)$.

These may not achieve economic objectives, however. For example, the second outcome is inefficient in both states. Further, the first, although efficient in both states, may not achieve the preferred distribution of value. For example, suppose we are interested in some notion of equity and want the players to share equally in the value created by the WiFi system. Note that the joint value created by installing the WiFi system is 8 in state H and 4 in state L. If we want this value to be divided equally between the players, then we want to implement v given by $v(\mathrm{H}) = (4, 4)$ and $v(\mathrm{L}) = (2, 2)$. So, we should look further to characterize the entire set of implementable value functions.

Let us next look at mechanisms that involve a message from just one of the players. That is, let us give one of the players an *option*. (Remember that the third party does not observe the state and therefore must rely on communication from the players.) To implement the desired v, we would have to give a player the choice between public action $d^{\mathrm{H}} = (1, (-4, 4))$ and public action $d^{\mathrm{L}} = (1, (2, -2))$. But clearly player 1 would always strictly favor d^{L} and player 2 would always strictly favor d^{H}, so neither player has an incentive to pick the socially preferred action in each state.

Next let us perform a more complete analysis by using the revelation principle. That is, examine a mechanism that requires the players to simultaneously report the state, and let us construct the mechanism so that the players report truthfully in each state. The mechanism (game form) will look like this:

1 \ 2	H	L
H	install ($a = 1$), transfer $\tau^H = -4$	?
L	?	install ($a = 1$), transfer $\tau^L = 2$

What should go in the off-diagonal elements of this matrix? We want the players to be punished in these cells, because reaching one of these means that one of the players deviated. The best way to punish is to specify $a = 0$ (no installation) because it leads to the lowest value for the players. Let us try $a = 0$ and $t = \vec{0}$, to get the following message game form:

1 \ 2	H	L
H	install ($a = 1$), transfer $\tau^H = -4$	not ($a = 0$), transfer $t = 0$
L	not ($a = 0$), transfer $t = 0$	install ($a = 1$), transfer $\tau^L = 2$

Note that this game form implies the following game in state H:

1 \ 2	H	L
H	4, 4	0, 0
L	0, 0	10, – 2

It is an equilibrium for the players to select (H, H) in this state (in fact, uniquely so).

Note also that the game form implies the following game in state L:

1 \ 2	H	L
H	– 4, 8	0, 0
L	0, 0	2, 2

It is an equilibrium for the players to select (L, L) in this state (in fact, uniquely so).

Thus, the prescribed game form implements (uniquely) the desired value function v given by $v(\mathrm{H}) = (4, 4)$ and $v(\mathrm{L}) = (2, 2)$.

4 More General Tools

This section reviews some basic theorems that simplify the determination of implementable value functions and identify properties of the set.

Result 3: [Maskin 1999, Watson 2007] Value function v is implementable if and only if (i) for every $\theta \in \Theta$, there is an outcome $w \in W$ such that $w(\theta) = v(\theta)$; and (ii) for every pair of states $\theta, \theta' \in \Theta$, there is an outcome $\hat{w} \in W$ such that $v_1(\theta) + v_2(\theta') \geq \hat{w}_1(\theta) + \hat{w}_2(\theta')$. Also, the set V of implementable value functions is closed under constant transfers.

1 \ 2	θ	θ'
θ	w	$\underline{\hat{w}}$
θ'	$\hat{w}$	w'

Thus, characterizing the implementable set amounts to examining the lowest (most severe) punishment values that can be imposed in the off-diagonal cells of the message game. For the cell corresponding to (θ', θ), this is

$$P(\theta, \theta') \equiv \min_{w \in W} w_1(\theta) + w_2(\theta').$$

The transfer is neutral in the calculation. In the case of public actions, where the external enforcer takes action $a \in A$, we can express P in terms of u:

$$P(\theta, \theta') \equiv \min_{a \in A} u_1(a, \theta) + u_2(a, \theta').$$

Using P, condition (ii) in Result 3 is that for every pair of states $\theta, \theta' \in \Theta$ we have

$$v_1(\theta) + v_2(\theta') \geq P(\theta, \theta').$$

Note that this imposes two inequalities for each pair of states. For example, suppose H and L are states. Then condition (ii) requires $v_1(\mathrm{H}) + v_2(\mathrm{L}) \geq P(\mathrm{H}, \mathrm{L})$ as well as $v_1(\mathrm{L}) + v_2(\mathrm{H}) \geq P(\mathrm{L}, \mathrm{H})$. That is, the condition is sensitive to the order of the states.

