



Game Theory Cheat Sheet

☆ 223 stars  43 forks

 Starred

 Watch ▾

 **Code**  Issues  Pull requests 1  Actions  Projects  Wiki  Security 



medvedev1088 Fix typo ...

on 29 Jun 2018

🕒 60

[View code](#)



README.md

Releases

Game Theory Cheat Sheet

No releases published

Game theory is the study of mathematical models of conflict and cooperation between intelligent rational decision-makers. Game theory is mainly used in economics, political science, and psychology, as well as logic, computer science and biology.

Table of Contents

- [Game Representations](#)
- [Classification of Games](#)
- [Commonly Known Games](#)
- [Nash Equilibrium](#)

No packages published

- Extensive-Form Games

- Repeated Games

- Bayesian Games

Contributors 2

- Coalitional Games



medvedev-evgeny

- Social Choice



medvedev1088 Evgeny Medvedev

- Mechanism Design

- VCG

- Cryptoeconomics

- Resources

Game Representations

Players: who are the decision makers?

- People? Governments? Companies? Somebody employed by a Company?...

Actions: what can the players do?

- Enter a bid in an auction? Decide whether to end a strike? Decide when to sell a stock? Decide how to vote?...

Payoffs: what motivates players?

- Do they care about some profit? Do they care about other players?...

Normal form

- Lists what payoffs players get as a function of their actions.
- Does not incorporate any notion of sequence, or time, of the actions of the players.
- Usually represented by a matrix.

	Stay silent	Betray
Stay silent	-1,-1	-3,0
Betray	0,-3	-2,-2

Extensive form

- Includes timing of moves.

- Usually represented by a tree.

Extensive Form Games

Positive affine transformation: $au + b$, where $a > 0$ and b is any real number. Expected utilities are identical to positive affine transformations.

Classification of Games

<p>Simultaneous games (a.k.a. Strategy games) - games where both players move simultaneously, or if they do not move simultaneously, the later players are unaware of the earlier players' actions (making them effectively simultaneous).</p> <p>Usually normal form is used to represent simultaneous games.</p>	<p>Sequential games (a.k.a. Extensive game) - games where later players have some knowledge about earlier actions.</p> <p>Usually extensive form is used to represent sequential games.</p>
<p>Cooperative games - games where the players are able to form binding commitments externally enforced (e.g. through contract law).</p>	<p>Non-cooperative games - games where players cannot form alliances or if all agreements need to be self-enforcing (e.g. through credible threats).</p>
<p>Zero-sum games - games in which each participant's gain or loss of utility is exactly balanced by the losses or gains of the utility of the other participants.</p>	<p>Non-zero-sum games - games in which the interacting parties' aggregate gains and losses can be less than or more than zero.</p>
<p>Perfect information games - games in which all players know the moves previously made by all other players.</p>	<p>Imperfect information games - games in which some players don't know the moves previously made by other players.</p>
<p>Complete information games - games in which all players know the strategies and payoffs</p>	<p>Incomplete information games - games in which some players</p>

available to the other players.	don't know the strategies or payoffs available to the other players.
Finite games - games that last for finite number of moves.	Infinite games - games that last for infinite number of moves.

Commonly Known Games

Prisoner's Dilemma

	Stay silent	Betray
Stay silent	-1,-1	-3,0
Betray	0,-3	-2,-2

- There is 1 NE:
 - Both players betray.
- This is a [dominant strategies NE](#).
- It is also the only non [Pareto optimal outcome](#) in this game.

Coordination Games

- [Battle of the Sexes](#)
- [Stag Hunt](#)
- [Pure Coordination Game](#)

Matching Pennies

	Play heads	Play tails
Play heads	1,-1	-1,1
Play tails	-1,1	1,-1

- There is 1 NE:
 - Mixed strategies NE: each player chooses heads or tails with probability 0.5.
- It's a zero-sum game.

