**Strumenti della Teoria dei Giochi per l'Informatica** **A.A. 2009/10**

Lecture 13-14: 29-30 Aprile 2010

**Algorithmic Mechanism Design**

Docente *Vincenzo Auletta* Note redatte da: *Alessandra Scafuro*

# 1 Introduzione

Nella prima parte del corso ci siamo occupati dello studio degli equilibri. Quella parte della Teoria dei Giochi si occupava sostanzialmente di analizzare delle situazioni, formalizzandole come giochi in modo da poterne prevedere l'evoluzione e la convergenza ad una qualche situazione di equilibro, consistente con un certo concetto di soluzione. Pertanto, l'obiettivo primario era di formalizzare e analizzare il gioco per capire quale concetto di soluzione soddisfa.
In questa seconda parte del corso, invece, ci occupereremo di Mechanism Design, una branca della Teoria dei Giochi e delle Teoria Economica, che non si occupa di analizzare oggettivamente un gioco ma piuttosto di progettare meccanismi che conducano i giocatori a fare delle scelte che siano in accordo con il concetto di soluzione che ci siamo prefissi di ottenere. In un certo senso, il progettista prende parte attiva al gioco, modificandone le regole in modo che determinati equilibri siano ottenibili. Per esempio, faremo in modo che in molti giochi esistano delle soluzioni in equilibrio rispetto a strategie dominanti (cosa non vera in generale) e che le soluzioni in equilibrio siano "buone" dal punto di vista sociale.
Per iniziare, esaminiamo il setting del Mechanism Design andando a sottolineare le differenze con quello della Teoria dei Giochi. Per farlo, utilizzeremo il framework della Teoria delle Scelta Sociali. In un problema di Scelta Sociale abbiamo $n$ giocatori, egoisti e razionali che devono scegliere uno di $m$ outcome, chiamati alternative. Ciascun giocatore ha una sorta di classifica su queste alternative e questa classifica è privata. Ogni giocatore fornisce la sua classifica ad una entità chiamata appunto meccanismo, che prese tutte le classifiche le aggrega dando in output una classifica che deve soddisfare al meglio le preferenze dei giocatori. Per questo motivo la funzione di aggregazione viene definito *funzione di benessere sociale.* L'obiettivo di ciascun giocatore è di indurre il meccanismo a dare in output una classifica il più possibile simile alla sua classifica personale. Per ottenere ciò, il giocatore potrebbe mentire dando in input al meccanismo una classifica diversa da dalle sue preferenze reali. Questo framework è estremamente generale e trova applicazione nei campi più disparati. Ad esempio:

**Elezioni.** Gli elettori hanno diverse preferenze sui candidati ed ognuno vota tenendo presente delle scelte degli altri per fare eleggere il candidato preferito. L'insieme dei voti decide il candidato eletto.

**Mercati.** La teoria economica assume l'esistenza del mercato ma abbiamo solo delle interazioni tra operatori regolate da protocolli. In queste interazioni ogni partecipante ha le sue preferenze che portano ad una distribuzione di beni e denaro.

**Aste.** Il venditore deve vendere un bene ad un insieme di acquirenti . Le loro preferenze sul bene sono integrate per decidere chi vince l'asta.

**Politica.** Governare può essere visto come un processo di integrazione delle preferenze dei cittadini al fine di prendere delle decisioni.

Il modello teorico fornito dal Mechanis Design può essere proficuamente utilizzato per studiare il funzionamento di Internet e delle moderne reti. L'interrete, infatti, è costituita da reti e server posseduti e gestiti da operatori diversi, i quali hanno i propri obiettivi (diversi) e le loro preferenze. Ogni protocollo che deve operare sulle reti deve essere visto come una aggregazione delle preferenze dei partecipanti.

**Framework della Teoria delle Scelte Sociali** In questa sezione definiamo più formalmente il nuovo setting del Mechanism Design e la notazione che verrà usata nelle sezioni successive. come detto nell'introduzione, utilizzeremo il framework della Teoria delle Scelte Sociali. Un problema di Mechanism Design è rappresentato da:

**Insieme delle alternative.** $A = \{a, b, c....\}$

**Insieme dei giocatori.** $\{i\}_{1 \geq i \leq n}$. Ciascun giocatore ha un ordine di preferenze tra l'insieme delle alternative. Indichiamo la relazione di preferenze dell'$i$-esimo giocatore con il simbolo $>_i = b > a > d > \ldots > \ldots$. Questa relazione è privata. Il profilo $>_i'$ inviato dal giocatore $i$ al meccanismo potrebbe essere differente da quello reale.

