Mixed-Nash Equilibrium eg:

	Party	¬Party
Clown	(6,0)	(1,1)
¬Clown	(4,4)	(2,3)

$0p + (1-p)4 = 1p + (1-p)3$ Solve for p. This gives us $(\frac{1}{2}Clown, \frac{1}{2}\neg Clown)$

$6q + (1-q)1 = 4q + (1-q)2$ Solve for q. This gives us $(\frac{1}{3}Party, \frac{2}{3}\neg Party)$

Normal-Form Games: Matrix representation of a game Strategy dominance Iterated dominance Nash equilibrium Pareto Optimality: Optimal from point of view of an outside observer Strictly dominant: $>$ for any strat profile of other players. Weakly dom: $>$ for at least one strat profile. Very weak dom: $\geq$

ORIGAMI: Lower everything. First set highest equal to the utility of the next highest attacher's undefended utility.

Eg:

	t_1	t_2	t_3	t_4
Covered	(1,0)	(3,0)	(8,0)	(8,1)
Uncovered	(-1,1)	(0,2)	(0,4)	(-4,4)

$-1r_4 + (1-r_4)4 = 0r_3 + (1-r_3)4 = 2$

We set 2 here because it is the next lowest, on t_2. If we have resources left over then we include t_2 and do

$-1r_4 + (1-r_4)4 = 0r_3 + (1-r_3)4 = 0r_2 + (1-r_2)2 = 0$

and solve with $r_2 + r_3 + r_4 = n_{resources}$. Set everything in terms of one variable and then plug in here.

Normal Form Game RPS example

	Paper	Rock	Scissor
Paper	(0,0)	(7,-7)	(-1,1)
Rock	(-7,7)	(0,0)	(2,-2)
Scissor	(1,-1)	(-2,2)	(0,0)

We use the form $1 - p_1 - p_2$. So something like $P_2(paper) = p_1 0 + 7p_2 - (1 - p_1 - p_2) = -1 + p_1 + 8p_2$ Follow similar pattern. Then let $P_2(paper) = P_2(Rock) = P_2(Scissor)$ and solve for probabilities.

Extensive Form Game: Tree representation is node is state and branch is action. Final branch is terminating action. Leaf is reward. $G = (N, A, H, Z, \chi, \rho, \sigma, u)$ N =set of players; A =set of actions, H =set of non-terminal choice nodes. Z =set of terminal nodes (disjoint from H $\chi : G \to 2^A$ action function which assigns to each choice node a set of possible actions. $\rho : H \to N$ player function which assigns each choide node a player $i \in N$ who chooses an action at that node. $\sigma : HxA \to H \cup Z$ is the successor function. Maps a choice node and action to new choice node $u = (u_1, \cdots, u_n)$ where $u_i : Z \to \mathbb{R}$ is real valued utility for player i on terminal nodes Z

Player set N: {Entrant, Incumbent}

Action set A: {Enter, Stay out, Accommodate, Fight}

Choice set H: {n_1, n_2}

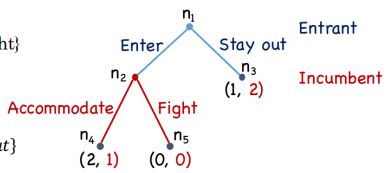
Terminal set Z: {n_3, n_4, n_5}

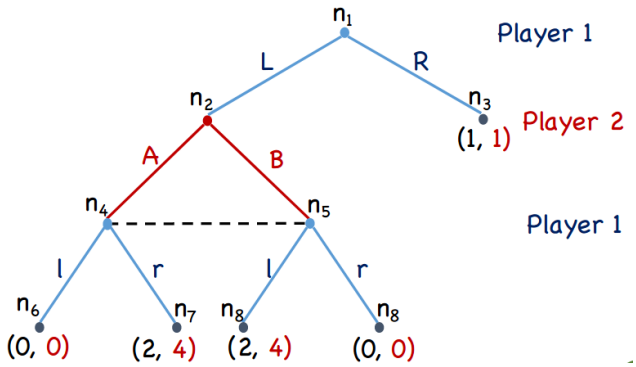
Action function $\chi: H \mapsto 2^A$ $\chi(n_1) = \{Enter, Stay\ out\}$