Battle of the Sexes

	Go to opera	Go to football
Go to opera	3,2	0,0
Go to football	0,0	2,3

- There are 3 NE:
 - Both players go to opera
 - Both players go to football
 - Mixed strategies NE: He goes to opera with probability $\frac{3}{5}$, She goes to opera with probability $\frac{2}{5}$.

Stag Hunt

	Hunt stag	Hunt hare
Hunt stag	2,2	0,1
Hunt hare	1,0	1,1

- There are 3 NE:
 - Both players hunt stag
 - Both players hunt hare
 - Mixed strategies NE: He hunts stag with probability $\frac{1}{2}$, She hunts stag with probability $\frac{1}{2}$.

Pure Coordination Game

	Go to Party	Stay Home
Go to Party	10,10	0,0
Stay Home	0,0	5,5

- There are 3 NE:
 - Both players go to party
 - Both players stay home
 - Mixed strategies NE: He goes to party with probability $\frac{1}{3}$, She goes to party with probability $\frac{1}{3}$.

Free Money

	Play heads	Play tails
Play heads	1,1	0,0
Play tails	0,0	0,0

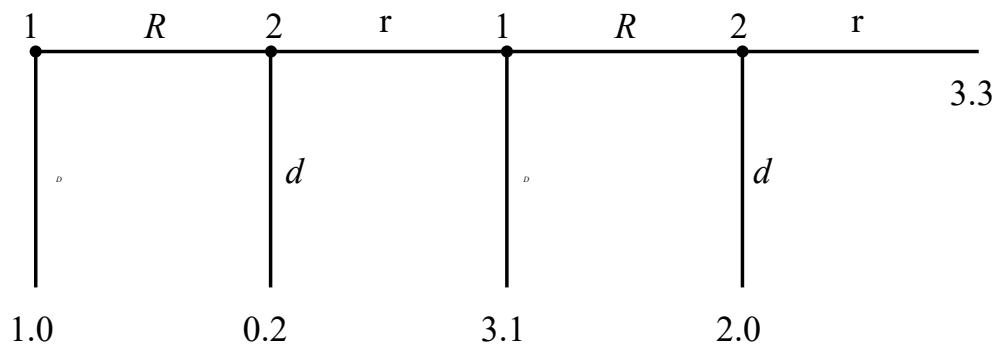
- There are 2 NE (a rare case when there is an even number of NE, see [The Oddness Theorem](#)):
 - Both players play heads
 - Both players play tails

Chicken Game

	Swerve	Straight
Swerve	0,0	-1,1
Straight	1,-1	-10,-10

- Anti-coordination games - it is mutually beneficial for the players to play different strategies.
- There are 3 NE:
 - He swerves, she goes straight
 - He goes straight, she swerves
 - Mixed strategy NE: He swerves with probability $9/10$, She swerves with probability $9/10$

Centipede Game



- SPE: each player chooses to defect at every opportunity.

Nash Equilibrium

Nash Equilibrium - a set of strategies, one for each player, such that no player has incentive to change his strategy given what the other players are doing.

Nash Equilibrium (alternative definition) - a set of strategies, one for each player, such that every player's strategy is the best response to what the other players are doing.

Best Response - a strategy such that a player cannot gain more utility from switching to a different strategy, given what all other players are doing.

Mixed Strategy - a probability distribution over two or more pure strategies, that is, the players choose randomly among their options in equilibrium.

Mixed Strategy Nash Equilibrium - a set of mixed strategies, one for each player, such that no player has incentive to change his strategy given what the other players are doing.

Dominant Strategy - a strategy that is always better than any other strategy, for any profile of other players' actions.

- Strictly Dominant Strategy - same as Dominant Strategy
- Weakly Dominant Strategy - a strategy that is always better than or equal to any other strategy, for any profile of other players' actions.

Dominant Strategy Nash Equilibrium (equilibrium in dominant strategies) - a Nash equilibrium in which all strategies are strictly dominant. If it exists can be found by elimination of strictly dominated strategies.

Dominated Strategy - a strategy, such that, regardless of what any other players do, the strategy earns a player a smaller payoff than some other strategy.