**Profilo di preferenze.** Sia $L$ l'insieme di tutti le possibili relazioni d'ordine sull'insieme delle alternative $A$. ($L$ è isomorfo all'insieme delle permutationi di $A$). Un profilo di preferenze è un insieme delle relazioni di prefenza degli $n$ giocatori: $\pi = (>_1, >_2, \ldots, >_n) \in L^n$

**Funzione di benessere sociale.** La funzione di benessere sociale $F$ è la formalizzazione del comportamento del meccanismo. L'input di $F$ è rappresentato da tutti i profili dichiarati dai giocatori. L'output di $F$ è ancora una relazione di preferenza che è un'aggregazione delle preferenze dei giocatori verso una decisione congiunta. Indichiamo con $>_S = F(>_1, >_2, \ldots, >_n)$, dove la $S$ in pedice indica la relazione di preferenza sociale. Formalmente: $F : L^n \rightarrow >_S$.

**Funzione di scelta sociale.** La funzione di scelta sociale $f$ prende in input un profilo di preferenze $\pi = (>_1, >_2, \ldots, >_n)$ e restituisce un'alternativa $k \in A$. Formalmente: $f : L^n \rightarrow k$.

**Obiettivo del meccanismo.** La funzione sociale deve essere tale da indurre tutti i giocatori a dichiarare la reale relazione di preferenze.

**Paradosso di Condorcet.** Il seguente esempio, noto come il paradosso di Condorcet, dal nome del marchese francese che lo ha proposto alla fine del XVIII secolo, evidenzia come possa essere impossibile avere una funzione di scelta sociale che accontenti la maggior parte dei giocatori. L'esempio è il seguente. Consideriamo lo scenario di un'elezione. Ci sono tre candidati, le alternative $a, b$ e $c$, e ci sono tre votanti, i giocatori $p_1, p_2$ e $p_3$. Le rispettive relazioni di preferenze sono le seguenti.

$p_1$ $a >_1 b >_1 c$
$p_2$ $b >_2 c >_2 a$
$p_3$ $c >_3 a >_3 b$

Consideriamo la funzione di scelta sociale che sceglie il vincitore dell'elezione. Notiamo che chiunque scegliamo come vincitore, ci sarà una maggioranza di elettori che preferirebbe un altro candidato al vincitore scelto da $f$. Il paradosso di Condorcet evidenzia le debolezze intrinsiche nella scelta della maggioranza come funzione di scelta sociale. Funzioni alternative sono state proposte, ma se da un lato non cadono nel paradosso di Condorcet, dall'altro incoraggiano la votazione strategica (esprimere preferenze differenti da quelle reali per influenzare la funzione di aggregazione).

**Proprietà della funzione di benessere sociale** Sarebbe auspicabile che la funzione di benessere sociale soddisfacesse le seguenti proprietà:

**Unanimità.** se tutti i giocatori hanno espresso la stessa relazione tra due alternative allora queste due alternative saranno poste nello stesso ordine anche nella relazione data in output dalla funzione sociale. Formalmente:

$$\text{se} \quad \forall i \quad a >_i b \quad \text{allora} \quad a >_S b \tag{1}$$

**Indipendenza dalle alternative irrilevanti.** Questa proprietà mette in relazione una coppia di alternative in due profili di preferenza differenti. In particolare quello che si richiede ò che, se una coppia di alternative è posta nello stesso ordine di preferenza da tutti i giocatori in due profili differenti, indipendentemente dalle posizioni in cui sono poste le altre alternative, si ha che anche nelle relazioni d'ordine date in output da $F$ avendo in input i due profili, deve valere lo stesso ordine preferenza.