Player function $\rho: H \mapsto N$ $\rho(n_1) = Entrant$

Successor function $\sigma: H \times A \mapsto H \cup Z$ $\sigma(n_1, Enter) = n_2$

Utility function $u = (u_1, u_2, \ldots, u_n)$ $u_1(n_4) = 2, u_2(n_4) = 1$





Realization plan for player 1

$$r_1() = 1, r_1(L) + r_1(R) = r_1()$$
$$r_1(Ll) + r_1(Lr) = r_1(L)$$
$$r_1(), r_1(L), r_1(R) \geq 0$$
$$r_1(Ll), r_1(Lr) \geq 0$$

Behavioral Strat for player 1

$$\beta_1(L) = r_1(L), \beta_1(R) = r_1(R)$$
$$\beta_1(l) = \frac{r_1(Ll)}{r_1(L)}, \beta_1(r) = \frac{r_1(Lr)}{r_1(L)}$$

Security Games:

Entries: $C_T^K T = \frac{T!}{K!(T-K)}$ K =number of resources, T = number of targets. $K \leq T$ Compact form is size $4T$

Stackelberg Security Games (SSGs): Limited resource allocation, defender commits to a (mixed) strategy, attacker observes and responds

Attacker: $U_{att}^{cov}(t_i) < U_{att}^{uncov}(t_i)$

Defender: $U_{def}^{cov}(t_i) > U_{def}^{uncov}(t_i)$

Strong SSE: Attacker plays a best response against defender's mixed strat. When ties exist, attacker chooses optimal strategy *for the defender*

A pair of strategies $(x, BR_{att}(x))$ forms a SSE IFF

- defender plays a best response $EU_{def}(x, BR_{att}(x)) \geq EU_{def}(x', BR_{att}(x'))\forall x'$
- Attacker plays best response $EU_{att}(x, BR_{att}(x)) \geq EU_{att}(x, t_i)\forall t_i$
- Attacker breaks ties in favor of defender

Theorem: SSE **always** exists

Weak SSG: a pair of strategies $(x, BR_{att}(x))$ forms a weak SE IFF

- defender plays a best response $EU_{def}(x, BR_{att}(x)) \geq EU_{def}(x', BR_{att}(x'))\forall x'$
- Attacker plays best response $EU_{att}(x, BR_{att}(x)) \geq EU_{att}(x, t_i)\forall t_i$
- Attacker does **not** break ties in favor of defender. Attacker chooses strategy which is **worst** for defender

Theorem: Weak SSE does **not** always exist.

Note: Nash strategies ARE NOT Stackelberg strategies. In zero-sum security games Maximin, Minimax, Nash, and Stackelberg strats are equivalent.

Behavioral Game Theory:

BRASS: Bounded rationality (ϵ). Think like ORIGAMI but within an ϵ bound. $EU_{att}(t_i) \geq EU_{att}(t_j) + \epsilon$

Assumption: Attacker perceives any ϵ-difference in utility and plays utility maximizing strategy.