- Strictly Dominated Strategy - same as Dominated Strategy
- Weakly Dominated Strategy - a strategy, such that, regardless of what any other players do, the strategy earns a player a smaller than or equal to some other strategy payoff.

Subgame Perfect Equilibrium

Bayes-Nash Equilibrium

The Core (analogous to Nash equilibrium for coalitional games)

Nash equilibrium for different types of games:

	Simultaneous	Sequential
Complete	Nash	Subgame Perfect Nash
Incomplete	Bayesian Nash	Perfect Bayesian Nash

Nash's Theorem

Every finite, non-cooperative game of two or more players has a mixed strategy Nash equilibrium. (John Nash, 1950)

Pareto Optimality

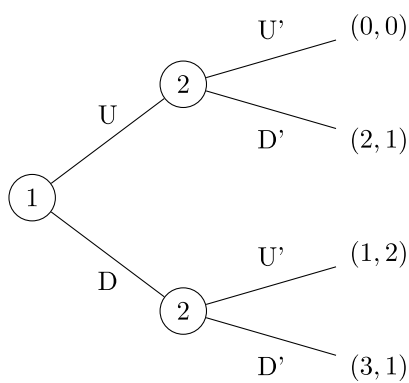
Pareto-optimal Outcome - an outcome, such that there is no other outcome that Pareto-dominates it. An outcome o Pareto-dominates outcome o' if it's at least as good for every player as outcome o' , and there is some player who strictly prefers o to o' .

The Oddness Theorem

Almost all finite games have a finite number of solutions, and that number is also odd. (Robert Wilson, 1971)

Extensive-Form Games

Perfect Information Extensive-form Games



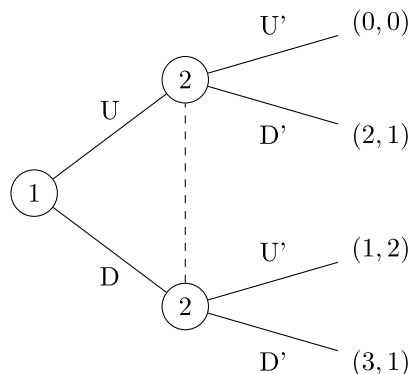
All players know the moves previously made by all other players.

Every game in extensive form can be converted into normal form. The reverse transformation is not always possible, e.g. matching pennies cannot be written as a perfect information extensive form game.

Theorem

Every perfect information game in extensive form has a pure strategy Nash equilibrium.

Imperfect Information Extensive-form Games



Player 2 doesn't know the move made by Player 1.

Backward Induction - identify the equilibria in the bottom-most trees, and adopt these as one moves up the tree.

Subgame Perfect Equilibrium - Nash equilibrium that represents a Nash equilibrium of every subgame in the original game. It's a refinement of the Nash equilibrium that eliminates non-credible threats.

Non-credible Threat - a threat made by a player in an extensive form game which would not be in the best interest for the player to carry out. The hope is that the threat is believed in which case there is no need to carry it out. While Nash equilibria may depend on non-credible threats, Backward Induction eliminates them.

Repeated Games

Repeated Game - an extensive form game that consists of a number of repetitions of some base game, called a stage game. The stage game is usually one of the well-studied 2-person games.

Discount Factor - a number between 0 and 1 that represents the time value of consumption and probability of continuation. A higher discount factor means more patience and higher chance of surviving into the next period.

The One-Shot Deviation Principle - in finite or infinitely repeated games with discounting, a set of strategies is a subgame perfect equilibrium iff no player can profitably deviate from his strategy at a single stage and maintain his strategy everywhere else.

Grim Trigger

- Play some strategy profile with higher payoff than NE
- If anyone deviates then play NE forever.

The strategy profile where everyone plays grim trigger is a subgame perfect equilibrium.

Tit-for-Tat - a strategy in the infinitely repeated prisoner's dilemma:

- Begin by cooperating
- For all future periods, copy opponent's strategy from the previous period.

The strategy profile where everyone plays tit-for-tat is not a subgame perfect equilibrium.