Result 4: [Watson 2006, 2007] For any two states $\theta, \theta' \in \Theta$ and for any given $\underline{w} \in W$, there is an implementable value function $v\colon \Theta \to \mathbb{R}^2$ such that

(i) $v_1(\theta) + v_2(\theta') = \underline{w}_1(\theta) + \underline{w}_2(\theta')$, and

(ii) $v_1(\theta'') + v_2(\theta'') = \max_{w \in W} w_1(\theta'') + w_2(\theta'')$ for every $\theta'' \in \Theta$.

Exercise 3: Calculate and represent the set of implementable value functions for the example of Section 3.

This section concludes with a note about a special subclass of settings in which implementation has a strong form. Let $U(a, \theta)$ be defined as the joint value of productive action a in state θ, which the transfer obviously does not affect. That is,

$$U(a, \theta) \equiv u_1(a, \theta) + u_2(a, \theta).$$

The following restriction will at first appear to be of very limited interest and applicability, but it will turn out to be relevant in later sections.

> A mechanism-design problem is said to exhibit **constant state-contingent joint values** if for every $\theta \in \Theta$ and every $a, a' \in A$, it is the case that $U(a, \theta) = U(a', \theta)$.

Result 5: Consider a mechanism-design problem that exhibits constant state-contingent joint values. If a mechanism f implements value function v, then f strongly implements v in the sense that, in every state θ, every Nash equilibrium of the message game yields payoff vector $v(\theta)$.[4]

Exercise 4: Prove Result 5. Hint: Utilize a standard result on exchangeable and equivalent Nash equilibria in strictly competitive games. Note that the mixed extension of a constant-sum game is also constant sum and hence strictly competitive.

This result does not hold without constant state-contingent joint values. The following exercise demonstrates that strong implementation is generally incompatible with direct-revelation mechanisms (so the revelation principle does not hold for strong implementation).

[4]This result actually holds under the slightly more general condition of the "linear payoff property," which just allows for different (fixed) marginal values of money for the two players.

Exercise 5: Consider a public-action, complete-information, mechanism-design setting with $n = 2$ (two players), state space $\Theta \equiv \{\mathrm{L}, \mathrm{H}\}$, and public action space $P \equiv \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}$. Suppose that preferences are given by utility function $\mu : P \times \Theta \to \mathbb{R}^n$ that takes values according to the following table.

$$\mu(\cdot,\cdot) = \begin{array}{c|c|c|} & \mathrm{L} & \mathrm{H} \\ \hline \mathrm{a} & 1,1 & 1,1 \\ \hline \mathrm{b} & 2,2 & 2,2 \\ \hline \mathrm{c} & -1,4 & -1,4 \\ \hline \mathrm{d} & 4,-1 & 4,-1 \\ \hline \mathrm{e} & 0,0 & 3,4 \\ \hline \mathrm{f} & 4,3 & 0,0 \\ \hline \end{array}$$

Note that this is a setting with *no transfers*. Suppose that you want to implement value function v such that $v(\mathrm{L}) = (1,1)$ and $v(\mathrm{H}) = (2,2)$.

(a) Find a direct-revelation mechanism that implements v (weakly) with truthful reporting.

(b) Prove that v cannot be strongly implemented by a direct-revelation mechanism. Ignore mixed strategy equilibria.

(c) Find a mechanism that strongly implements v. Again, ignore mixed strategy equilibria.

Exercise 6: Consider all public-action, mechanism-design settings with two players and just two states. Can you find necessary and sufficient conditions for a value function to be strongly implementable by a direct revelation mechanism and with truthful reporting in equilibrium? Consider separately three different solution concepts: pure strategy Nash equilibrium, dominance, and rationalizability. How would the conditions be altered in the case of transferable utility?

5 Contracting and Renegotiation

[From here, these notes are in more of a summary form, rather than with complete text.]

A *contract* is an agreement about behavior that parties intend to enforce. There are two types of enforcement:

- **External enforcement** – An external enforcer (whose incentives we do not study) takes actions as a function of verifiable information, as directed by the contract.
- **Self-enforcement** – Each contracting party has an incentive to abide by the agreement when he/she anticipates that the others will as well.

Contractual Settings with Investment, Trade, and Renegotiation

[Describe the general timing with individual trade actions. Define forcing contracts. Show that forcing contracts effectively make the trade outcome a public action. (Pull this material from Buzard-Watson.) Say that part of the literature focuses on forcing contracts for simplicity (or looks at the case in which productive actions are public).]

Interpretation of the Example as a Contracting Problem

The framework described above can be interpreted as a contractual setting, whereby the players form a contract before the state is realized. The contract's externally enforced part is the mechanism (the strategy of the external enforcer and commitment to the communication scheme). Self-enforcement means that the players' strategies (in the communication phase) form an equilibrium.

