Communication and Mechanism Design

Sandeep Baliga and Tomas Sjöström
Northwestern and Rutgers

January 2011 © Sandeep Baliga and Tomas Sjöström. Based on book project *Conflict and Cooperation*. If you use this material, cite the book.
Asymmetric Information and War

- Recall asymmetric information provides one possible source of transactions costs and hence war: “[R]ational leaders may be unable to locate a mutually preferable negotiated settlement due to private information about relative capabilities or resolve and incentives to misrepresent such information” Fearon.

- We earlier presented a signaling model where incomplete information lead to war.

- Is it possible to design a game such that war can be avoided, even though agents have private information? Or is it always the case that there is war?
  - Myerson-Satterthwaite study these two question in a model of bilateral trade.
Mechanism Design: Basic Model

- Based on Bester and Wärneryd [1] and Fey and Ramsay [2].
- Suppose each agent \( i \in \{A, B\} \) has a set of possible types \( T = [\bar{t}, t] \) representing his toughness. Each agent’s type is his private information. The agents’ types are chosen independently from a distribution \( F \).
- The agents can either fight or reach a peaceful agreement. If they reach an agreement they share a joint surplus of size 1. In the event of conflict, the joint surplus is \( 1 - c \), where \( c \equiv c_A + c_B > 0 \). That is, a share \( c \) is destroyed in a conflict. Therefore, conflict is Pareto inefficient in the classical sense. Player \( i \) wins conflict with probability \( p_i \) with \( p_A + p_B = 1 \).
  (Alternative model where \( c_i \) is private cost to player \( i \) of fighting gives similar results)
- If the agents’ knew each others’ types, they would always be able to avoid a conflict (Coase Theorem). However, they do not know each others’ types.
Mechanism Design with Incomplete Information about Costs

- Suppose $p_A$ and $p_B$ are common knowledge but $c_A$ and $c_B$ are private information.

- Then, there is a simple mechanism that guarantees no conflict. Each player announces “participate” or “not participate”. If both say participate, the outcome is a division of the surplus $p_A$ and $p_B = 1 - p_A$. If either says not participate, the outcome is conflict.
Mechanism Design with Incomplete Information about Relative Power

If there is a conflict, agent A’s probability of winning is $p_A(t_A, t_B)$, where $p_A$ is monotonic in both arguments:

$$\frac{\partial p_A(t_A, t_B)}{t_A} > 0, \quad \frac{\partial p_A(t_A, t_B)}{t_B} < 0$$

Agent B’s probability of winning is

$p_B(t_A, t_B) = 1 - p_A(t_A, t_B)$.

Define

$$\hat{c} \equiv 1 - \frac{1}{\int_T [p_A(\bar{t}, s) + p_B(s, \bar{t})] dF(s)}$$

The monotonicity of $p_A$ implies that $0 < \hat{c} \leq \frac{1}{2}$.

Using the revelation principle, we can restrict our attention to incentive-compatible revelation mechanisms. A revelation mechanism specifies a probability of conflict $\pi(t_A, t_B)$, and agent A’s share of the surplus $\alpha(t_A, t_B)$ if there is no conflict, for each type-profile.
A mechanism is *peaceful* if $\pi(t_A, t_B) = 0$ for all $(t_A, t_B) \in T \times T$. The following result shows that conflict can be avoided only if it is sufficiently costly.

**Theorem**

*A peaceful, incentive compatible and individually rational mechanism exists if and only if $c \geq \hat{c}$.***
Consider a peaceful, incentive compatible and individually rational mechanism. If A is of type $t'_A$ and announces that he is of type $t_A$, then his expected payoff is

$$U_A(t_A | t'_A) = \int_T \alpha(t_A, s) dF(s) \quad (2)$$

Incentive compatibility (IC) requires

$$U_A(t'_A | t'_A) \geq U_A(t_A | t'_A)$$

Using (2), we get

$$\int_T \alpha(t_A', s) dF(s) \geq \int_T \alpha(t_A, s) dF(s)$$

for all $(t_A, t'_A)$. This implies there is a constant $K_A$ such that

$$U_A(t_A | t_A) = \int_T \alpha(t_A, s) dF(s) = K_A \quad (3)$$

for all $t_A \in T$. 

Mechanism Design: Proof
The intuition behind (3) is the following: if there is no fighting in equilibrium, then all of A’s types must expect the same share of the surplus (otherwise they would all pretend to be the type which gets the highest share). Since (3) holds for all $t_A$, we can take expectations:

$$K_A = \int_T \left( \int_T \alpha(t, s) dF(s) \right) dF(t) \quad (4)$$
Mechanism Design: Proof c’td

The individual rationality (participation) constraint requires that

$$U_A(\bar{t}|\bar{t}) = K_A \geq \int_T p_A(\bar{t}, s)(1 - c)\,dF(s) \tag{5}$$

where the right hand side is the expected payoff the toughest type $\bar{t}$ expects from fighting. Applying the same reasoning to player B, there is $K_B$ such that

$$K_B = \int_T \left(\int_T (1 - \alpha(s, t))\,dF(s)\right)\,dF(t) \tag{6}$$

for all $t_A \in T$, and

$$U_B(\bar{t}|\bar{t}) = K_B \geq \int_T p_B(s, \bar{t})(1 - c)\,dF(s) \tag{7}$$

Now, add (5) and (7) and use $K_A + K_B = 1$ (which follows from (4) and (6)) to obtain

$$1 \geq \int_T p_A(\bar{t}, s)(1 - c)\,dF(s) + \int_T p_B(s, \bar{t})(1 - c)\,dF(s)$$

which is the same as $c \geq \hat{c}$. 
Conversely, suppose $c \geq \hat{c}$. For all $(t_A, t_B)$, define the following mechanism:

$$\pi(t_A, t_B) = 0$$

$$\alpha(t_A, t_B) = \int_T p_A(\bar{t}, s)(1 - c) dF(s)$$

This mechanism is peaceful, incentive-compatible and individually rational.