The Folk theorem

- Take a Nash equilibrium from the stage game
- Consider any alternative set of strategies such that the expected utility for those strategies in the stage game is strictly greater for all players
- If discounting factor is sufficiently high, a subgame perfect equilibrium exists in which players use those alternative strategies on the equilibrium path.

Implications of the folk theorem:

- If the purpose of equilibrium is to predict or explain outcomes, subgame perfect equilibrium predicts and explains everything
- Predicting and explaining everything is essentially predicting and explaining nothing

Stochastic Game - a generalization of repeated games.

- agents repeatedly play games from a set of normal-form games.
- the game played at any iteration depends on the previous game played and on the actions taken by all agents in that game.

Learning in Repeated Games:

Fictitious Play: each player maintains explicit belief about the other players.

- Initialize beliefs about the opponent's strategies.
- Each turn:
 - Play best response to the assessed strategy of the opponent
 - Observe the opponent's actual play and update beliefs accordingly.

No-regret Learning:

Bayesian Games

Bayesian Game (Incomplete Information Game) - a game in which players have incomplete information on the other players' strategies and payoffs, but, they have beliefs with known probabilities. It can be modeled as a normal form game with the difference that each player has multiple types with known probabilities (called a common prior beliefs).

Bayes-Nash Equilibrium (Bayesian Nash Equilibrium) - a set of strategies, one for each type of player, such that no type has incentive to change his or her strategy given the beliefs about the types and what the other types are doing.

3 stages of a Bayesian game:

- **Ex-ante** - the player knows nothing about anyone's actual type.
- **Interim** - the player knows her own type but not the types of the other players.
- **Ex-post** - the player knows all players' types. Making choices at this stage for players is equivalent to making choices in complete information game.

Ex-ante Dominated Strategy - a strategy for a player such that an alternative strategy for that player provides a greater payoff for that player regardless of all other players' strategies

Interim Dominated Strategy - a strategy for a type such that an alternative strategy for that type provides a greater payoff for that type regardless of all other players' strategies

Interim dominated strategy implies ex-ante dominated strategies. The reverse is not always true.

The Purification Theorem

Almost all mixed strategy Nash equilibria in a complete information game are the limit of pure strategy Bayesian Nash equilibria in an incomplete information game that converges to the complete information game.

Such games are called Bayesian because of the probabilistic analysis inherent in the game. Players have initial beliefs about the type of each player (where a belief is a probability distribution over the possible types for a player) and can update their beliefs according to Bayes' rule as play takes place in the game, i.e. the belief a player holds about another player's type might change on the basis of the actions they have played.

Bayes' Theorem

$P(A|B) = P(B|A) * P(A) / P(B)$, where

- $P(A|B)$ is a conditional probability: the likelihood of event A occurring given that B is true.
- $P(B|A)$ is also a conditional probability: the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other; this is known as the marginal probability.

Bayes' Rule

Bayes' theorem in odds form is:

$O(A1:A2|B) = O(A1:A2) * \Lambda(A1:A2|B)$, where

$\Lambda(A1:A2|B) = P(B|A1) / P(B|A2)$ - Bayes factor or likelihood ratio, and

$O(A1:A2) = P(A1) / P(A2)$ - odds between to events.

So the rule says that **the posterior odds are the prior odds times the Bayes factor**, or in other words, posterior is proportional to prior times likelihood.

Coalitional Games

Coalitional Game is given by specifying a value for every coalition. Formally, the coalitional game consists of a finite set of players N , called the **grand coalition**, and a characteristic function $v: 2^N \rightarrow \mathbb{R}$ from the set of all possible coalitions of players to a set of payments that satisfies $v(\emptyset)=0$.

Two ways for allocating payoffs:

- Shapley value: based on marginal contributions - what does each player contribute to each possible coalition;
 - "fair" distribution.
- Core: based on coalitional threats - each coalition must get at least what it can generate alone;
 - "stable" distribution.