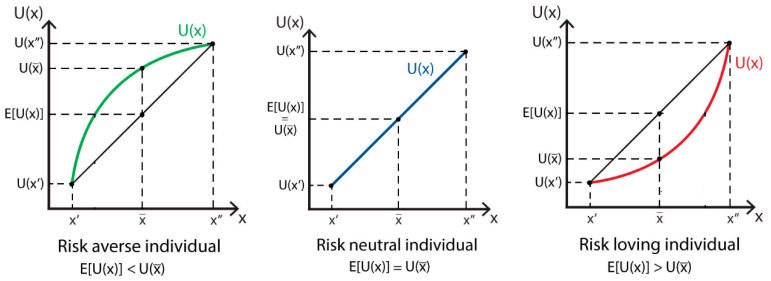
$$\max_{v,x,z,q} v_d ef$$
$$s.t. z_t \in \{0,1\}\forall t$$
$$\Sigma z_t = 1$$
$$q_t \in \{0,1\}\forall t$$
$$\Sigma q_t \geq 1$$
$$x_t \in [0,1]\forall t$$
$$\Sigma x_t \leq K$$
$$0 \leq v_{att} - U_{att}(t, x_t) \leq (1 - z_t)M, \forall t$$
$$\epsilon(1 - q_t) \leq v_{att} - U_{att}(t, x_t) \leq \epsilon + (1 - q_t)M, \forall t$$
$$v_{def} - U_{def}(t, x_t) \leq (1 - q_t)M, \forall t$$

GUARD/COBRA: observation uncertainty (anchoring bias) (Continuation from BRASS) No observation may lead to **anchoring bias** on uniform distribution. Optimize against attacker's belief $x'_t = \alpha\frac{1}{N} + (1-\alpha)x_t$ (N =#targets, $\alpha \in [0,1]$ decreases with # observations)

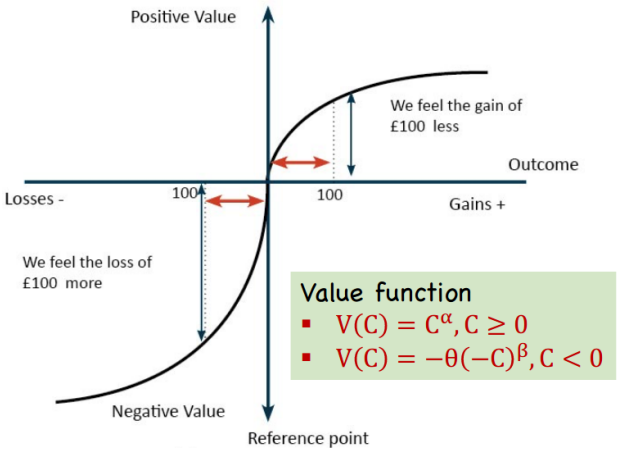
$$\max_{v,x,z} v_{def}$$
$$s.t. z_t \in \{0,1\} \forall t$$
$$\Sigma_t z_t = 1$$
$$x_t \in [0,1], \forall t$$
$$\Sigma_t x_t \leq K$$
$$0 \leq v_{att} - U_{att}(t, x'_t) \leq (1-z_t)M, \forall t$$
$$x'_t = \alpha\frac{1}{N} + (1-\alpha)x_t, \forall t$$
$$v_{def} - U_{def}(t, x_t) \leq (1-z_t)M, \forall t$$

Quantal Response: λ $q(t_t, x) = \frac{e^{\lambda U_{att}(t_i,x)}}{\Sigma_j e^{\lambda U_{att}(t_j,x)}}$ λ is the error level (rationality of attacker; $\lambda = 0$: attacks targets uniformly; $\lambda =$ inf: perfectly rational)

Subjective Utility: Linear comb of features and a weight for the attacker's preference. $SU_{att}(t_i, \vec{x}) = \Sigma w_i f_i(t_i)$ *replace Quantal Response s/U/SU/g,* $\lambda = 1$

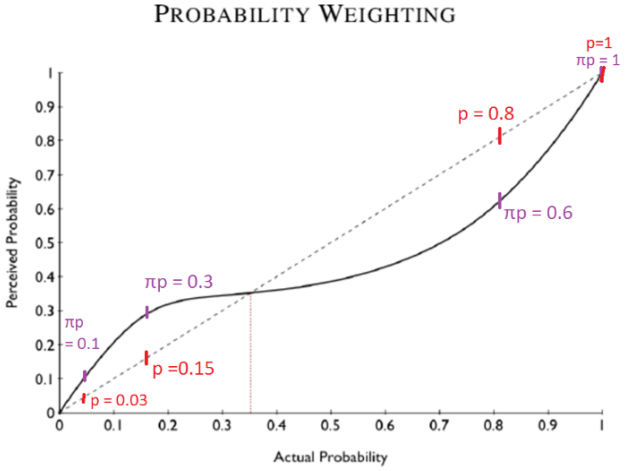


Prospect Theory: How to decide btwn alternatives that involve risks and uncertainty. Value function: values of outcome. Weight function: perceived probabilities.



