Homo Ex Machina

The Man from the Machine

Pedro Ferreira

Intelligence

Biological Intelligence: Biopsychological potential to solve problems.

- **Analytical Intelligence:**
  - Abstraction - Create complex world representations.
  - Learning - Inform future decisions based on past experience.
- **Emotional Intelligence:**
  - Self-awareness - Understanding your mental states.
  - Self-management - Managing one's mental states.
- **Social Intelligence:**
  - Social awareness - Awareness of others' mental states.
  - Relationship Management - Ability to change others' mental states.
- **Creative Intelligence**:
  - Going beyond what is given to generate novel and interesting ideas.
  - Defining intelligence trait of humans, compared to other animals.

e.g. solving a hard equation

e.g. dealing with stress of solving a hard equation

e.g. dealing with stressed people who are solving hard equations

e.g. simplify by changing to spherical coordinates

Artificial Intelligence: "(...) A machine [that behaves] in ways that would be called intelligent if a human were so behaving."

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century.* New York, NY: Basic Books.

McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.

Artificial Intelligence

Boring if-else statements

Machine Learning

This Room

	Expert Systems	Analytical AI	Human-inspired AI	Humanized AI	Homo Sapiens
Analytical Intelligence	X	✓	✓	✓	✓
Emotional Intelligence	X	X	✓	✓	✓
Social Intelligence	X	X	X	✓	✓
Creative Intelligence	X	X	X	X	✓

Supervised Learning
Self-Supervised Learning
Reinforcement Learning

My Work

- Risk Assessment
- Theory of Mind

Kaplan, A., & Haenlein, M. (2018). *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*. Business Horizons.

Why humanized AI matters

	Analytical AI	Human-inspired AI	Humanized AI
Universities	Virtual teaching assistants able to answer student questions and tailor reactions to individual data.	AI-based career services able to identify emotions to improve interview techniques of students.	Robo-teachers animating a student group by acting as moderator and sparring partners.
Corporations	Robo-advisors leveraging automation and AI algorithms to manage client portfolios.	Stores identifying unhappy shoppers via facial recognition at checkouts to trigger remedial actions.	Virtual agents dealing with customer complaints and addressing concerns of unhappy customers.
Governments	Automation systems to set the brightness of streetlights based on traffic and pedestrian movements.	Virtual army recruiters interviewing and selecting candidates based on emotional cues.	AI systems able to psychologically train soldiers before entering a war zone.

Increasingly complex AI can replace humans in more boring/dangerous/time-consuming tasks.

Kaplan, A., & Haenlein, M. (2018). *Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence*. Business Horizons.

Risk Assessment

Option 1

- 1M€ with 100%

Option 2

- 2M€ with 50%
- 0€ with 50%

Risk: The possibility of losing something of value.

Risk Assessment

Expected Utility Theory (EUT)

A theory about deciding **optimally**.

Cumulative Prospect Theory (CPT)

A theory about deciding under **risk.**

- Most **AI agents** are built to perform optimally and so **use EUT**.
- Experiments indicate **people use CPT rather than EUT.**

Theory of Mind

Ability to attribute **mental states** to others and to **realize they may be different** from our own.

Mental States

Beliefs
Goals
Emotions
Knowledge

Why "**theory**"?

- Mental states are **not directly observable.**
- Makes **predictions** about the behavior of other agents.

Enables complex social behavior:

- Common sense ideas about others
- Taking perspectives
- Presuming intent
- Inferring emotions

Level-K Model

A recursive theory of mind model.
"I think that you think that I think that..."

D. Premack and G. Woodruff. *Does the chimpanzee have a theory of mind?* Behavioral and Brain Sciences, 1978.

Motivation

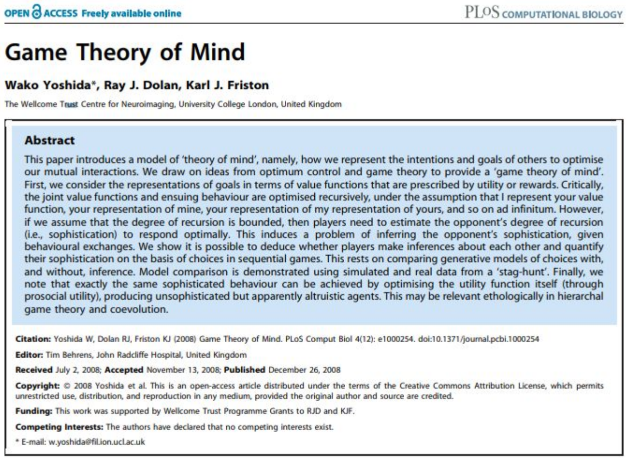
- AI models are being developed to decide optimally, using **expected utility theory** (EUT).
- People **do not seem to decide optimally using EUT**.
- People predict behavior of others with a cognitive mechanism called **theory of mind** (ToM).

We should create AI models that decide like people.

- Create agents that use **cumulative prospect theory** instead for expected utility theory
- Create agents that use **Level-K model**, a recursive model of theory of mind.

The Question

Does coordination happen among risk-sensitive agents equipped with ToM?



OPEN ACCESS Freely available online PLoS COMPUTATIONAL BIOLOGY

Game Theory of Mind

Wako Yoshida*, Ray J. Dolan, Karl J. Friston

The Wellcome Trust Centre for Neuroimaging, University College London, United Kingdom

Abstract

This paper introduces a model of 'theory of mind', namely, how we represent the intentions and goals of others to optimise our mutual interactions. We draw on ideas from optimum control and game theory to provide a 'game theory of mind'. First, we consider the representations of goals in terms of value functions that are prescribed by utility or rewards. Critically, the joint value functions and ensuing behaviour are optimised recursively, under the assumption that I represent your value function, your representation of mine, your representation of my representation of yours, and so on ad infinitum. However, if we assume that the degree of recursion is bounded, then players need to estimate the opponent's degree of recursion (i.e., sophistication) to respond optimally. This induces a problem of inferring the opponent's sophistication, given behavioural exchanges. We show it is possible to deduce whether players make inferences about each other and quantify their sophistication on the basis of choices in sequential games. This rests on comparing generative models of choices with, and without, inference. Model comparison is demonstrated using simulated and real data from a 'stag-hunt'. Finally, we note that exactly the same sophisticated behaviour can be achieved by optimising the utility function itself (through prosocial utility), producing unsophisticated but apparently altruistic agents. This may be relevant ethologically in hierarchal game theory and coevolution.

Citation: Yoshida W, Dolan RJ, Friston KJ (2008) Game Theory of Mind. PLoS Comput Biol 4(12): e1000254. doi:10.1371/journal.pcbi.1000254

Editor: Tim Behrens, John Radcliffe Hospital, United Kingdom

Received July 2, 2008; **Accepted** November 13, 2008; **Published** December 26, 2008

Funding: This work was supported by Wellcome Trust Programme Grants to RJD and KJF.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: w.yoshida@fil.ion.ucl.ac.uk

Yoshida et. al. shows **promoted coordination** among **risk-insensitive (EUT)** agents **equipped with ToM.**

W. Yoshida, R. J. Dolan, and K. J. Friston. *Game theory of mind.* PLoS computational biology 4.12 (2008): e1000254.

Roadmap

Game Theory and Coordination
Expected Utility Theory
Cumulative Prospect Theory
Example - Stag Hunt

Coordination over time
Markov Games
Level-K Theory of Mind
Example - Grid Stag Hunt

Conclusion and Future Work

Game Theory

The study of strategic reasoning between rational decision-makers.

A game is a metaphor for conflict between agents.

Game = Agents + Actions + Information + Rewards

Solving a game usually means finding the **Nash Equilibrium**:
The set of strategies from which no agent would be better off by unilaterally switching.

Applications

Evolutionary models, board games, mechanism design, voting systems, war, public choice, social dilemmas, climate change, animal territorial distribution, multi-agent systems, bargaining, social network formation, disaster relief, and many more...