The Shapley Value allocates the value of a group according to marginal contribution calculations.

$$\phi_i(v) = \frac{1}{|N|!} \sum_R [v(P_i^R \cup \{i\}) - v(P_i^R)]$$

where the sum ranges over all $|N|!$ orders R of the players and P_i^R is the set of players in N which precede i in the order R .

Theorem

For any coalitional game, there is a unique payoff division (the Shapley Value) that divides the full payoff of the grand coalition and that satisfies the 3 axioms

- Symmetry: if players are interchangeable they should be getting the same allocation
- Dummy Player: if the amount that a player contributes to any coalition is 0, then his allocation is 0
- Additivity: if we can separate the game into two subgames such that the value in every coalition is the sum of values in the two subgames, then the allocation in the

game should be equal to the sum of allocations in the two subgames.

The Core - the set of payoff vectors under which no coalition has a value greater than the sum of its members' payoffs. Therefore, no coalition has incentive to leave the grand coalition and receive a larger payoff.

- It's **analogous to Nash equilibrium**, except that it allows deviations by groups of agents.

A game is **simple** if for all coalitions the value of the coalition is either 0 or 1.

A player is a **veto** player if the value of all coalitions that don't involve the player is 0.

Theorem

In a simple game the core is empty iff there is no veto player. If there are veto players, the core consists of all payoff vectors in which the nonveto players get 0.

A game is **convex** if its characteristic function v is supermodular:

$$v(S \cup T) + v(S \cap T) \geq v(S) + v(T), \forall S, T \subseteq N.$$

that is, "the incentives for joining a coalition increase as the coalition grows".

Theorem

Every convex game has a nonempty core.

In every convex game, the Shapley value is in the core.

Social Choice

Social Choice Function - a function that, given a set of linear orderings on the outcomes, tells which outcome should be chosen.

Social Welfare Function - a function that, given a set of linear orderings on the outcomes, tells which ordering should be chosen.

Voting Schemes:

- **Plurality** - pick the outcome which is most preferred by the most people
- **Cumulative voting**
 - distribute e.g., 5 votes each
 - possible to vote for the same outcome multiple times
- **Approval voting** - vote for as many outcomes as you "like" (used e.g. in electing new members into a club or society)
- **Plurality with elimination** ("instant runoff", "transferable voting")
 - if some outcome has a majority, it is the winner
 - otherwise, the outcome with the fewest votes is eliminated (may need some tie-breaking procedure)
 - repeat until there is a winner.
- **Borda Rule, Borda Count**
 - assign each outcome a number.
 - the most preferred outcome gets a score of $n - 1$, the next most preferred gets $n - 2$, down to the n th outcome which gets 0.
 - sum scores for each outcome, and choose one with highest score.
- **Successive elimination**
 - in advance, decide an ordering of alternatives
 - everyone votes for the first or second, and the loser is eliminated.

Condorcet winner - an outcome that is preferred to every other outcome in pairwise majority-rule comparison. It doesn't exist when there is a **Condorcet cycle** e.g. a situation when A defeats B, B defeats C and C defeats A.

Condorcet consistency - if there is a Condorcet winner it must be selected by the social choice function.

Arrow's Impossibility Theorem

Any social welfare function over 3 or more outcomes that is Pareto efficient and independent of irrelevant alternatives is dictatorial. (Kenneth Arrow, 1951)

Social welfare function is **Pareto efficient** if whenever all agents agree on the ordering of two outcomes, the social welfare function selects that ordering.

Social welfare function is **independent of irrelevant alternatives** if the selected ordering between two outcomes depends only on the relative orderings they are given by the agents.

Social welfare function is **dictatorial** if there exists a single agent whose preferences always determine the social ordering.

Muller-Satterthwaite Impossibility Theorem

Any social choice function that is weakly Pareto efficient and monotonic is dictatorial. (definitions are similar to the corresponding terms for social welfare function)

In advance, decide an ordering of outcomes (e.g. according to left-right political spectrum for political parties)

A group of agents is said to have **single-peaked preferences** if:

- Each agent has an ideal outcome in the set; and
- For each agent, outcomes that are further from his ideal outcome are preferred less.