Let us reinterpret the example of Section 3 as a contracting problem with specific investment. Suppose that the state is determined by strategic choices that a player makes. In particular, assume that player 2 must choose an investment, high (H) or low (L), and this determines the state. Suppose that the payoffs are the same as before, except that player 2 pays an additional cost of c if he/she makes the high investment ($\theta = \mathrm{H}$). Assume that $c \in (0, 4)$, which means high investment is efficient. The timing of interaction is:

- **Period 0**
  - The players write a contract $\mathcal{C}$, specifying a mechanism (message game) to be enforced in period 1.
  - Player 2 selects an investment $\theta \in \{\mathrm{L}, \mathrm{H}\}$ and pays c in the case of H.

- **Period 1**
  - The players send messages (reports of the state).
  - The external enforcer compels the public action $d = (a, t)$ as a function of the message profile, as directed by the contract.

In this kind of contracting problem, we have *specific investment* because player 2's investment is valuable only in the relationship with player 1. (There isn't another party with whom player 2 could interact and extract some of the value of her investment.) The investment is *unverifiable*, so there is no way to contract directly on θ and force player 2 to pick the efficient investment. Furthermore, the investment affects the benefits of the productive action a, so there may be a way to structure how the public action is selected that effectively rewards player 2 for choosing the efficient investment level.

Note that the value functions we analyzed in previous sections are, in this example, continuation values from the start of Period 1. Assume that these values do not incorporate the cost c; that is, the value functions are gross of Period-0 investment costs.

Now the key issue for implementation is whether it is possible to implement a value function v satisfying $v_2(\mathrm{H}) - c \geq v_2(\mathrm{L})$, for otherwise player 2 will not have the incentive to choose high investment.

Recall our calculation of the set of implementable value functions. Note that desired value functions are implementable and so we predict a good outcome for this contractual relationship.

Note that we have not modeled the initial contracting process, but we have assumed that the players maximize their joint value in the contracting phase (as our standard bargaining solutions would predict).

Renegotiation

Typically, contracting parties have opportunities to renegotiate their contracts. In other words, the players replace their original contract with a modified one. For the framework we have been studying, this means that the players can agree to change the mechanism at some time.

We can incorporate renegotiation using our standard model of bargaining. In settings with complete information, as we have here, it is most convenient to use a cooperative solution concept. This bargaining solution represents a noncooperative model of bargaining that would be inconvenient to specify in the middle of a model that is already becoming complicated. We define the disagreement point to be an equilibrium continuation value under the outstanding (original) contract. The standard bargaining solution predicts that the players will select a new contract in order to divide the surplus of renegotiation in proportion to their bargaining weights $\pi = (\pi_1, \pi_2)$, where $\pi_1, \pi_2 \geq 0$ and $\pi_1 + \pi_2 = 1\ldots$

[Discuss the importance of fixed bargaining weights.]

Times at which renegotiation can occur:

- Before the state is realized (ex ante),
- Just after the state is realized and before players send messages (interim), and
- After the communication phase but before the physical outcome/public action (ex post).

There are variants of interim and ex post that depend on a more detailed account of trade (more on this later).

In the case of interim renegotiation, the disagreement point in a given state is whatever continuation value would have been implemented in this state by the originally specified mechanism. In the case of ex post renegotiation, the disagreement point depends on both the state and on the actual messages sent by the players in the communication phase (on or off the equilibrium path). It is the payoff vector of the outcome specified by the original contract given these messages. In both interim and ex post renegotiation, by selecting a new mechanism the players are essentially selecting the outcome directly.

Let us concentrate on ex post renegotiation in the contractual relationship that is based on the example of Section 3. The timing is:

- **Period 0**
  - The players write a contract, specifying a mechanism (message game) to be enforced in period 1.
  - Player 2 selects $\theta \in \{\mathrm{L}, \mathrm{H}\}$.

- **Period 1**
  - The players send messages (reports of the state).
  - The parties can renegotiate the contract.
  - The external enforcer compels the public action $d = (a, t)$ as a function of the message profile, as directed by the current contract.

Note that we are continuing to assume here that the productive action at the end is taken by an external enforcer—that is, it is a public action.