Coordination and Stag Hunt

Coordination game is a game with multiple deterministic Nash equilibria in which players choose the same or corresponding strategies.

Stag Hunt

The paradigmatic example of a coordination dilemma.

1 \ 2	Stag	Hare
Stag	5,5	0,1
Hare	1,0	1,1

Coordination and Stag Hunt

Coordination game is a game with multiple deterministic Nash equilibria in which players choose the same or corresponding strategies.

Stag Hunt
The paradigmatic example of a coordination dilemma.

If decisions are **deterministic**:

(Stag,Stag) and (Hare,Hare) are **deterministic NEs.**

Under EUT: $V^{EUT}(a) = \mathbb{E}_{a'}[u(a,a')]$ $\quad Pr(Stag) = p$

If decisions are **stochastic**:

$$V^{EUT}(Stag) = 5p + 0(1-p) = 5p$$
$$V^{EUT}(Hare) = 1p + 1(1-p) = 1$$
$$p = \frac{1}{5}$$

1 \ 2	Stag	Hare
Stag	5,5	0,1
Hare	1,0	1,1

[(Stag,⅕;Hare,⅘),(Stag,⅕;Hare,⅘)]
is the **stochastic NE.**

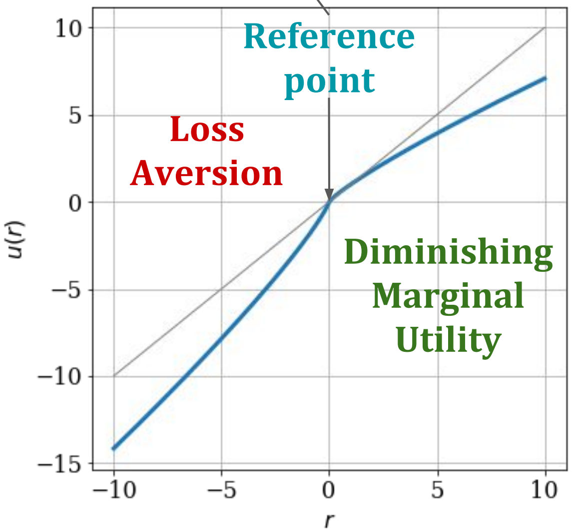
Prospect Theory

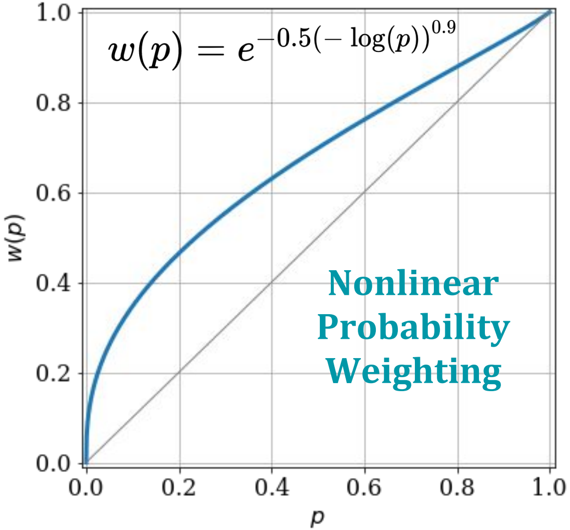
$$R = \text{"reward"} \qquad \text{support}\{R\} = \{r_{-m}, \ldots, r_n\} \qquad P(R = r_i) = p_i$$

this is a prospect

$$\underbrace{([r_{-m}, p_{-m}], \ldots, [r_k, p_k]}_{\text{Losses}}, \underbrace{[r_{k+1}, p_{k+1}], \ldots, [r_n, p_n])}_{\text{Gains}}$$

$$V^{PT} = \sum_{i=k+1}^{n} u^+(r_i)w(p_i) + \sum_{i=-m}^{k} u^-(r_i)w(p_i)$$





Example - A roll of a die

$$\underbrace{([1,1/6],[2,1/6],[3,1/6],[4,1/6],[5,1/6],[6,1/6])}_{\textbf{Gains}}$$

$$V^{PT} = \sum_{k=1}^{6} u(k) w\left(\frac{1}{6}\right) = w\left(\frac{1}{6}\right)(u(1)+u(2)+u(3)+u(4)+u(5)+u(6))$$

$$u(x) = x^{0.85}$$
$$w(x) = e^{-0.5(-log(x))^{0.9}}$$

$$V^{PT} \approx 7.35 > 6$$

A gift of 7 would be refused if the opportunity cost was not being able to accept this gamble.

Does not satisfy first-order stochastic dominance.
Cumulative Prospect Theory solves this.

Cumulative Prospect Theory

$$\underbrace{([r_{-m}, p_{-m}], \dots, [r_k, p_k]}_{\text{Losses}}, \uparrow \underbrace{[r_{k+1}, p_{k+1}], \dots, [r_n, p_n])}_{\text{Gains}}$$

Losses — Reference point — Gains

$$V^{PT} = \sum_{i=k+1}^{n} u^+(r_i)w(p_i) + \sum_{i=-m}^{k} u^-(r_i)w(p_i)$$

$\longrightarrow$

$$V^{CPT} = \sum_{i=k+1}^{n} u^+(r_i)[w^+(P(R \geq r_i)) - w^+(P(R > r_i))] + \sum_{i=-m}^{k} u^-(r_i)[w^-(P(R \geq r_i)) - w^-(P(R > r_i))]$$

Same as PT but transforms cumulative probabilities instead.

Coordination with CPT

$$V^{CPT}(a) = \sum_i u^+(r_i)[w^+(P(R(a) \geq r_i)) - w^+(P(R(a) > r_i))]$$
$$+ \sum_{i=-m}^{k} u^-(r_i)[w^-(P(R(a) \geq r_i)) - w^-(P(R(a) > r_i))]$$

		2: Stag	2: Hare
1	Stag	5,5	0,1
1	Hare	1,0	1,1

$$V^{CPT}(Stag) = 5w^+(p)$$
$$V^{CPT}(Hare) = 1$$

$p_{CPT} = 0.028$ $\quad p_{EUT} = 0.2$

Coordination increases

$$\text{Total Reward}(p) = \mathbb{E}[r_1(p_1, p_2) + r_2(p_1, p_2) | p_1 = p_2 = p]$$
$$= (5+5)p^2 + 2p(1-p) + 2(1-p)^2$$

$$\text{Total Reward}(p_{EUT}) = 2$$
$$\text{Total Reward}(p_{CPT}) \approx 1.95$$

Total reward decreases only slightly

Roadmap

Game Theory and Coordination
Expected Utility Theory
Cumulative Prospect Theory
Example - Stag Hunt

Coordination over time
Markov Games
Level-K Theory of Mind
Example - Grid Stag Hunt

Conclusion and Future Work

Sequential Decision-making

Normal-Form Game = Agents + Actions + Rewards + Single Simultaneous Decision

Many real-world **scenarios are stochastic** in nature and require **time** to be taken into consideration.

Markov Game = **Markov Decision Process** + Agents

Markov Decision Process = **Markov Chain** + Actions + Rewards

Adds control

Adds motivation

Markov Chain = Model of discrete-time stochastic environment.

Markov Chain

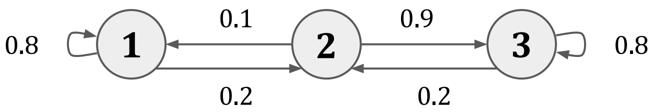
Markov Chain: Sequence of random variables $\{S_t\}_t$ with the first-order Markov property.