Median voting - the median of the most preferred outcomes is selected.

Theorem

With median voting a condorcet winner always exists if there is an odd number of voters.

Mechanism Design

Mechanism Design (also called Inverse Game Theory) - a field in game theory that focuses on designing the game structure e.g. choosing the actions available to players and mappings of action profiles to outcomes, so as to optimise for certain **qualities** e.g. incentive compatibility, Pareto efficiency, individual rationality etc.

Game Setting - components of the game that we, as game designers don't get to control, e.g. a set of agents, a set of outcomes, common priors etc.

Mechanism - components of the game, which, when added to a corresponding game setting, turn it into a game. An example is a set of available actions for agents and mapping of action profiles to outcomes.

- Mechanism for a Bayesian Game Setting
- Transferable Utility Mechanism

Mechanism for a Bayesian Game Setting

Bayesian Game Setting - a tuple (N, O, Θ, p, u) :

- N - a finite set of n agents
- O - a set of outcomes
- $\Theta = \Theta_1 \times \dots \times \Theta_n$ - a set of possible joint type vectors
- p - a probability distribution on Θ (common prior)
- $u = (u_1, \dots, u_n)$, where $u_i: O \times \Theta \Rightarrow \mathbb{R}$ - utility function for each player i .

Mechanism for a Bayesian Game Setting - is a mechanism where the designer gets to specify the action sets for the agents and the mapping to outcomes, over which agents have utility. Thus it is a pair (A, M) , where

- $A = A_1 \times \dots \times A_n$, where A_i - the set of actions available to agent $i \in N$
- $M: A \Rightarrow \Pi(O)$ maps each action profile to a distribution over outcomes.

Given a Bayesian game setting (N, O, Θ, p, u) , a mechanism (A, M) is an **implementation in dominant strategies** of a social choice function C (over N and O) if for any vector of utility functions u , the game has an **equilibrium in weakly-dominant strategies**, and in any such equilibrium a we have $M(a) = C(u)$

Given a Bayesian game setting (N, O, Θ, p, u) , a mechanism (A, M) is an **implementation in Bayes-Nash equilibrium** of a social choice function C (over N and O) if there exists a **Bayes-Nash equilibrium** of the Bayesian game (N, A, Θ, p, u) , such that for every type profile $\theta \in \Theta$ and every action profile $a \in A$ that can arise given type profile θ in this equilibrium, we have $M(a) = C(u(\cdot, \theta))$

Transferable Utility Mechanism

Agents have **quasilinear preferences with transferable utility** in an n -player Bayesian game when the set of outcomes is:

$$O = X \times \mathbb{R}^n$$

for a set X , if the utility of an agent i given joint type θ can be written:

$$u_i(o, \theta) = u_i(x, \theta) - p_i,$$

where $o = (x, p)$ is an element of O , and $u_i: X \times \Theta \Rightarrow \mathbb{R}$.

The corresponding game setting is called **Quasilinear Setting**.

A **direct mechanism in a quasilinear setting** $(N, O = X \times \mathbb{R}^n, \Theta, p, u)$ is a pair (χ, ρ) specifying a basic outcome $\chi(\theta)$ and a profile of payments $\rho(\theta) = (p_1(\theta), \dots, p_n(\theta))$.

- χ defines the "nonmonetary" outcome
- p_i defines the "monetary" payment (possible negative) that an agent i makes to the mechanism

Private Values

Preferences have **private values**, or satisfy conditional utility independence, if each agent i 's utility function does not depend on the other agents' types, i.e. it can be written as $u_i(o, \theta_i)$.

An agent's type becomes their **valuation function**: i 's value for choice $x \in X$ is $v_i(x) = u_i(x, \theta_i)$.

- v_i is the maximum amount i would be willing to pay to get x

Alternative definition of a **direct mechanism with private values**:

- ask agents to declare valuation functions $v_i: X \Rightarrow \mathbb{R}$

Mechanism Qualities

Direct Mechanism - a mechanism where the set of joint actions is equal to the set of joint types, i.e. the agents have to declare their types to the mechanism.