Here is a trick (from Maskin and Moore 1999) for incorporating the renegotiation element: Redefine the outcome functions to represent the ex post renegotiation activity. We shall use the standard bargaining solution, so that the players select an ex post efficient public action in each state and they divide the surplus of renegotiation according to their bargaining weights. Designate by z the new outcome function and let Z be the set of functions derived in this way. In the example, we have

- Public action $d = (1, t)$ implies the payoff outcome z given by $z(\mathrm{H}) = (8 + t_1, t_2)$ and $w(\mathrm{L}) = (t_1, 4 + t_2)$, and
- Public action $d = (0, t)$ implies the payoff outcome z given by $z(\mathrm{H}) = (8\pi_1 + t_1, 8\pi_2 + t_2)$ and $w(\mathrm{L}) = (4\pi_1 + t_1, 4\pi_2 + t_2)$.

To define the set of renegotiation-derived outcome functions in general, let $\gamma(\theta)$ denote the maximum joint value in state θ:

$$\gamma(\theta) \equiv \max_{a \in A} U(a, \theta) = \max_{w \in W} w_1(\theta) + w_2(\theta).$$

Further, let $r(w, \theta) \equiv \gamma(\theta) - w_1(\theta) - w_2(\theta)$. Then the set Z is defined as follows:

$$Z \equiv \{z \mid \text{there exists } w \in W \text{ such that } z(\theta) = w(\theta) + \pi r(w, \theta) \text{ for all } \theta \in \Theta\}.$$

We then proceed to analyze the contracting problem using the techniques reviewed in the previous sections — that is, calculating the set of implementable value functions and examining whether this set contains functions that support efficient outcomes from the beginning of the contractual relationship. Ex post renegotiation is incorporated by using Z to represent outcomes, rather than by using W.

Note that the sets Z and W are not necessarily related by inclusion. In particular, it is not necessarily the case that $Z \subset W$. Therefore, it may not be obvious whether renegotiation imposes an additional constraint on the contracting problem or expands the implementable set. In fact, it is the former, as the following result implies.

Result 6: [Renegotiation-Proofness Principle] Consider a mechanism-design setting with ex post renegotiation. If a value function v is implementable then it can be implemented using a direct-revelation mechanism in which players report truthfully and there is no renegotiation on the equilibrium path. That is, each on-diagonal element of the game form specifies an outcome that is efficient in the relevant state. The same result holds for the case of only interim renegotiation.

Let V^{EPF} be the set of implementable value functions when using Z as the set of outcomes, and let V^{N} be the set of implementable value functions when using W as the set of outcomes. That is, V^{EPF} incorporates ex post renegotiation using Maskin and Moore's (1999) construction, whereas V^{N} assumes no renegotiation. Result 6 clearly implies that $V^{\mathrm{EPF}} \subset V^{\mathrm{N}}$. That is, opportunities to renegotiate impose a constraint that narrows the implementable set.[5] This is a manifestation of the *hold-up problem*.

[5]"F" is in the superscript here to follow the notational convention in Watson (2007), which will

Exercise 7: Prove Result 6.

A nice by-product of ex post renegotiation is that its presence in the model implies a strong form of implementation.

Result 7: [Unique Implementation, Maskin and Moore 1999] With ex post renegotiation, the mechanism-design problem exhibits constant state-contingent joint values, and so implementation is unique in value function terms.

To illustrate the mechanism-design tools, we can perform a thorough analysis of the example in Section 3. Here are the steps:

- Redefine the outcome function by incorporating the implications of renegotiation,
- Calculate the set of implementable value functions, and
- Compare this set with the implementable set in the case without renegotiation.

[Go through the calculations.]

[Prove informally that in the public-action setting, the set of implementable value functions would not change if we allow private messages. We have to model the renegotiation non-cooperatively. Anticipating the other player will report truthfully, a player could choose any message and force the other player to know that this message was chosen by making an offer that gives the other player a huge amount if any other message were chosen. (Check notes and students work on this exercise.)]

---

be handy later. It is not difficult to see that the set of implementable value functions in the case of interim renegotiation, V^{I}, is intermediate between V^{EPF} and V^{N}.

6 Public-Action Models in the Hold-Up Literature

An active area of the contract-theory literature is the examination of hold-up problems in relationships with specific investment, of which the illustrative example in these notes is one case. The following examples summarize the results of the recent papers that specify public-action models. The next section reviews the more recent literature that criticizes public-action modeling and clarifies the results.

Each of the following examples features a procurement relationship, where a buyer (player 1) interacts with a seller (player 2). Assume that ex post renegotiation is possible and the timing is exactly as described in Section 5. I will keep it simple by assuming that, in all cases, the seller makes the investment at the end of Period 0. Furthermore, for now let us assume that this investment precisely determines the state. That is, there is no exogenous uncertainty in the model. The examples below thus differ solely on the basis of the benefits and costs of the investment, and the variety of public actions.