First-order Markov property: Future depends only on the present.

$$\mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t, S_{t-1} = s_{t-1}, \ldots, S_0 = s_0) = \mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t)$$

State space

$$S = \{1, 2, 3\}$$



Transition Probability Matrix

$$[P]_{s_t, s_{t+1}} = \mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t)$$

$$P = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0 & 0.9 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

Transition Step

$$\rho_{t+1} = \rho_t P$$

$$\rho_t = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$$\rho_{t+1} = \begin{bmatrix} 0.1 & 0 & 0.9 \end{bmatrix}$$
$$\rho_{t+2} = \begin{bmatrix} 0.08 & 0.2 & 0.72 \end{bmatrix}$$
$$\vdots$$

Stationary distribution

$$\rho_\infty = \rho_\infty P$$

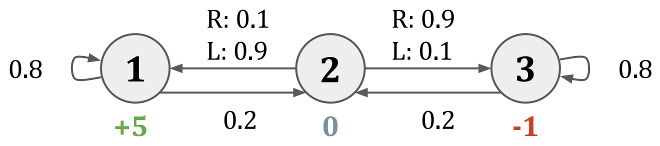
$$\rho_\infty = \left[\frac{1}{12}, \frac{2}{12}, \frac{9}{12} \right]$$

Markov Decision Process

Markov Decision Process = Markov Chain + Actions + Rewards

State space

$$S = \{1, 2, 3\}$$



Action space

$$A = \{L, R\}$$

Transition Probability Matrices

$$[P^a]_{s_t, s_{t+1}} = \mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t, a_t = a)$$

$$P^L = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.9 & 0 & 0.1 \\ 0 & 0.2 & 0.8 \end{bmatrix} \quad P^R = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.1 & 0 & 0.9 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

Given a policy, MDP is a MC

$$\mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t, \pi) = \sum_{a \in A} \mathbb{P}(S_{t+1} = s_{t+1} | S_t = s_t, a_t = a)\pi(a|s_t)$$

Policy

$\pi(a|s) =$ Probability of choosing action a, when in state s.

Problem

Find a policy that maximizes a value functional $V(s, \pi)$

EUT: $V^{\text{EUT}}(s, \pi) = \mathbb{E}_{s_{t+1} \sim p(\cdot | s_t, \pi(s_t))} \left[\sum_{t=0}^{\infty} \beta^t r(s_t, \pi(s_t)) \middle| s_0 = s \right]$

CPT: **heavy math ahoy!**

Markov Decision Process with CPT-Value

$$V^{\text{CPT}}(s,\pi) = \int_0^{\infty} w^+ \left(\sum_{a \in A(s)} P_s^a \left(u^+((r(s) + \beta V^{\text{CPT}}(S,\pi) - b)_+) > \epsilon\right) \pi(a|s) \right) d\epsilon$$
$$- \int_0^{\infty} w^- \left(\sum_{a \in A(s)} P_s^a \left(u^-((r(s) + \beta V^{\text{CPT}}(S,\pi) - b)_-) > \epsilon\right) \pi(a|s) \right) d\epsilon$$

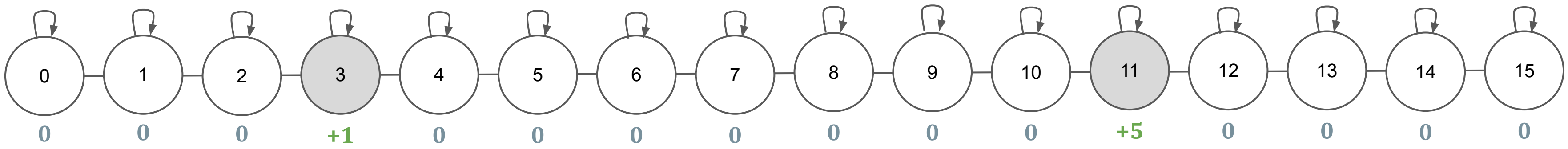
$$(\cdot)_+ = \max(0, \cdot)$$
$$(\cdot)_- = -\min(0, \cdot)$$

$$V^{\text{CPT}*}(s) = \max_{\pi} V^{\text{CPT}}(s,\pi), \text{ for all s.}$$

A. Ruszczyński. *Risk-averse dynamic programming for markov decision processes*. Mathematical Programming, 125 (2010), pp. 235–261.
K. Lin. *Stochastic Systems with Cumulative Prospect Theory,* PhD Thesis, 2013.

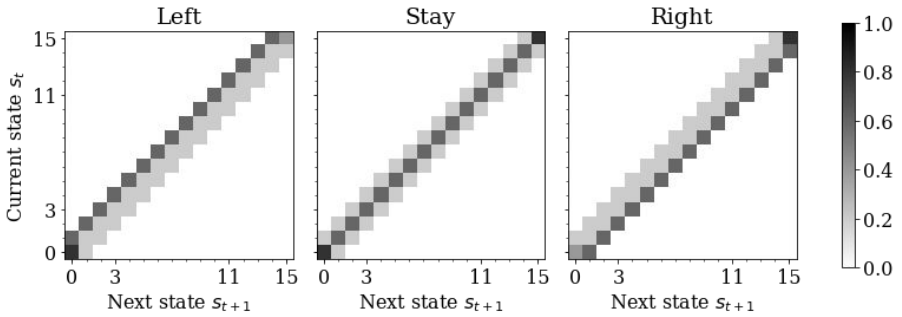
Example

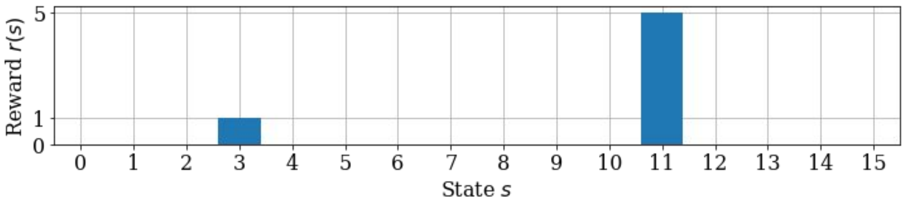
One hunter | Prey: **hares** (3) and **pigs** (11). Pigs are **fatter** than hares and **as easy to catch**.



$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}$$

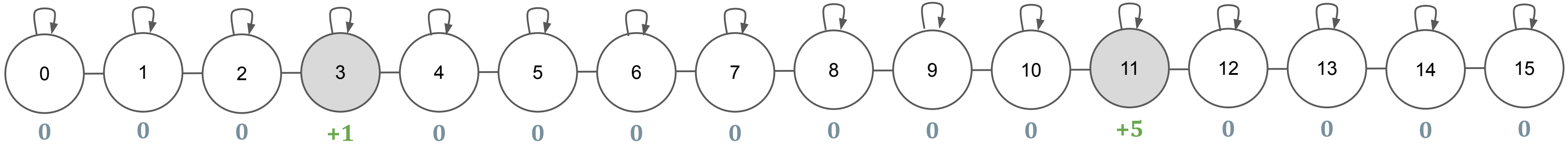
$$A = \{Left, Stay, Right\}$$

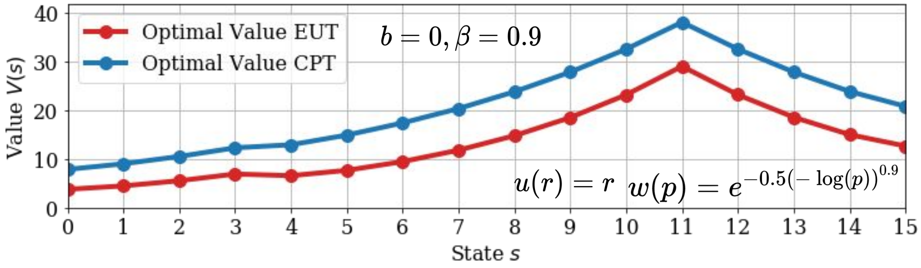


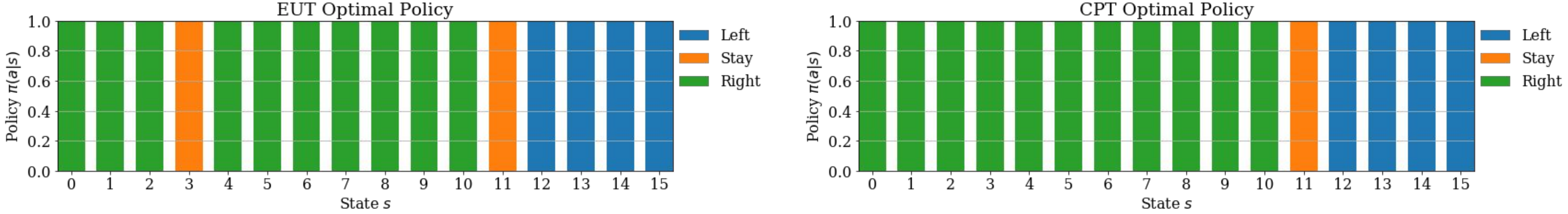


Example

One hunter | Prey: **hares** (3) and **pigs** (11). Pigs are **fatter** than hares and **as easy to catch**.