$$\begin{array}{l} \forall \quad a, b \quad \forall \quad (>_1, >_2, \ldots, >_n) \quad \text{e} \quad (>'_1, >'_2, \ldots, >'_n) \\ \quad \text{siano} \quad F(>_1, >_2, \ldots, >_n) = \quad >_S \quad F(>'_1, >'_2, \ldots, >'_n) = \quad >'_S \quad \text{allora} \\ \text{se} \quad \forall \quad i \quad a >_i b \Leftrightarrow a >'_i b \quad \Rightarrow \quad a >_S b \Leftrightarrow a >'_S b \end{array} \tag{2}$$

**Dittatura** $F$ è una dittatura quando restituisce sistematicamente una relazione d'ordine identica a quella di un solo giocatore, chiamato per questo dittatore, indipendentemente dalle preferenze degli altri giocatori. Formalmente:

$$\exists i \quad \text{t.c.} \quad \forall >_1, >_2, \ldots, >_n \quad F(>_1, >_2, \ldots, >_n) = \quad >_i \tag{3}$$

Ci aspettiamo che una funzione sociale è una buona funzione sociale se soddisfa le proprietà **??** e **??** e ovviamente non sia una dittatura (proprietà **??**). Il prossimo teorema dimostra che queste semplici richieste sono inconciliabili, nel senso che non è possibile ottenerle tutte e tre contemporaneamente.

## 2 Teorema di impossibilità di Arrow

Il teorema di Arrow dimostra che l'unica funzione sociale che soddisfa le proprietà di unanimità e di indipendenza dalle alternative irrilevanti è una dittatura. Prima di enunciare il teorema, mostriamo il lemma di pairwise neutrality che sarà usato nella dimostrazione.

**Lemma 1 (Pairwise neutrality)** *Sia $F$ una funzione di benessere sociale che soddisfa le proprietà di unanimità e di i.i.a. Siano $(>_1, >_2, \ldots, >_n)$e $(>'_1, >'_2, \ldots, >'_n)$due profili di preferenze ed esistono $a, b, c, d \in A$ tali che $\forall i \quad a >_i b \Leftrightarrow c >'_i d$. Allora, se indichiamo con $>_S = F(>_1, >_2, \ldots, >_n)$e $>'_S = F(>'_1, >'_2, \ldots, >'_n)$, si ha che $a >_S b \Leftrightarrow c >'_S d$.*

*Proof.* Per ipotesi esistono due profili di preferenze $(>_1, >_2, \ldots, >_n)$e $(>'_1, >'_2, \ldots, >'_n)$tali che vale:

$$\forall i \quad a >_i b \Leftrightarrow c >'_i d \tag{4}$$

vogliamo dimostrare che per le relazioni $>_S$=F$(>_1, >_2, \ldots, >_n)$e $>'_S$=F$(>'_1, >'_2, \ldots, >'_n)$deve valere che:

$$a >_S b \Leftrightarrow c >'_S d. \tag{5}$$

Per semplicità consideriamo soltanto il caso in cui $b \neq c$ e dimostriamo solo l'implicazione $\Rightarrow$ (per l'altro verso la dimostrazione è analoga).

Consideriamo la relazione del generico giocatore $i$, $>_i$e costruiamo a partire da $>_i$una nuova relazione $\bar{>}_i$dove $c$ è posta immediatamente prima di $a$ e $d$ è posta immediatamente dopo $b$. Quindi, $\bar{>}_i$sarà di uno di questi due tipi:

$$\begin{pmatrix} \vdots \\ c \\ a \\ \vdots \\ b \\ d \\ \vdots \end{pmatrix} \qquad\qquad \begin{pmatrix} \vdots \\ b \\ d \\ \vdots \\ c \\ a \\ \vdots \end{pmatrix}$$

Pertanto, per costruzione valgono le seguenti condizioni:

$$c \bar{>}_i a \quad e \quad b \bar{>}_i d$$

Poichè $a$ e $b$ sono in $\bar{>}_i$ nella stessa relazione in cui erano in $>_i$, abbiamo che