Own Investment

In the own-investment case — studied by Chung (1991), Rogerson (1992), Aghion, Dewatripont, and Rey (1994), Nöldeke and Schmidt (1995), Edlin and Reichelstein (1996) and others — the seller's investment affects her cost of trade later. Contributors to this part of the literature have shown that own-investment problems can generally be solved so that efficient investment and trade are achieved (an exception is under a different heading below). Furthermore, fairly simple contracts often suffice.

Here is a very simple version. Let the seller's (player 2's) investment determine the state $\theta \in \Theta \equiv \mathbb{R}_+$. That is, player 2 selects the state. Assume that the immediate cost of investment is θ. The public action is written (a, t) as usual, with t denoting the transfer vector. Suppose that $A \equiv \{0, 1\}$, where $a = 1$ denotes "trade" and $a = 0$ denotes "no trade." If the trade does not occur then both players get zero, gross of the investment cost. If trade occurs then player 1 obtains a benefit of 1 and player 2 pays the cost $c(\theta)$. That is, $u(0, \theta) = (0, 0)$ and $u(1, \theta) = (1, -c(\theta))$ for all θ. The function c is continuous, decreasing, and positive.

Exercise 8: Analyze this example of own-investment through the following steps. If you need to add assumptions, be careful to precisely describe them.

(a) Calculate the efficient level of investment and trade. Under what conditions is it efficient to have strictly positive investment and trade?

(b) Show that a simple contract that requires trade at a fixed price (with no messages) implements a value function that leads to efficient investment and trade, without renegotiation occurring on the equilibrium path. Start with the case in which $c(0) < 1$, and then consider the additional complications that arise if $c(0) > 1$.

(c) Consider a variation of the model in which the player 1's benefit of trade is randomly determined at the time that player 2 selects θ. Let player 1's trade benefit be denoted b and suppose that it is uniformly distributed on $[0, 1]$. Assume that b is commonly observed by the players but is unverifiable to the external enforcer. What is the *state* in this setting?

Furthermore, expand the set of trade actions to include fractional trades, so that $A \equiv [0, 1]$. Let player 1's benefit and player 2's cost of trade be given by ab and $ac(\theta)$, respectively. Determine the efficient level of investment and trade. Find a simple (no message) contract that leads to efficient investment and trade for the case in which $c(0) < 1$. Feel free to make additional assumptions, such as on differentiability.

Cross Investment

In the cross-investment case — introduced to the literature by Che and Hausch (1999), who call this "cooperative investment" — the seller's investment affects the buyer's benefit of trade and may also affect her cost of trade later. Che and Hausch showed that ex post renegotiation (the threat of hold up) greatly constrains the contracting problem, so that inefficiency is inevitable. Furthermore, the optimal contract is often a simple "null" contract, whereby the players defer any commitment until after the seller makes the investment choice.

Here is a very simple version of Che and Hausch's (1999) model. As above, let the seller's (player 2's) investment determine the state $\theta \geq 0$ and let this measure the cost of investment. Suppose that $A \equiv \{0, 1\}$ as before, where $a = 1$ denotes "trade" and $a = 0$ denotes "no trade." If the trade does not occur then both players get zero, gross of the investment cost. If trade occurs then player 1 obtains a benefit of $b(\theta)$ and player 2 pays the cost $c(\theta)$. Thus, $u(0, \theta) = (0, 0)$ and $u(1, \theta) = (b(\theta), -c(\theta))$ for all θ. Assume that c and b are twice continuously differentiable and satisfy $b(\theta) - c(\theta) \geq 0$ for all θ, $b' \geq 0$, $c' \leq 0$, $b'' < 0$, $c'' > 0$, $b'(0) - c'(0) > 1$, and $b'(\theta) - c'(\theta) \to 0$ as $\theta \to \infty$.

Exercise 9: Analyze this example of cross investment by taking the following steps:

(a) Provide an expression that identifies the efficient level of investment, θ^*, and trade.

(b) Consider the null contract that specifies $a = 0$ and $t = 0$ with no messages. Examine and describe how this contract would be renegotiated in each state, and describe the resulting value function $\hat{v}$ that is implemented. Calculate player 2's optimal investment choice, $\hat{\theta}$. How does $\hat{\theta}$ compare with the efficient investment θ^*?

(c) Consider any two states θ^1 and θ^2 such that $\theta^1 > \theta^2$. Calculate $P(\theta^1, \theta^2)$ and $P(\theta^2, \theta^1)$. Describe the public actions that achieve the minima. Remember that there is ex post renegotiation, so you have to calculate the z functions.