Markov Game with CPT-Value

Markov Game = Markov Decision Process + Agents

$$P_{i,s,+}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon)$$

$$V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(s) = \int_0^\infty w_i^+ \left(\sum_{a_i \in A_i(s)} \sum_{\boldsymbol{a}_{-i} \in A_{-i}(s)} P_s^{a_i,\boldsymbol{a}_{-i}} \left(u_i^+ \left((r_i(s) + \beta_i V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(S) - b_i)_+ \right) > \epsilon \right) \boldsymbol{\pi}_{-i}(\boldsymbol{a}_{-i}|s) \pi_i(a_i|s) \right) d\epsilon$$
$$- \int_0^\infty w_i^- \left(\sum_{a_i \in A_i(s)} \sum_{\boldsymbol{a}_{-i} \in A_{-i}(s)} P_s^{a_i,\boldsymbol{a}_{-i}} \left(u_i^- \left((r_i(s) + \beta_i V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(s') - b_i)_- \right) > \epsilon \right) \boldsymbol{\pi}_{-i}(\boldsymbol{a}_{-i}|s) \pi_i(a_i|s) \right) d\epsilon$$

$$P_{i,s,-}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon)$$

$$V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(s) = \int_0^\infty w_i^+ \left(\sum_{a_i \in A_i(s)} P_{i,s,+}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon) \pi_i(a_i|s) \right) d\epsilon - \int_0^\infty w_i^- \left(\sum_{a_i \in A_i(s)} P_{i,s,-}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon) \pi_i(a_i|s) \right) d\epsilon$$

Same as single agent,
given the joint policy of other agents.

Use Level-K model to get policies of other agents.

Level-K model

$$V_1^{\pi_1,\pi_2}(s) = \int_0^\infty w_1^+ \left(\sum_{a_1 \in A_1(s)} P_{1,s,+}^{a_1,\pi_2}(\epsilon)\pi_1(a_1|s) \right) d\epsilon - \int_0^\infty w_1^- \left(\sum_{a_i \in A_1(s)} P_{1,s,-}^{a_1,\pi_2}(\epsilon)\pi_1(a_1|s) \right) d\epsilon$$

2 agent scenario

Agent 1 assumes stereotype policy $\pi_2^{(0)}$

Agent 2 assumes stereotype policy $\pi_1^{(0)}$

$$\pi_2^{(0)} \qquad \pi_1^{(0)}$$

$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(0)}}(s) = \pi_1^{(1)}(s) \qquad \pi_2^{(1)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(0)},\pi_2}(s)$$

$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(1)}}(s) = \pi_1^{(2)}(s) \qquad \pi_2^{(2)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(1)},\pi_2}(s)$$

$$\vdots \qquad \vdots$$

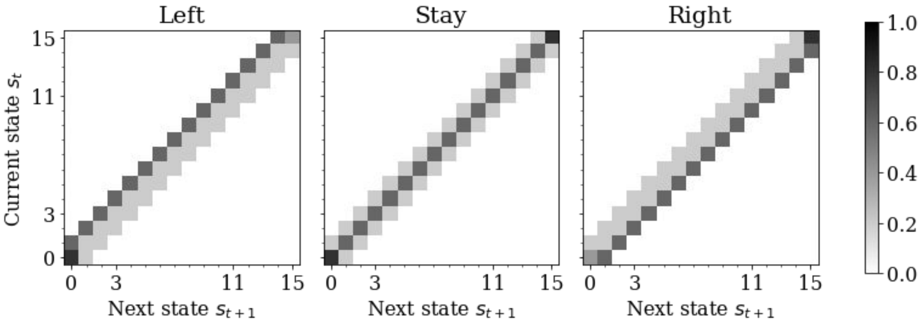
$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(k-1)}}(s) = \pi_1^{(k)}(s) \qquad \pi_2^{(k)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(k-1)},\pi_2}(s)$$

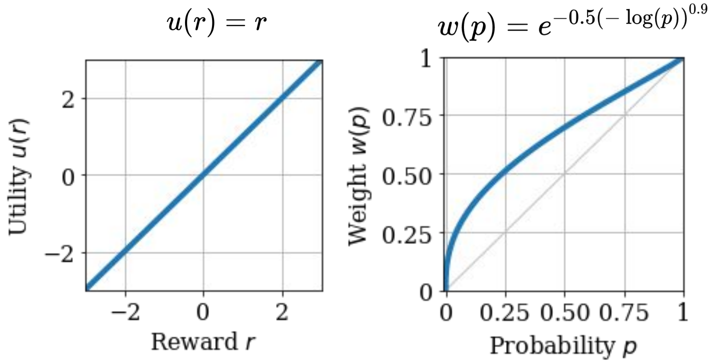
Recap

- Game Theory
  - CPT increases coordination in the Stag Hunt game.
- Decision over time
  - Markov Chain: Describes discrete-time stochastic environments.
  - Markov Decision Process: **One** agent finds a policy that maximizes value.
    - EUT and CPT are different. CPT allowed to escape local maximum.
  - Markov Game: **Multiple** agents find corresponding policies that maximize corresponding values.
    - Level-K allows agents to assume behavior and find increasingly sophisticated policies.

How does coordination change in a Markov game where agents use CPT and ToM?

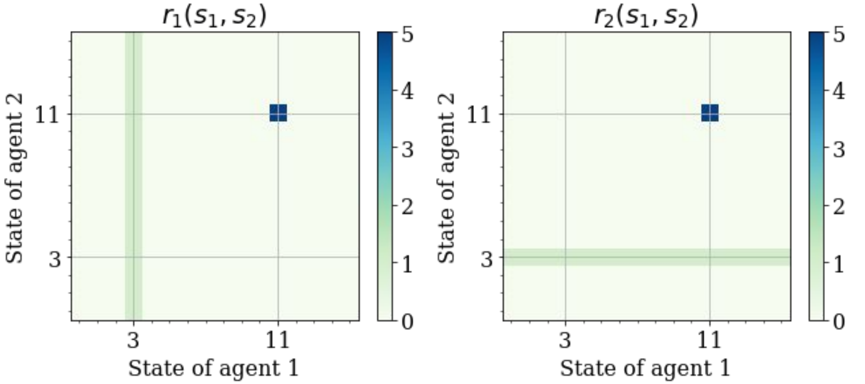
Example - Grid Stag Hunt





Actions are simultaneous and agents can not change state of others.

$$P^{a_1,a_2} = \frac{I \otimes P_1^{a_1} + P_2^{a_2} \otimes I}{2}$$

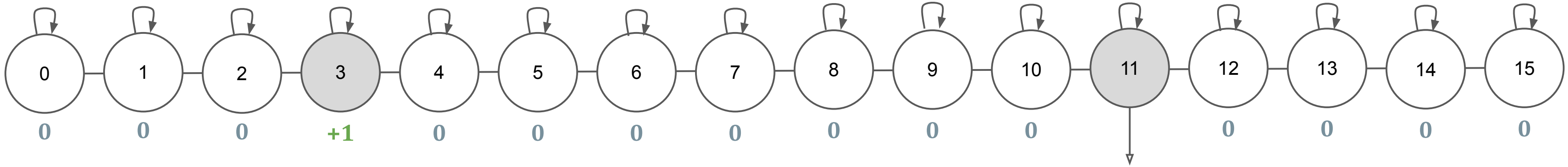


Stereotype policies are assumed uniform for both agents.