$$a \bar{>}_i b \quad \Leftrightarrow \quad c \bar{>}_i d \tag{6}$$

Dalle equazioni **??** e **??** otteniamo che

$$\forall i \quad c \bar{>}_i d \Leftrightarrow c >'_i d$$

Inoltre, applicando la proprietà di i.i.a (prop. **??**) otteniamo che

$$c \bar{>}_S d \Leftrightarrow c >'_S d, \tag{7}$$

dove $\bar{>}_S$= F$(\bar{>}_{S,1}, \bar{>}_{S,2}, \ldots, \bar{>}_{S,n})$.
Consideriamo ora le relazioni di preferenze prodotte dalla funzione di benessere sociale $F$ quando riceve in input i tre profili $(>_1, >_2, \ldots, >_n)$, $(>'_1, >'_2, \ldots, >'_n)$ e $(\bar{>}_1, \bar{>}_2, \ldots, \bar{>}_n)$. Ricordiamo che, per ipotesi, la $F$ soddisfa le proprietà di unanimità e di i.i.a.
Prima di tutto osserviamo che per costruzione per ogni giocatore $i$ vale che $a >_i b \Leftrightarrow a \bar{>}_i b$. Quindi, per la proprietà di i.i.a. di $F$ abbiamo che

$$a >_S b \Leftrightarrow a \bar{>}_S b \tag{8}$$

Inoltre, dalla **??** segue che per ogni giocatore $i$ vale che $a \bar{>}_i b \Leftrightarrow c \bar{>}_i d$. Quindi, per la proprietà di i.i.a. di $F$ abbiamo che

$$a \bar{>}_S b \Leftrightarrow c \bar{>}_S d \tag{9}$$

Mettendo insieme tutte le proprietà dimostrate otteniamo che

$$\begin{aligned} a >_S b \quad &\Leftrightarrow \quad a\bar{>}_S b \quad (\text{per } \textbf{??}) \\ &\Leftrightarrow \quad c\bar{>}_S d \quad (\text{per } \textbf{??}) \\ &\Leftrightarrow \quad c >'_S d \quad (\text{per } \textbf{??}) \end{aligned}$$

e questo chiude la prova.
Siamo ora nelle condizioni di enunciare e provare il Teorema di Impossibilità di Arrow.

**Theorem 2 (Impossibilità di Arrow)** *Qualunque funzione di scelta sociale $F$ che soddisfa la proprietà di unanimità (prop. **??**) e i.i.a. (prop. **??**) è una dittatura (prop. **??**).*

*Proof.* Sia $F$ una funzione di benessere sociale che aggrega le relazioni di preferenza di $n$ giocatori e sia $>_S$ la relazione di preferenza prodotta dalla $F$ avendo ricevuto in input un profilo di relazioni di preferenza $(>_1, >_2, \ldots, >_n)$. Vogliamo far vedere che esiste un indice $i^*$ tale che $>_S = >_{i^*}$.
Fissiamo arbitrariamente due outcome $a, b \in A$ e per ogni $i = 0, 1, \cdots, n$ costruiamo un profilo di preferenze $\pi^i = (\succ^i_1, \succ^i_2, \cdots, \succ^i_n)$ in cui tutti i giocatori fino al giocatore $i$-esimo preferiscono l'outcome $a$ rispetto a $b$ ($a \succ^i_j b$ for $j \leq i$), mentre tutti i restanti giocatori preferiscono $b$ ad $a$ ($b \succ^i_j a$ for $j > i$).
Sia $\succ^i = F(\pi^i)$ la relazione di preferenza prodotta dalla funzione $F$ sul profilo $\pi^i$. Osserviamo che in $\pi^0$ tutti i giocatori preferiscono $b$ ad $a$ e quindi per la proprietà di unanimità $b \succ^0 a$. In maniera analoga, in $\pi^n$ tutti i giocatori preferiscono $a$ a $b$ e quindi $a \succ^n b$. Se le relazioni prodotte dalla $F$ su $\pi^0$ e $\pi^n$ danno preferenze opposte tra gli outcome $a$ e $b$ allora ci deve essere un indice $i^*$ tale che