(d) From your answer to part (c), give a uniform upper bound on $v_2(\theta^1) - v_2(\theta^2)$ for all implementable value functions.

(e) Find conditions under which this upper bound is no greater than $\hat{v}_2(\theta^1) - \hat{v}_2(\theta^2)$. Explain why you can then conclude that the null contract is the best contract under these conditions, although it does not induce the efficient investment.

Complexity/Ambivalence

In the case of a complex assortment of possible goods to trade — a setting put forth by Segal (1999) and Hart and Moore (1999), and elaborated on by Reiche (2006) — the seller's investment affects her own cost of trade but there are many possible types of goods that can be traded and they have wildly different costs and benefits. These authors showed that ex post renegotiation causes a problem because the players can hold up each other by strategically declaring that they should trade the wrong type of good. Inefficient outcomes are unavoidable.

Here is a very simple version of Segal's (1999) model. As before, the state represents the seller's (player 2's) investment, which is either "high" (state H) or "low" (state L). The high investment entails a cost c, which is immediately paid by the seller. Low investment is costless. The public action is a transfer and productive action a, which is either a^h or a^l. Function u is defined by:

$u(a^h, \mathrm{H}) = (10, 0)$, $u(a^l, \mathrm{H}) = (22, -22)$, $u(a^h, \mathrm{L}) = (0, 0)$, and $u(a^l, \mathrm{L}) = (10, -8)$. As with the other examples, assume that ex post renegotiation is possible.

Exercise 10: For this example, calculate the set of implementable value functions (continuation values from just after the state is realized) and determine whether the efficient investment level can be supported. Comment on the sense in which this is a case of self investment.

When Investment Produces Direct Benefits

If player 2's investment gives a direct benefit to player 1 (a benefit that is independent of the productive action a that occurs later), then it becomes more difficult to give player 2 the incentive to invest. This is a direct result of the investment being unverifiable. Ellman (2006) looks at this issue, at least indirectly, by examining variations in a form of specificity.

7 Modeling the Technology of Trade

[This section has not yet been written but is the most important part. . . Read and be prepared to discuss Watson (2007), Buzard and Watson (2009), Evans (2008), and Watson and Wignall (2009).]

8 Hold Up in Dynamic Settings

Durable Trading Opportunities

[Watson-Wignall, Evans]

Durable Investment Opportunities

[Che-Sakovics]

9 Solutions to Exercises

Solution to Exercise 5:

(a) Specify b for message profile (H,H) and a otherwise. (b) We need to use e or f in at least one of the off-diagonals, because otherwise the message games played in the two states are identical (and thus have multiple equilibria with different payoffs). But you can see that specifying either e or f will disrupt incentives and not implement v. (c) The following mechanism (writhen in terms of the public action specified for each message profile) will work:

$1\backslash 2$	L	H	K
L	a	a	e
H	a	b	d
K	c	f	a

Solution to Exercise 8:

(a) Given θ, it is efficient to trade if and only if $1 \geq c(\theta)$. Because c is decreasing and continuous, there is a number $\underline{\theta}$ such that $c(\theta) > 1$ for $\theta < \underline{\theta}$ and $c(\theta) \leq 1$ for $\theta \geq \underline{\theta}$. That is, trade is ex post efficient if $\theta \geq \underline{\theta}$ and not trading is efficient if $\theta < \underline{\theta}$. Strictly positive investment and trade are optimal if there exists $\theta > 0$ such that $1 - c(\theta) - \theta > 0$, and then the efficient level of investment solves $\max_\theta 1 - \theta - c(\theta)$. Let θ^* denote the solution. It is clear that $\theta^* > \underline{\theta}$, because otherwise we would have $1 - \theta - c(\theta) < 0$ since $c(\theta) > 1$.

(b) Assume that strictly positive investment is optimal, so the problem is interesting. Let the contract specify $a = 1$ with transfer t. There are no messages. Note that renegotiation will occur if $\theta < \underline{\theta}$ and that the renegotiation surplus is then $c(\theta) - 1$. Thus, in state θ, the implemented payoff vector is

$$v(\theta) = (1, -c(\theta)) + t + \pi \max\{c(\theta) - 1, 0\}.$$

Player 2 maximizes

$$-\theta + t_2 - c(\theta) + \pi_2 \max\{c(\theta) - 1, 0\}.$$

Note that for $\theta < \underline{\theta}$,

$$-\theta + t_2 - c(\theta) + \pi_2 \max\{c(\theta) - 1, 0\} = -\theta + t_2 - \pi_1 c(\theta) - \pi_2 < t_2 - 1,$$

because for such a value of θ we have $c(\theta) > 1$ (and we use the facts that $\pi_1 + \pi_2 = 1$ and $\theta \geq 0$). We know that $-\theta^* - c(\theta^*) > -1$, so

$$-\theta^* + t_2 - c(\theta^*) + \pi_2 \max\{c(\theta^*) - 1, 0\} = -\theta^* + t_2 - c(\theta^*) > t_2 - 1.$$