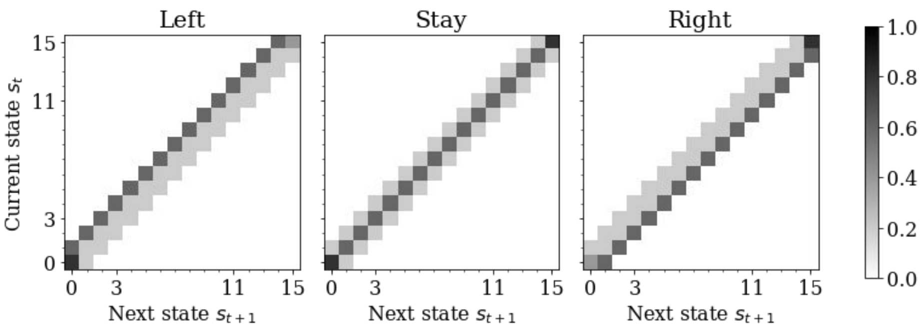
$$b_1 = b_2 = 0, \beta_1 = \beta_2 = 0.9$$

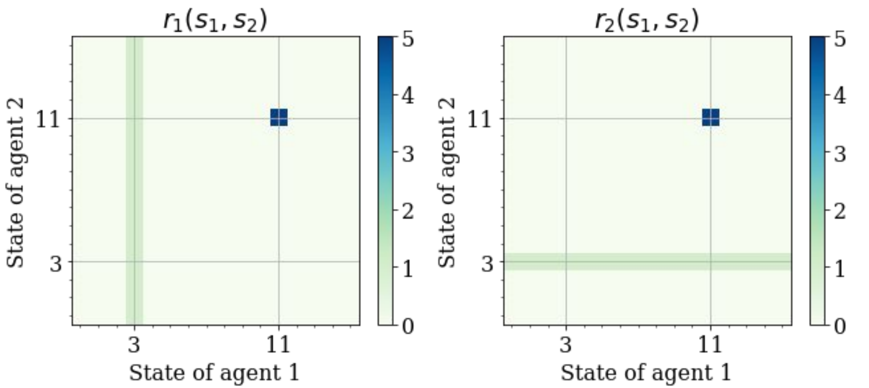
Example - Grid Stag Hunt

Two hunters | Prey: **hares** (3) and **stags**(11). Stags are better but require **coordination**.