$$a \succ^{i^*} b, \qquad b \succ^{i^*-1} a.$$

Vogliamo provare che, dato che $F$ soddisfa le proprietà di unanimità e i.i.a, allora il giocatore $i^*$ che abbiamo individuato, 'e un dittatore. Per fare questo dobbiamo provare che per ogni coppia di alternative, la relazione scelta dalla funzione di benessere sociale tra queste due alternative è la stessa esistente in $>_{i^*}$.
Consideriamo una coppia di alternative $c, d \in A$ con $c \neq d$ e, senza perdita di generalità assumiamo che $c >_{i^*} d$. Consideriamo, inoltre, un terzo outcome $e$ che sia diverso sia da $c$ che da $d$. Per la proprietà i.i.a. possiamo spostare $e$ in un qualunque punto delle relazioni di preferenze $>_i$ dei giocatori senza modificare la relazione di preferenza tra $c$ e $d$ prodotta dalla F. Pertanto modifichiamo le relazioni di preferenza del profilo $(>_1, >_2, \ldots, >_n)$ in modo da avere $e$ al primo posto nella relazione $>_j$, per ogni $j < i^*$, all'ultimo posto della relazione $>_j$, per ogni $j > i^*$ e compresa tra $c$ e $d$ nella relazione $>^{i^*}$.
Osserviamo che il profilo di strategie modificato è tale che

$$\forall i \quad e >_i c \Leftrightarrow a \succ^{i^*} b$$

da cui per il Lemma **??** ricaviamo che

$$e >_S c \Leftrightarrow a >^{i^*} b. \tag{10}$$

Analogamente, per il profilo $\pi^{i^*-1}$ abbiamo che

$$\forall i \quad e >_i d \Leftrightarrow a \succ^{i^*-1} b$$

e per il Lemma **??** ricaviamo che

$$d >_S e \Leftrightarrow a >^{i^*-1} b. \tag{11}$$

Poichè, per costruzione, $a >^{i^*} b$ e $b >^{i^*-1} a$, dalle **??** e **??** otteniamo che $c >_S e$ e $e >_S d$, da cui per transitività ricaviamo che $c >_S d$. In conclusione, abbiamo provato che

$$\forall c \neq d \quad c >_S d \Leftrightarrow c >_{i^*} d$$

e quindi $i^*$ è un dittatore.

□

## 3 Le funzioni di scelta sociale e il Teorema di Gibbard-Satterthwaite

Nella sezione precedente abbiamo fornito un importante risultato negativo riguardante le funzioni di benessere sociale. In questa sezione cercheremo di capire se questo risultato di impossibilità vale anche per le funzioni di scelta sociale.
Diamo prima alcune definizioni che riguardano le funzioni di scelta sociale, e poi mostriamo che purtroppo, anche le funzioni di scelta sociale obbediscono alla stessa regola di impossibilità.
Ricordiamo che una funzione di scelta sociale $f : (>_1, >_2, \ldots, >_n) \to k$ con $k \in A$.

**Definition 3** *Una funzione di scelta sociale $f$ è strategicamente manipolabile dal giocatore $i$ se, dati $(>_1, >_2, \ldots, >_n)$ e $>_i$, esiste un outcome $a' \in A$ con $f(>_1, >_2, \ldots, >_n) = a$ e $f(<_1, \ldots, <_{i-1}, <'_i, \ldots, <_n) = a'$ tale che $a' >_i a$*

**Definition 4** *Una funzione di scelta sociale $f$ è compatibile agli incentivi se non è strategicamente manipolabile.*

**Definition 5** *Una funzione di scelta sociale $f$ è monotona se $f(>_1, >_2, \ldots, >_n) = a \neq a' = f(>_1, \ldots, >_{i-1}, >'_i, \ldots, >_n)$ implica che $a >_i a'$ e $a' >'_i a$.*

**Definition 6** *Una funzione di scelta sociale $f$ è una dittatura se esiste un player $i$ e $\forall \pi = (>_1, >_2, \ldots, >_n)$, $\forall b \neq a$, $a >_i b \Rightarrow a = f(>_1, >_2, \ldots, >_n)$ ed il giocatore $i$ è detto dittatore.*

E' possibile caratterizzare le funzioni di scelta sociale compatibili agli incentivi in termini di monotonicità.

**Theorem 7** *$f$ è una funzione compatibile agli incentivi se e solo se è monotona.*

Sfortunamente anche per le funzioni di scelta sociale vale un risultato di impossibilità analogo a quello ottenuto da Arrow.