Therefore, player 2 will not select $\theta < \underline{\theta}$, so her optimal investment choice is that which maximizes $-\theta + t_2 - c(\theta)$ over $\theta \geq \underline{\theta}$, which is θ^*.

(c) Consider the case in which $c(0) < 1$ so $\underline{\theta} = 0$. Note that the state is now given by (θ, b). Clearly it is ex post efficient to have $a = 1$ if $b - c(\theta) \geq 0$ and $a = 0$ if $b - c(\theta) < 0$. Efficiency never requires $a \in (0, 1)$. Thus, for a given θ, trade is efficient if and only if $b \geq c(\theta)$. The efficient investment level θ^* is that which solves

$$\max_{\theta \geq 0} -\theta + \int_{c(\theta)}^{1} [b - c(\theta)]db.$$

Assume that $\theta^* > 0$. So we can use calculus, also assume that c is twice continuously differentiable, with $c' < 0$ and $c'' > 0$. Then the solution to the maximization problem above is characterized by the following first-order condition:

$$c'(\theta^*)[1 - c(\theta^*)] = -1 \tag{2}$$

Note that, from our assumptions, $c'(\theta)[1 - c(\theta)]$ is strictly increasing in θ and equals -1 at $\theta^* > 0$.

Consider a simple, no-message contract that specifies some trade action $\hat{a} \in [0, 1]$ and a transfer t. Note that if $\hat{a} \in (0, 1)$ then renegotiation will occur in almost every state. In general, if $b - c(\theta) < 0$ then the parties will renegotiate to choose $a = 0$. Likewise, they will select $a = 1$ if $b - c(\theta) \geq 0$. Note that the renegotiation surplus in the first case is $\hat{a}(c(\theta) - b)$ and in the second case it is $(1 - \hat{a})(b - c(\theta))$. This implies that the implemented value function is given by:

$$v(\theta, b) = (\hat{a}b, -\hat{a}c(\theta)) + t + \pi \max\{\hat{a}(c(\theta) - b), (1 - \hat{a})(b - c(\theta))\}.$$

Taking the expectation over b and subtracting player 2's investment cost, we see that player 2's expected value is

$$-\theta - \hat{a}c(\theta) + t_2 + \pi_2 \int_0^{c(\theta)} \hat{a}(c(\theta) - b)db + \pi_2 \int_{c(\theta)}^{1} (1 - \hat{a})(b - c(\theta))db.$$

Player 2 will choose the $\hat{\theta}$ that maximizes this, which yields the following first-order condition:

$$-1 - \hat{a}c'(\hat{\theta}) + \pi_2 \hat{a}c'(\hat{\theta})c(\hat{\theta}) - \pi_2(1 - \hat{a})c'(\hat{\theta})[1 - c(\hat{\theta})] = 0,$$

which simplifies to

$$c'(\hat{\theta})[\hat{a}(1 - \pi_2 c(\hat{\theta})) + (1 - \hat{a})\pi_2(1 - c(\hat{\theta}))] = -1.$$

Observe that

$$\hat{a}(1 - \pi_2 c(\theta) + (1 - \hat{a})\pi_2(1 - c(\theta))$$

is increasing in θ, so $\hat{\theta}$ is well defined. A sufficient condition for concavity of the problem (the second-order condition) is that $[c'(\theta)]^2 \leq c''(\theta)(1 - c(\theta))$ for all θ.[6]

[6] To show this, take the second derivative observe that it is highest at $\hat{a} = 0$, so setting $\hat{a} = 0$ yields an upper bound.

By our continuity assumptions, the solution to player 2's investment problem is continuous in $\hat{a}$. Note that at $\hat{a} = 1$ the solution is given by

$$c'(\hat{\theta})[1 - \pi_2 c(\hat{\theta})] = -1.$$

Furthermore, at $\hat{a} = 0$ the solution is given by

$$c'(\hat{\theta})\pi_2[1 - c(\hat{\theta})] = -1.$$

Comparing these with Equation 2 that characterizes θ^*, we conclude that $\hat{\theta} > \theta^*$ in the first case and $\hat{\theta} < \theta^*$ in the second case. By continuity, there is a number $\hat{a} \in (0, 1)$ that makes $\hat{\theta} = \theta^*$.