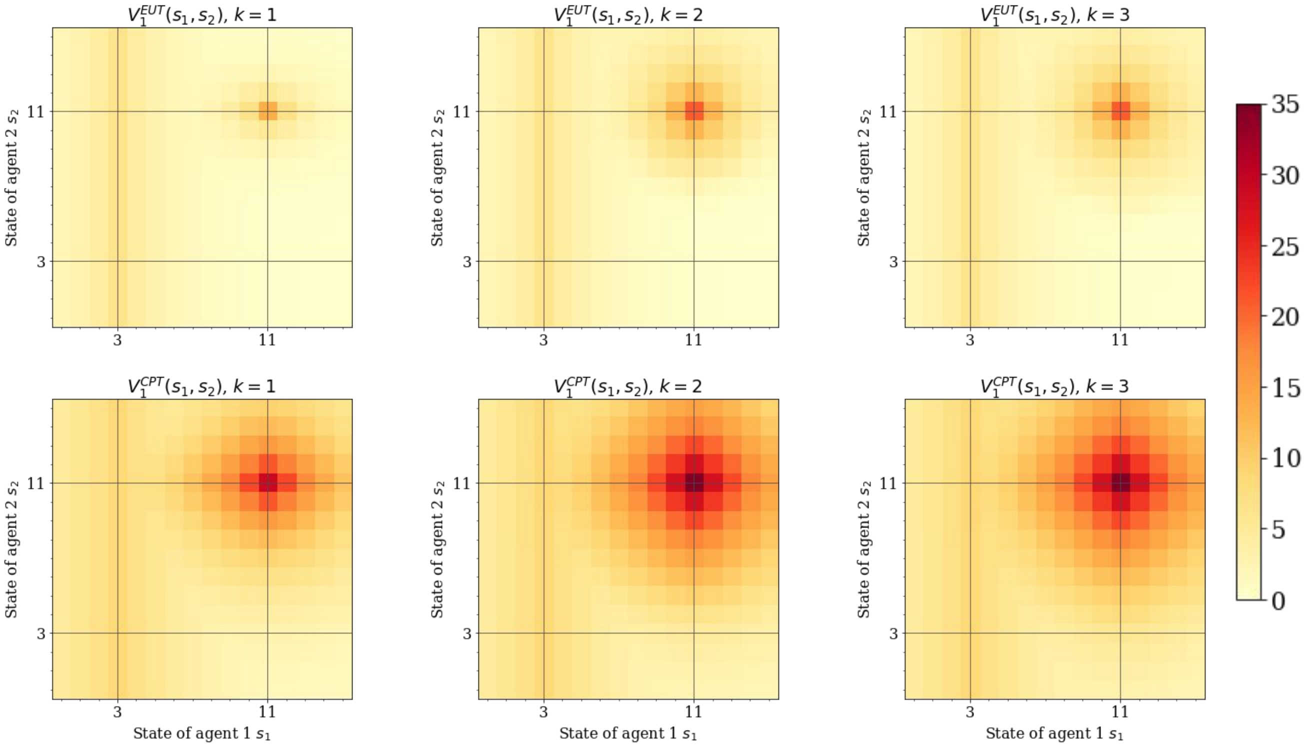


+5 if **both** agents are here,
0 otherwise



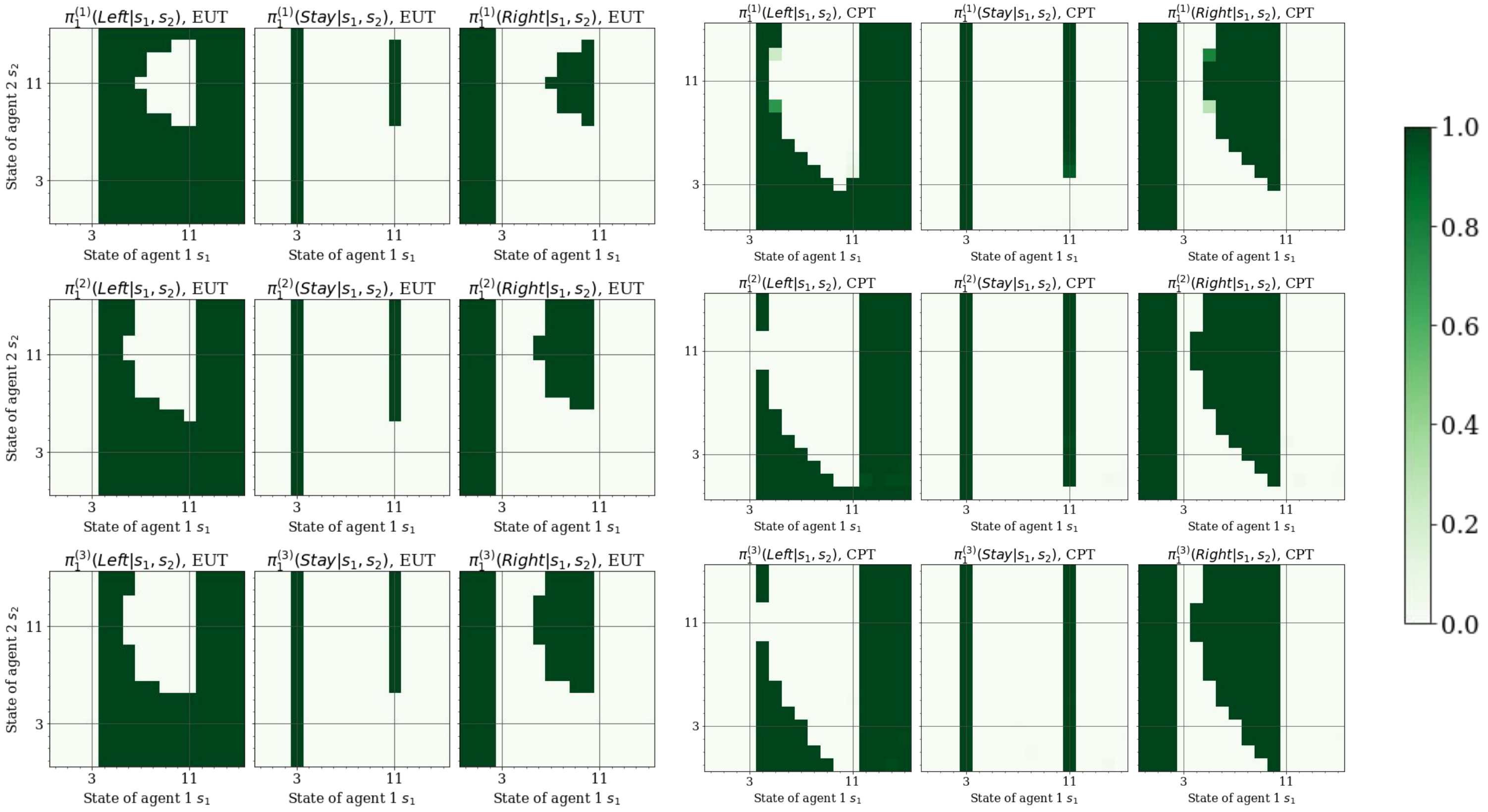


Values



$$b_1 = b_2 = 0, \beta_1 = \beta_2 = 0.9 \qquad u(r) = r \quad w(p) = e^{-0.5(-\log(p))^{0.9}}$$

Policies



$b_1 = b_2 = 0, \beta_1 = \beta_2 = 0.9$

$u(r) = r \quad w(p) = e^{-0.5(-\log(p))^{0.9}}$

Agent Distribution

$$P_{s,s'}^{\pi_1^{(k_1)},\pi_2^{(k_2)}} = \sum_{\substack{a_1 \in A_1(s) \\ a_2 \in A_2(s)}} P_{s,s'}^{a_1,a_2} \pi_1^{(k_1)}(a_1|s)\pi_2^{(k_2)}(a_2|s)$$

Conditioned on the policies, a Markov Game becomes a Markov Chain.

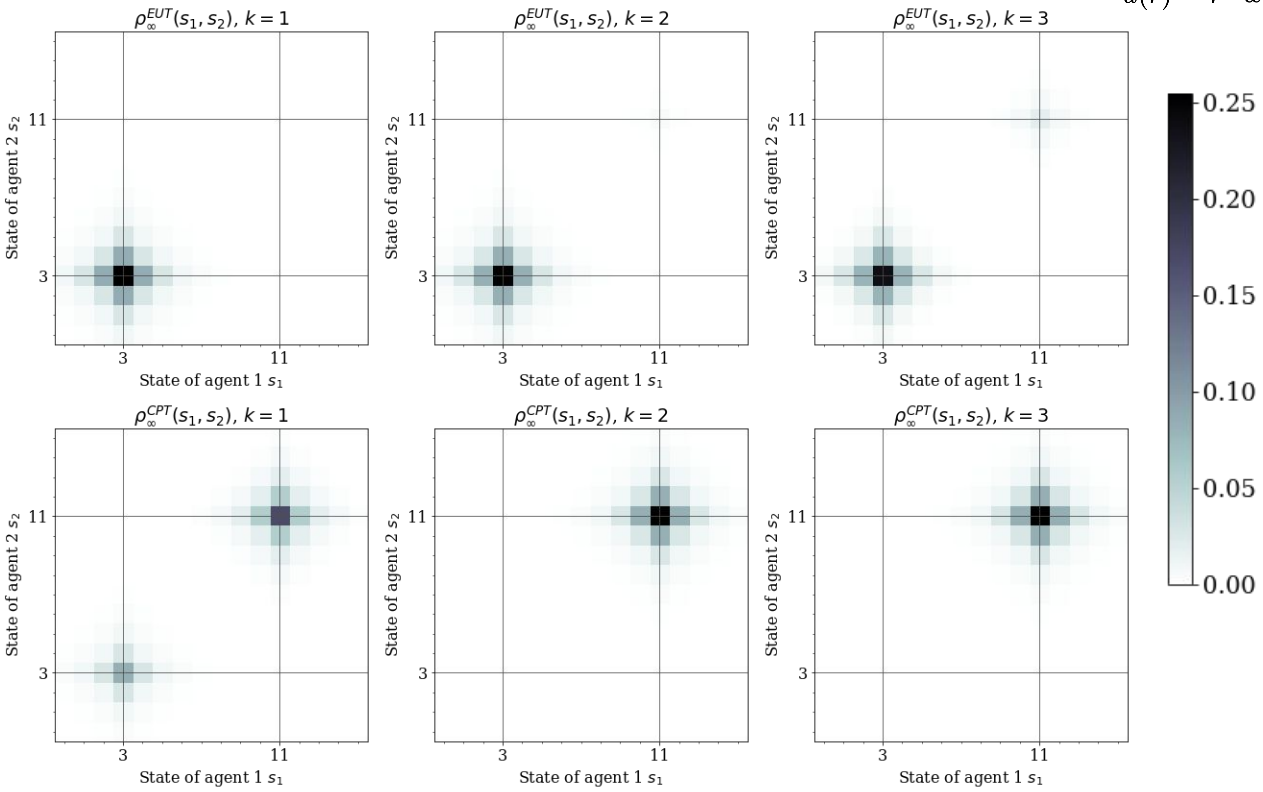
$$\rho_\infty^{(k_1,k_2)} = P^{\pi_1^{(k_1)},\pi_2^{(k_2)}} \rho_\infty^{(k_1,k_2)}$$

Stationary distribution of agents indicates equilibrium.

Agent Distribution

$$b_1 = b_2 = 0, \beta_1 = \beta_2 = 0.9$$

$$u(r) = r \quad w(p) = e^{-0.5(-\log(p))^{0.9}}$$



Roadmap

Theory of Mind
Definition
Importance

Game Theory and Coordination
Expected Utility Theory
Cumulative Prospect Theory
Example - Stag Hunt

Coordination over time
Markov Games
Level-K Theory of Mind
Example - Grid Stag Hunt

Conclusion and Future Work

Conclusion & Future Work

Risk-sensitive agents equipped with ToM show **increased coordination**.

- ➔ Improving the algorithmic performance of CPT value to scale with the state and action spaces.
- ➔ Apply to current and new applications where agents represent humans.
- ➔ Sensitivity analysis of all parameters is required to understand the limitations of the model.
  - ◆ E.g. understanding short-term vs long-term risk.
- ➔ Add more cognitive biases and mechanisms to improve descriptive power of the model.

HERE BE DRAGONS

Saint Petersburg Paradox

Saint Petersburg Paradox

How much would you be willing to pay to get into this gamble?

A coin is tossed repeatedly until, at the k-th toss, it comes up Heads. You get $\$2^k$.