**Theorem 8** *Una funzione di scelta sociale $f$ su $|A| \geq 3$ è compatibile agli incentivi se e solo se è una dittatura.*

I risultati di Arrow e Gibbard-Satterthwaite ci dicono praticamente che l'unico "meccanismo" che soddisfa le proprietà che auspichiamo è proprio la dittatura. Questi risultati che apparentemente chiudono definitivamente la discussione sulla progettazione di meccanismi, sono in realtà il punto di partenza da cui si sviluppa la teoria del Mechanism Design. Il Mechanism Design infatti fornisce una serie di tecniche usate per progettare meccanismi che implementano funzioni di scelta sociale compatibili agli incentivi e che non siano dittature.

# 4 Meccanismi con pagamenti

In questa sezione mostriamo una delle tecniche usate (l'introduzione del concetto del pagamento) e i risultati positivi raggiunti (i meccanismi di VCG).

**Introduzione di denaro** La prima osservazione riguarda le relazioni di preferenza dei giocatori. Una relazione di preferenza rappresenta un ordinamento sul grado di apprezzamento da parte di un giocatore degli outcome del gioco. Questo modo di rappresentare le preferenze fornisce solo informazioni qualitative (un outcome piace più di un altro). Nel nuovo modello si richiede ai giocatori di fornire anche informazioni *quantitative* riguardo le proprie preferenze (quanto un outcome è più gradito di un altro). Ogni giocatore dovrà dichiarare quanto apprezza ciascun outcome. Una maniera per descrivere il grado di apprezzamento di un outcome consiste nell'indicare quanto il giocatore è disposto a pagare affinchè venga ottenuto. Pertanto introduciamo il denaro come scala numerica rispetto alla quale confrontare le preferenze.
Assumiamo di avere per ogni giocatore $i$ una funzione di valutazione

$$v_i : A \Rightarrow \mathbb{R} \quad \text{tale che} \quad \forall a, b \in A \quad a >_i b \quad \Leftrightarrow v_i(a) >_i v_i(b)$$

Se assumiamo che il denaro è trasferibile possiamo considerare meccanismi che richiedono ai giocatori di pagare un costo, che indichiamo con $p$, dipendente dall'outcome del meccanismo e della valutazione dello specifico giocatore rispetto a quell'outcome.
Definiamo pertanto l'*utilità* dell'$i$-esimo giocatore rispetto all'outcome $a$ nel modo seguente:

$$u_i = v_i(a) + p_i(a)$$

Ciascun giocatore si comporta strategicamente per massimizzare la sua utilità. Questo significa che potrebbe fornire al meccanismo valutazioni diverse da quelle reali.

**L'asta di Vickrey** Il meccanismo proposto da Vickrey per lo scenario dell'asta è un semplice esempio di applicazione del mechanism design nel modello modificato con i pagamenti. Lo scenario dell'asta è il seguente. Ci sono $n$ giocatori, un oggetto da assegnare, $n$ offerte e un venditore che prende in input le offerte degli $n$ giocatori e deve scegliere il giocatore che si aggiudica l'oggetto.
Ciascun giocatore ha una sua valutazione dell'oggetto $v_i$. Il venditore rappresenta il meccanismo, cioè la funzione di scelta sociale che, prese in input le valutazioni, stabilisce il vincitore assegnandogli un pagamento. L'utilità del giocatore $i$ è data dalla differenza tra la sua valutazione dell'oggetto $v_i$ e il pagamento assegnatogli dal meccanismo $p_i$. Se il giocatore non vince l'asta, invece, la sua utilità è 0. Essendo razionale, un giocatore fornirà al meccanismo una valutazione, possibilmente diversa da quella reale, che massimizzi la sua utilità minimizzando il pagamento richiestogli. Pertanto, l'obiettivo del progettista del meccanismo è creare una funzione di scelta sociale tale che ciascun giocatore è incentivato a fornire la sua vera valutazione, cioè ottiene la sua massima utilità quando fornisce le valutazioni reali al meccanismo. Un meccanismo che soddisfa questa proprietà è l'asta del secondo prezzo di Vickrey.