Solution to Exercise 9:

(a) Trade is always ex post efficient. The efficient investment level θ^* solves $\max_\theta b(\theta) - c(\theta) - \theta$, which has the first-order condition $b'(\theta^*) - c'(\theta^*) = 1$.

(b) Since trade is always efficient, the null contract would be renegotiated to have $a = 1$. The surplus of renegotiation in state θ is $b(\theta) - c(\theta)$, so the implemented value function is given by:

$$\hat{v}(\theta) = \pi[b(\theta) - c(\theta)].$$

Player 2 chooses θ to solve $\max \pi_2[b(\theta) - c(\theta)] - \theta$, yielding the first-order condition $b'(\hat{\theta}) - c'(\hat{\theta}) = 1/\pi_2$. Because $b - c$ is strictly concave, $b' - c'$ is strictly decreasing and we conclude that $\hat{\theta} < \theta^*$ assuming $\pi_2 < 1$. If $\pi_2 = 1$ then efficient investment and trade is achieved.

(c) We take as given $\theta^1 > \theta^2$. Consider $P(\theta^1, \theta^2)$. Trade action $a = 1$ implies a punishment value of $b(\theta^1) - c(\theta^2)$, because there is no renegotiation surplus in either state. Trade action $a = 0$ implies a punishment value of

$$\pi_1[b(\theta^1) - c(\theta^1)] + \pi_2[b(\theta^2) - c(\theta^2)],$$

because the players get zero without trade but would renegotiate in both states. A few algebraic steps reveal that using $a = 0$ yields a strictly lower punishment value if and only if $\pi_1[c(\theta^2) - c(\theta^1)] < \pi_2[b(\theta^1) - b(\theta^2)]$.

Next consider $P(\theta^2, \theta^1)$. Trade action $a = 1$ implies a punishment value of $b(\theta^2) - c(\theta^1)$, because there is no renegotiation surplus in either state. Trade action $a = 0$ implies a punishment value of

$$\pi_1[b(\theta^2) - c(\theta^2)] + \pi_2[b(\theta^1) - c(\theta^1)],$$

because the players get zero without trade but would renegotiate in both states. A few algebraic steps reveal that using $a = 0$ yields a strictly lower punishment value if and only if $\pi_1[c(\theta^2) - c(\theta^1)] > \pi_2[b(\theta^1) - b(\theta^2)]$.

To summarize, if $\pi_1[c(\theta^2) - c(\theta^1)] < \pi_2[b(\theta^1) - b(\theta^2)]$ then we have

$$P(\theta^1, \theta^2) = \pi_1[b(\theta^1) - c(\theta^1)] + \pi_2[b(\theta^2) - c(\theta^2)] \text{ and } P(\theta^2, \theta^1) = b(\theta^2) - c(\theta^1).$$

On the other hand, if $\pi_1[c(\theta^2) - c(\theta^1)] \geq \pi_2[b(\theta^1) - b(\theta^2)]$ then we have

$$P(\theta^1, \theta^2) = b(\theta^1) - c(\theta^2) \text{ and } P(\theta^2, \theta^1) = \pi_1[b(\theta^2) - c(\theta^2)] + \pi_2[b(\theta^1) - c(\theta^1)].$$

(d) Each implementable value function satisfies $v_1(\theta^1) + v_2(\theta^2) \geq P(\theta^1, \theta^2)$. Using the fact that $v_1(\theta^1) + v_2(\theta^1) = b(\theta^1) - c(\theta^1)$, this implies

$$v_2(\theta^1) - v_2(\theta^2) \leq b(\theta^1) - c(\theta^1) - P(\theta^1, \theta^2),$$

where $P(\theta^1, \theta^2)$ is determined from part (c).

(e) By the definition of $\hat{v}$, we see that

$$\hat{v}_2(\theta^1) - \hat{v}_2(\theta^2) = \pi_1[b(\theta^1) - c(\theta^1)] + \pi_2[b(\theta^2) - c(\theta^2)],$$

which equals $P(\theta^1, \theta^2)$ if $\pi_1[c(\theta^2) - c(\theta^1)] \leq \pi_2[b(\theta^1) - b(\theta^2)]$. Thus, the null contract is best for giving player 2 the incentive to invest if $c(\theta^2) - c(\theta^1)$ is small compared to $b(\theta^1) - b(\theta^2)$ (that is, the own-investment component is small relative to the cross-investment component) and/or if player 2's bargaining weight is large.