A mathematician: “Calculate the expected value of this gamble and pay less than that value.”

$$V(R) = \mathbf{E}[R] = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2^k}\right) = \sum_{k=1}^{\infty} 1 = \infty$$

St. Petersburg paradox

Saint Petersburg Paradox

$$V(R) = \mathbf{E}[R] = \sum_{k=1}^{\infty} 2^k \left(\frac{1}{2^k}\right) = \sum_{k=1}^{\infty} 1 = \infty$$

Daniel Bernoulli:

"The determination of the value of an item must not be based on the price, but rather on the utility it yields…"

$$V^{\text{EUT}}(R) = \mathbf{E}[u(R)] = \sum_{k=1}^{\infty} u\left(2^k\right) \left(\frac{1}{2^k}\right)$$

Utility function

$$u(x) = \log(x)$$

$$V^{\text{EUT}}(R) = \sum_{k=1}^{\infty} \log\left(2^k\right) \left(\frac{1}{2^k}\right) = 2\log(2) \approx 1.39 < \infty$$

Sub-linear utility = risk aversion

Von Neumann-Morgenstern Axioms and Theorem

Von Neumann-Morgenstern Axioms and Theorem

von Neumann–Morgenstern axioms of choice:

- **Completeness**
  A preference ordering is complete iff, for any 2 outcomes X, Y, either $X \sim Y$ or $X \succ Y$ or $X \prec Y$.
- **Transitivity**
  For any 3 outcomes X, Y, Z, if $X \succeq Y$ and $Y \succeq Z$ then $X \succeq Z$.
- **Continuity**
  If $X \preceq Y \preceq Z$, then there exists a probability $p \in [0, 1]$ such that $pX + (1-p)Z \sim Y$.
- **Independence**
  If $X \preceq Y$, then for any Z and $p \in [0, 1]$, $pX + (1-p)Z \preceq pY + (1-p)Z$.

von Neumann-Morgenstern utility theorem:

If the preferences of an agent satisfy the 4 axioms above, there exists a function u such that for any two lotteries,

$$X \prec Y \qquad \text{if and only if} \qquad \mathbb{E}[u(X)] < \mathbb{E}[u(Y)]$$

Allais’ Paradox

Allais' Paradox

Choose A or B:

A

$0.11(\$1M) + 0.89(\$0)$

B

$0.1(\$5M) + 0.9(\$0)$

Allais' Paradox

A
0.11($1M) + 0.89($0)

B
0.1($5M) + 0.9($0)

Choose C or D:

C
$1M

D
0.1($5M) + 0.89($1M) + 0.01($0)

Allais' Paradox

A

$0.11(\$1M) + 0.89(\$0)$

B

$0.1(\$5M) + 0.9(\$0)$

$=0.11[10/11(\$5M)+1/11(\$0)]+0.89(\$0)$

C

$\$1M$

D

$0.1(\$5M) + 0.89(\$1M) + 0.01(\$0)$

Allais' Paradox

A

0.11($1M) + 0.89($0)

B

0.1($5M) + 0.9($0)

=0.11[10/11($5M)+1/11($0)]+0.89($0)

C

$1M

D

0.1($5M) + 0.89($1M) + 0.01($0)

Allais' Paradox

A

0.11($1M) + 0.89($0)

B

0.1($5M) + 0.9($0)

=0.11[10/11($5M)+1/11($0)]+0.89($0)

C

$1M

D

0.1($5M) + 0.89($1M) + 0.01($0)

=0.11[10/11($5M)+1/11($0)]+ 0.89($1M)

Allais' Paradox

A

$0.11(\$1M) + 0.89(\$0)$

B

$0.1(\$5M) + 0.9(\$0)$

$=0.11[10/11(\$5M)+1/11(\$0)]+0.89(\$0)$

C

$1M

D

$0.1(\$5M) + 0.89(\$1M) + 0.01(\$0)$

$=0.11[10/11(\$5M)+1/11(\$0)]+ 0.89(\$1M)$

Allais' Paradox

Violation of Independence Axiom if (A,D) or (B,C)

A

$0.11(\$1M) + 0.89(\$0)$

B

$0.1(\$5M) + 0.9(\$0)$

$=0.11[10/11(\$5M)+1/11(\$0)]+0.89(\$0)$

C

$1M

D

$0.1(\$5M) + 0.89(\$1M) + 0.01(\$0)$

$=0.11[10/11(\$5M)+1/11(\$0)]+ 0.89(\$1M)$

M. Allais. *Le Comportement de l'Homme Rationel devant le Risque, Critique des Postulates et Axiomes de l'École Americaine*, Econometrica, October 1953, 21, 503–46.
Donald G. Morrison. *On the Consistency of Preferences in Allais' Paradox*. Behavioral Science, September 1967, 12, 373–83.
Slovic, Paul and A. Tversky, *Who Accepts Savage's Axiom?*, Behavioral Science, November 1974, 19, 368–73.

EUT vs PT vs CPT
Die Cast Example

Comparison - A roll of a die

$$([1,1/6],[2,1/6],[3,1/6],[4,1/6],[5,1/6],[6,1/6])$$

$$\underbrace{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad}_{\textbf{Gains}}$$

$$V^{\mathrm{EUT}} = \sum_{k=1}^{6} u(k)\left(\frac{1}{6}\right) = \frac{1}{6}(u(1)+u(2)+u(3)+u(4)+u(5)+u(6))$$

$V^{EUT} \approx 2.85$

$$V^{PT} = \sum_{k=1}^{6} u(k)w\left(\frac{1}{6}\right) = w\left(\frac{1}{6}\right)(u(1)+u(2)+u(3)+u(4)+u(5)+u(6))$$

$V^{PT} \approx 7.35 > 6$

$$V^{CPT} = \sum_{k=1}^{6} u(k)[w(P(R \geq k)) - w(P(R > k))]$$

$$= u(1)(w(1)-w(5/6)) + u(2)(w(5/6)-w(4/6)) + u(3)(w(4/6)-w(3/6))$$

$$+ u(4)(w(3/6)-w(2/6)) + u(5)(w(2/6)-w(1/6)) + u(6)(w(1/6)-w(0))$$

$V^{CPT} \approx 3.48$

$$u(x) = x^{0.85} \qquad w(x) = e^{-0.5(-log(x))^{0.9}}$$

Stag Hunt Nash Equilibria

Stag Hunt

Rationality

A player wishes to maximize his utility.

Common Knowledge of Rationality

{Every player knows that}$^{\infty}$ every player is rational.

		2: Stag	2: Hare
1	Stag	10,10	0,2
1	Hare	2,0	2,2

Best Response (against $\boldsymbol{\pi}_{-i}$)

Policy $\pi_i^* \in BR(\boldsymbol{\pi}_{-i})$ which provide the most utility against $\boldsymbol{\pi}_{-i}$.

Nash Equilibrium

A joint policy $\boldsymbol{\pi}^*$ such that

$$\pi_i^* \in BR(\boldsymbol{\pi}^*_{-i})$$

for every agent i.

Deterministic Nash Equilibrium in Stag Hunt

1 \ 2	Stag	Hare
Stag	**10,10**	**0,2**
Hare	**2,0**	**2,2**

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

		2: Stag	2: Hare
1	Stag	10,10	0,2
1	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.

		2: Stag	2: Hare
1	Stag	10,10	0,2
	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.

		2: Stag	2: Hare
1	Stag	**10**,10	0,2
1	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

		2: Stag	2: Hare
1	Stag	<u>10</u>,10	0,2
1	Hare	2,0	<u>2</u>,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From 1's perspective:

If 2 chooses Stag, then 1 chooses Stag.
If 2 chooses Hare, then 1 chooses Hare.

From 2's perspective:

		2: Stag	2: Hare
1	Stag	10,10	0,2
1	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From 1's perspective:

If 2 chooses **Stag**, then 1 chooses **Stag**.
If 2 chooses **Hare**, then 1 chooses **Hare**.

From 2's perspective:

If 1 chooses **Stag**, then 2 chooses **Stag**.