**Definition 9 (Vickrey's second price auction)** *Sia i il vincitore dell'asta in quanto ha dichiarato la più alta valutazione $v_i$. Allora, il pagamento assegnato a i è:*

$$p = max_{i \neq j} w_j$$

**Proposition 10 (Vickrey)** *Per ogni $v_1, \ldots, v_i, \ldots, v_n$ (valutazioni reali) e per ogni $v_i'$. Sia $u_i$ l'utilità del giocatore $i$ se dichiara $v_i$ e sia $u_i'$ la sua utilità se dichiara $v_i'$. Allora, $u_i \geq u_i'$.*

*Proof.*

**Caso A: $i$ con valutazione $v_i$ sarebbe stato il vincitore.** In questo caso l'utilità di $i$ dichiarando $v_i$ è

$$u_i = v_i - p$$

Sia $v_i'$ la valutazione falsa fornita al meccanismo abbiamo i seguenti casi:

1. $v_i' > v_i$ in questo caso $i$ sicuramente è ancora il vincitore, e abbiamo che $u_i' = u_i$
2. $v_i' < v_i$ qui abbiamo che, se $i$ è ancora il vincitore allora $u_i' = u_i$, mentre se $i$ non è il vincitore si ha che $u_i' = 0 \leq u_i$

**Caso B: $i$ con valutazione $v_i$ avrebbe perso e $j$ con valutazione $v_j$ sarebbe stato il vincitore.** In questo caso l'utilità di $i$ dichiarando $v_i$ è $u_i = 0$ Sia $v_i'$ la valutazione falsa fornita al meccanismo, assumiamo che $v_i' > v_i$ (in caso $v_i' < v_i$ l'utilità di $i$ è ancora 0) abbiamo i seguenti casi:

- se $j$ è ancora il vincitore allora $u_i' = u_i = 0$
- se $i$ diventa il vincitore, allora $v_j$ sarebbe la seconda migliore offerta. Essendo $j$ vincitore se $i$ avesso giocato la sua reale valutazione, abbiamo che vale $v_j \geq v_i$, pertanto $u_i' = v_i - v_j \leq 0 \leq u_i$.

## 4.1 Meccanismi VCG

Un meccanismo $f$ è composto da una funzione di scelta di un'alternativa in $A$ e da una funzione di pagamento. Le preferenze di ciascun giocatore sono modellate dalla funzione di valutazione

$$v_i : A \Rightarrow \mathbb{R}$$

Indichiamo con $V_i$ l'insieme di tutte le possibili valutazioni del giocatore $i$, per cui abbiamo $v_i \in V_i$ e $V_i \subseteq \mathbb{R}^A$. Inoltre indichiamo con $v = (v_1, \ldots, v_n)$ il vettore $n$-dimensionale delle valutazioni di tutti i giocatori, mentre indichiamo con $v_{-i} = (v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ il vettore di $n-1$ elementi costituito da tutte le valutazioni esclusa quella del giocatore $i$. Infine denotiamo $V = V_1 \times, \ldots, \times V_n$ l'insieme di tutte le possibili valutazioni.

**Definition 11 (Meccanismo a rivelazione diretta.)** *Un meccanismo a rivelazione diretta è rappresentato da una funzione di scelta sociale $f : V = V_1 \times, \ldots, \times V_n \to A$ , e un vettore di funzioni di pagamento $\{p_1, p_2, \ldots, p_n\}$, dove $p_i : V_1 \times, \ldots, \times V_n \to \mathbb{R}$*

**Definition 12 (Meccanismo compatibile agli incentivi)** *Un meccanismo $(f, \{p_1, p_2, \ldots, p_n\})$ è compatibile agli incentivi se per ogni giocatore $i$, per ogni $v_1 \in V_1, \ldots, v_n \in V_n$ e per ogni $v_i' \in V_i$, sia $a = f(v_i, v_{-i})$ e sia $a' = f(v_i', v_{-i})$, allora si ha che $v_i(a) - p_i(v_i, v_{-i}) \geq v_i(a') - p_i(v_i', v_{-i})$*

Pertanto, definire un meccanismo significa definire:

1. la funzione di scelta sociale $f$.
2. la funzione dei pagamenti $p$.