Risk aversion: *convexity.*

Endowment effect: value things we own more highly. Loss is strong.



$\pi(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{\frac{1}{\gamma}}}$
Underweight of high probability. Overweight of small probability.

Expected Utility:
$U_{att}(t_i, \vec{x} = U_{att}^{uncov}(1 - x_{t_i}) + U_{att}^{cov} x_{t_i}$

Prospect Utility:
$U_{att}^{PT}(t_i, \vec{x}) = V(U_{att}^{uncov})\pi(1 - x_{t_i}) + V(U_{att}^{cov})\pi(x_{t_i})$

Multi-Agent Learning:
$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha(r_t + \gamma \max_{a'} Q(s_{t+1}, a')$
Eg: where $\alpha = \frac{1}{2}$, $\gamma = \frac{1}{2}$

s_t	a_t	s_{t+1}	r_t
A	D	B	2
B	D	C	3

$Q(A, D) = (1-\frac{1}{2})0 + \frac{1}{2}(2 + \frac{1}{2}\max_{a' \in \{D,U\}} Q(B, a')) = 1.$
$Q(B, D) = (1-\frac{1}{2})0 + \frac{1}{2}(3 + \frac{1}{2}\max_{a' \in \{D,U\}} Q(C, a')) = 1.5$
$Q(s_t, a_t) = 0 \because$ we have no stored value in that element of the matrix (initialized to 0). $2/3 \because$ that is the reward in the table. In both cases the $\max_{a'} = 0 \because$ again there were no values in the states (B then C) and so the best action we could take would give us 0 reward.

Nash Equilibrium: Players do not want to deviate from their actions even when the other player changes their strategy. *NE is actions not a cell*

Using *Transition and Reward* functions $\hat{T}(s, a, s')$ and $\hat{R}(s, a, s')$ $\hat{T}$ we look up chance of going to state s' from iterated $Q(s_t, a_t)$. $\hat{R}$ we just lookup on the table.

Optimal Policy: $\hat{V}^{\pi_0}(s) = \hat{T}(s, \pi s, s')(\hat{R}(s, a, s') + \gamma \hat{V}^{\pi_0}(s)$
Eg:

a	a	s'	$\hat{T}$	$\hat{R}$
A	U	A	1	12
A	D	B	$\frac{1}{2}$	2
B	U	B	1	8
B	D	C	1	-6
C	D	C	1	12

$\hat{V}^{\pi_0}(B) = \hat{T}(B, \pi_0(B), B)(8 + \frac{1}{2}V^{\pi_0}(B) = 8 + \frac{1}{2}V^{\pi_0}(B)$
$= 8\frac{1}{1-\frac{1}{2}} = 16$ $(1 + x + x^2 \cdots = \frac{1}{1-x})$

Exploration: $f(u, n) = u + \frac{k}{n}$ where u =utility, n =#visits, k =some const. Replace when calculating values (this offsets). Random exploration has higher regret than exploration function.

MDP: given a present state, the future and the past are independent.

Value Iteration
$V_{k+1}(s) \leftarrow \max_a \Sigma_{s'} T(s, a, s')[R(s, a, s') + \gamma V_k(s')]$ Complexity $O(S^2 A)$

Policy Iteration:
Evaluation $V_{k+1}^{\pi_i}(s) \leftarrow \Sigma_{s'} T(s, \pi_i(s), s')[R(s, \pi_i(s), s') + \gamma V_k^{\pi_i}(s')]$
Improvement with one step lookahead