Incentive Compatible (aka **Truthful** or **Strategy-proof**) **Mechanism** - a direct mechanism where declaring true type for every agent is a weakly-dominant strategy Nash equilibrium. In other words, every agent fare best or at least not worse by being truthful, regardless of what the others do.

- Dominant Strategy Incentive Compatible (DSIC) - same as Incentive Compatible
- Bayesian-Nash Incentive Compatible (BNIC) - a direct mechanism where declaring true type for every agent is a Bayes-Nash equilibrium. In other words, every agent fare best or at least not worse by being truthful, if all the others act truthfully.

A transferable utility mechanism is **strictly Pareto efficient**, or just **efficient**, if in equilibrium it selects the choice that maximizes the sum of agents' utilities, disregarding monetary payments.

A transferable utility mechanism is **budget balanced** when, regardless of the agents' types, in equilibrium the mechanism collects and disburses the same amount of money from and to the agents.

A transferable utility mechanism is **individual rational** when, in equilibrium no agent loses by participating in the mechanism, i.e. the valuation subtracting the payment for every agent is not negative.

- ex interim individual rational - holds for every possible valuation of the individual, but averages over the possible valuations of the other agents
- ex post - holds for every possible valuation of the individual and over the possible valuations of the other agents

A mechanism is **tractable** when for every possible valuation profile the mechanism mapping function can be computed in polynomial time.

A mechanism X is **revenue maximizing** when among the set of other mechanisms that satisfy the other constraints, the mechanism X maximizes the total payments made by agents in equilibrium.

- revenue minimizing is defined analogously

A mechanism is **maxmin fair** if it "makes the least-happy agent the happiest".

Revelation Principle

Any social choice function that can be implemented by any mechanism can be implemented by a truthful, direct mechanism.

In mechanism design, the revelation principle is of utmost importance in finding solutions. The researcher need only look at the set of equilibrium characterized by truthfulness. That is, if the mechanism designer wants to implement some outcome or property, he can restrict his search to mechanisms in which agents are willing to reveal their private information to the mechanism designer that has that outcome or property. If no such direct and truthful mechanism exists, no mechanism can implement this outcome/property. By narrowing the area needed to be searched, the problem of finding a mechanism becomes much easier.

Impossibility of General Dominant-Strategy Implementation (Gibbard-Satterthwaite Theorem)

For every social choice function, one of the following three things must hold:

- The social choice function output is limited to 2 alternatives only
- The social choice function is dictatorial, i.e. there exists a distinguished agent who can choose the winning outcome
- Truthful reporting of preferences is a dominant strategy for every agent.

Median voting in single-peaked domains is strategy-proof (any other statistics can be used instead of median e.g. max or min)

Trade is strategy-proof:

- An agent have a private value for buying (or selling) an indivisible good. He declares this value truthfully.
- Other agents declare whether they are willing to buy or sell at that price.

VCG

The **Vickrey-Clarke-Groves** (VCG) mechanism is a general way for self-interested agents to choose a social-welfare maximizing outcome. It works in quasilinear utility settings, i.e. where monetary payments are applied, although has some limitations (listed below).

Examples of where it can be used

- Privatization - a government privatizing a public utility like a power plant doesn't aim to maximize revenue, but to ensure that the right buyer wins.
- Building a bridge - businesses on both sides of a river need to decide whether to

build a bridge, and if so how to pay for it.

- Scheduling meetings between people who value times differently and may not tell the truth
- Buying a path in a network - shipping along privately owned railroads

Qualities

- truthful - has truth as a dominant strategy
- Pareto efficient - makes efficient choice (not including payments)

Under additional assumptions about the setting, can satisfy:

- Weak budget balance
- Interim individual rationality

History

- Vickrey was the first who defined them in auction settings - 2nd price auctions, Vickrey auctions
- Clarke generalized it to a more general class of settings and defined pivotal mechanisms
- Groves gave a more general class of such mechanisms.