**Definition 13 (Meccanismo Vickrey-Clarke-Groves)** *Un meccanismo* $(f, p_1, \ldots, p_n)$ *è detto meccanismo Vickrey-Clarke-Groves(VCG) se:*

- *Funzione sociale:* $f(v_1, \ldots, v_n) = argmax_{a \in A} \sum_i v_i(a)$
- *Funzione di pagamento:* $\forall i,\ p_i(v_i, v_{-i}) = h_i(v_{-i}) - \sum_{i \neq j} v_j(f(v_i, v_{-i}))\ \forall h : V_{-i} \to \mathbb{R}$

L'idea chiave del meccanismo VCG è che il pagamento calcolato per un singolo giocatore è determinato da una funzione che è indipendente dalla valutazione del giocatore stesso. In questo modo la valutazione del singolo giocatore non contribuisce a modificare il proprio pagamento. Inoltre, per come è definita la funzione sociale, si ha che essa massimizza il benessere sociale (somma le valutazioni dei singoli agenti). Queste due osservazioni, ci portano a concludere che l'unica azione strategica che massimizza l'utilità di ogni singolo giocatore, è dare alla funzione sociale la reale valutazione.

**Theorem 14** *Il meccanismo VCG è compatibile agli incentivi*

*Proof.* Fissiamo $i$,$v_{-i}, v_i$ e $v'_i$. Dobbiamo dimostrare che, per qualunque giocatore $i$ la cui reale valutazione è $v_i$, l'utilità che ottiene quando dichiara $v_i$ non è mai minore dell'utilità che otterrebbe dichiarando qualunque $v'_i \neq v_i$. Sia $a = f(v_i, v_{-i})$ e $a' = f(v'_i, v_{-i})$.

- Se $i$ dichiara $v_i$ ottiene un'utilità pari a $u_i = v_i(a) + \sum_{j \neq i} v_j(a) - h_i(v_{-i})$
- Se $i$ dichiara $v'_i$ ottiene un'utilità pari a $u_i = v_i(a') + \sum_{j \neq i} v_j(a') - h_i(v_{-i})$

Siccome $a = f(v_i, v_{-i})$ per definizione è tale che massimizza le utilità rispetto a tutte le possibili alternative si ha sempre che

$$v_i(a) + \sum_{j \neq i} v_j(a) - h_i(v_{-i}) \geq v_i(a') + \sum_{j \neq i} v_j(a') - h_i(v_{-i})$$

e la stessa diseguaglianza vale anche sottraendo $h_i(v_{-i})$ ad entrambe le espressioni.
Ci resta da determinare come viene calcolata la funzione $h_i$ e che impatto ha sul pagamento finale del giocatore. Intuitivamente vorremmo che ciascun giocatore paghi una somma di denaro al meccanismo, ma questa somma non deve mai superare l'utilità che il giocatore ottiene. In generale auspichiamo che siano soddisfatte le seguenti condizioni:

- Razionalità individuale(utilità non negativa): $\forall v_1, \ldots, v_n \quad u_i(v_1, \ldots v_n) \geq 0$
- Non trasferimento di denaro: $\forall v_1, \ldots, v_n \quad p_i(v_1, \ldots, v_n) \geq o$

Siamo ora pronti per mostrare la funzione $h_i$ proposta da Clarke che soddisfa entrambe le proprietà enunciate sopra.

**Definition 15 (Regola di pivot di Clarke)**

$$h_i(v_{-i}) = max_{b \in S} \sum_{j \neq i} v_j(b)$$

Un meccanismo VCG che utilizza la regola di Pivot di Clarke per i pagamenti soddisfa le proprietà di razionalità individuale e del non trasferimento.

$$
\begin{aligned}
p_i(v_1, \ldots, v_n) &= max_{b \in A} \sum_{i \neq j} v_j(b) - \sum_{i \neq j} v_j(a) \\
&= max_{b \in A} (\sum_{i \neq j} v_j(b) - v_j(a)) \\
&\geq 0
\end{aligned}
$$

$$
\begin{aligned}
u_i(v_1, \ldots, v_n) &= v_i \cdot max_{b \in A} \sum_{i \neq j} v_j(b) + \sum_{i \neq j} v_j(a) \\
&= \sum_{i} v_i(a) - max_{b \in A} (\sum_{i \neq j} v_j(b) \\
&\geq \sum_{i} (v_i(a) - v_i(b)) \\
&\geq 0
\end{aligned}
$$