Groves mechanism - a direct transferable utility mechanism such that:

- it selects the outcome that maximizes the total utility for all agents
- every agent:
 - gets paid everyone else's utility under the allocation that is actually chosen, except his own (but he gets that directly as utility)
 - gets charged an amount based on an arbitrary function of the values of the other agents

Vickrey-Clarke-Groves (VCG) mechanism (aka **Pivotal mechanism**) - a Groves mechanism such that:

- it selects the outcome that maximizes the total utility for all agents
- every agent pays his "social cost":
 - gets paid everyone else's utility under the allocation that is actually chosen, except his own (but he gets that directly as utility)

- gets charged everyone's utility in the world where he doesn't participate

In the VCG mechanism:

- agents who don't affect the outcome pay 0
- (pivotal) agents who make things worse for others by existing, pay more than 0
- (pivotal) agents who make things better for others by existing, get paid

Theorem

Truth telling is a dominant strategy under any Groves mechanism including the pivotal mechanism (a VCG mechanism)

Theorem (Green-Laffont)

Suppose that for all agents any utility function is possible. Then a Pareto efficient mechanism has truthful reporting as a dominant strategy for all agents and preferences only if it is Groves mechanism.

VCG Limitations

Privacy

VCG requires agents to fully reveal their private information. This private information may have value to agents that extends beyond the current interaction. For example, the agents may know that they will compete with each other again in the future.

Susceptibility to collusion

Agents can benefit by colluding. For example 2 agents can increase their valuation for an outcome, which will decrease their payment.

VCG is not frugal

The gap between agent's true cost and the payment they could receive under VCG is unbounded. VCG can end up paying arbitrarily more than an agent is willing to accept (or equivalently charging arbitrarily less than an agent is willing to pay).

Revenue Monotonicity Violated

Revenue always weakly increases as agents are added. An agent could pretend to be 2 agents and eliminate his payment. (Sybil attack)

Cannot Return All Revenue to Agents

We might want to find some way of returning the mechanism's profits back the agents. However, the possibility of receiving a rebate after the mechanism has been run changes the agents' incentives.

Theorem

The VCG mechanism is **ex-post individual rational** when the choice set monotonicity and no negative externalities properties hold.

An environment exhibits **choice-set monotonicity** if for all agents the set of outcomes that are achievable without that agent present is a weak subset of the set of outcomes that are possible when that agent is present.

An environment exhibits **no negative externalities** if for all agents and all choices that can be made without that agent, the agent's own valuation for each of these choices is non-negative.

Theorem

The VCG mechanism is weakly budget-balanced when the **no single-agent effect** property holds.

No single-agent effect - if I drop an agent i and then I pick some other choice instead without i , everybody other than i is at least as happy with the new choice as with the old choice.

Theorem (Krishna & Perry, 1998)

In any Bayesian game setting in which VCG is ex post individually rational, VCG collects at least as much revenue as any other efficient and ex interim individually-rational mechanism.

A useful corollary: VCG is as budget balanced as any efficient mechanism can be: it satisfies weak budget balance in every case where any dominant strategy, efficient and ex interim individually-rational mechanism is able to.

Cryptoeconomics

<https://github.com/medvedev1088/cryptoeconomics-cheat-sheet>

Resources

- Game Solver (2x2 matrix games): <http://mindyourdecisions.com/GameSolver.html>
- Game Theory 101: https://www.youtube.com/watch?v=NSVmOC_5zrE&list=PLKI1h_nAkaQoDzI4xDIXzx6U2ergFmedo
- Game Theory: <https://www.coursera.org/learn/game-theory-1>
- Game Theory II: Advanced Applications: <https://www.coursera.org/learn/game-theory-2>
- An Introduction to Decision Theory by Martin Peterson: <https://www.amazon.com/Introduction-Decision-Cambridge-Introductions-Philosophy/dp/0521716543>
- <http://www.gametheory.net/>
- A Crash Course in Mechanism Design for Cryptoeconomic Applications
<https://medium.com/blockchannel/a-crash-course-in-mechanism-design-for-cryptoeconomic-applications-a9f06ab6a976>