1 \ 2	**Stag**	**Hare**
Stag	10,10	0,2
Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.

		2: Stag	2: Hare
1	Stag	10,10	0,2
	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.
If **1** chooses **Hare**, then **2** chooses **Hare**.

		2: Stag	2: Hare
1	Stag	**10,10**	**0,2**
	Hare	**2,0**	**2,2**

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.
If **1** chooses **Hare**, then **2** chooses **Hare**.

		2: Stag	2: Hare
1	Stag	10,10	0,2
	Hare	2,0	2,2

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.
If **1** chooses **Hare**, then **2** chooses **Hare**.

1 \ 2	Stag	Hare
Stag	**10,10**	**0,2**
Hare	**2,0**	**2,2**

What is the Nash Equilibrium here?

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.
If **1** chooses **Hare**, then **2** chooses **Hare**.

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

What is the Nash Equilibrium here?

Two NEs: $\boldsymbol{\pi}' = (Stag, Stag)$ $\boldsymbol{\pi}'' = (Hare, Hare)$

Deterministic Nash Equilibrium in Stag Hunt

From **1**'s perspective:

If **2** chooses **Stag**, then **1** chooses **Stag**.
If **2** chooses **Hare**, then **1** chooses **Hare**.

From **2**'s perspective:

If **1** chooses **Stag**, then **2** chooses **Stag**.
If **1** chooses **Hare**, then **2** chooses **Hare**.

		2: **Stag**	2: **Hare**
1	**Stag**	**10,10**	**0,2**
	Hare	**2,0**	**2,2**

What is the Nash Equilibrium here?

Two NEs: $\boldsymbol{\pi}' = (Stag, Stag) \qquad \boldsymbol{\pi}'' = (Hare, Hare)$

What about stochastic policies?

Stochastic Nash Equilibrium in Stag Hunt

From **1**'s perspective, if **2** chooses **Stag** with probability q:

2 chooses a distribution π_2 that makes **1** indifferent between **Stag** and **Hare**

		2: Stag	2: Hare
1	Stag	10,10	0,2
	Hare	2,0	2,2

Stochastic Nash Equilibrium in Stag Hunt

From **1**'s perspective, if **2** chooses **Stag** with probability q:

2 chooses a distribution π_2 that makes **1** indifferent between **Stag** and **Hare**

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = \mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)]$$

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

Stochastic Nash Equilibrium in Stag Hunt

From **1**'s perspective, if **2** chooses **Stag** with probability q:

2 chooses a distribution π_2 that makes **1** indifferent between **Stag** and **Hare**

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = \mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)]$$

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = 10q + 0(1-q) = 10q$$

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

Stochastic Nash Equilibrium in Stag Hunt

From **1**'s perspective, if **2** chooses **Stag** with probability q:

2 chooses a distribution π_2 that makes **1** indifferent between **Stag** and **Hare**

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = \mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)]$$

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = 10q + 0(1-q) = 10q$$

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)] = 2q + 2(1-q) = 2$$

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

Stochastic Nash Equilibrium in Stag Hunt

From **1**'s perspective, if **2** chooses **Stag** with probability q:

2 chooses a distribution π_2 that makes **1** indifferent between **Stag** and **Hare.**

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = \mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)]$$

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Stag, a_2)] = 10q + 0(1-q) = 10q$$

$$\mathbb{E}_{a_2 \sim \pi_2(\cdot)}[u_1(Hare, a_2)] = 2q + 2(1-q) = 2$$

$$\Leftrightarrow q = \frac{1}{5}$$

$$\pi_2^* = \left(\frac{1}{5}, \frac{4}{5}\right)$$

		2: Stag	2: Hare
1	Stag	10,10	0,2
1	Hare	2,0	2,2

By symmetry, $\pi_1^* = \left(\frac{1}{5}, \frac{4}{5}\right)$.

Stochastic Nash Equilibrium: $\left(\left(\frac{1}{5}, \frac{4}{5}\right), \left(\frac{1}{5}, \frac{4}{5}\right)\right)$

Deterministic Nash Equilibria: $\left((0,1), (0,1)\right)$

$\left((1,0), (1,0)\right)$

Nash Equilibria in Stag Hunt

Stochastic Nash Equilibrium: $\left(\left(\frac{1}{5}, \frac{4}{5}\right), \left(\frac{1}{5}, \frac{4}{5}\right)\right)$

Deterministic Nash Equilibria: $\left((0, 1), (0, 1)\right)$

$\left((1, 0), (1, 0)\right)$

1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

Quantal Response Equilibrium

Quantal Response Equilibrium

For each agent i and each action j:

Assume all expected utilities are observed with some zero-mean **error** ε_{ij} : $\hat{u}_{ij} = \bar{u}_{ij} + \varepsilon_{ij}$

Assume players are **rational**; they will **choose action that maximizes observed expected utility.**
Player *i* will use the action *j* that $\bar{u}_{ij} + \varepsilon_{ij} \geq \bar{u}_{ik} + \varepsilon_{ik}, \forall_{k \in A_i}$.

This induces a stochastic policy with full support.

Let m_i be the size of player *i*'s action set. The preference shock region that player *i* chooses action *j* is

$$R_{ij}(\bar{u}_i(\boldsymbol{\pi}_{-i})) = \{\varepsilon_i \in \mathbb{R}^{m_i} : \bar{u}_{ij}(\boldsymbol{\pi}_{-i}) + \varepsilon_{ij} \geq \bar{u}_{ik}(\boldsymbol{\pi}_{-i}) + \varepsilon_{ik}, \forall k \in \{1, \ldots, m_i\}\}$$

Quantal Response Equilibrium

Let m_i be the size of player *i*'s action set. The preference shock region that player *i* chooses action *j* is

$$R_{ij}(\bar{u}_i(\boldsymbol{\pi}_{-i})) = \{\varepsilon_i \in \mathbb{R}^{m_i} : \bar{u}_{ij}(\boldsymbol{\pi}_{-i}) + \varepsilon_{ij} \geq \bar{u}_{ik}(\boldsymbol{\pi}_{-i}) + \varepsilon_{ik}, \forall k \in \{1, \ldots, m_i\}\}$$

The probability player *i* chooses action *j* is

$$\sigma_{ij}(\boldsymbol{\pi}_{-i}) = \int_{R_{ij}(\bar{u}_i(\boldsymbol{\pi}_{-i}))} f_i(\varepsilon_i) d\varepsilon_i$$

statistical reaction function (or quantal response function) → $\sigma_{ij}(\boldsymbol{\pi}_{-i})$

Joint p.d.f of player *i*'s preference shocks → $f_i(\varepsilon_i)$

In a normal-form game, a quantal response equilibrium is a joint policy $\boldsymbol{\pi}^*$ such that,

$$\pi^*_{ij} = \sigma_{ij}(\boldsymbol{\pi}^*_{-i}), \forall (i, j) \in N \times \{1, \ldots, m_i\}$$

Quantal Response Equilibrium

Which distribution for the errors should we choose? Draw inspiration from behavioral choice theory.

Assume, for every player and every action, ε_{ij} are i.i.d. and follow a $\text{Log-Weibull}(0, \lambda)$ distribution.

$$\sigma_{ij}(\bar{u}_i(\boldsymbol{\pi}_{-i})) = \frac{e^{\lambda \bar{u}_{ij}(\boldsymbol{\pi}_{-i})}}{\sum_{k=1}^{m_i} e^{\lambda \bar{u}_{ik}(\boldsymbol{\pi}_{-i})}}$$

This leads to the Logistic QRE:

$$\pi^*_{ij}(\bar{u}_i(\boldsymbol{\pi}^*_{-i})) = \frac{e^{\lambda \bar{u}_{ij}(\boldsymbol{\pi}^*_{-i})}}{\sum_{k=1}^{m_i} e^{\lambda \bar{u}_{ik}(\boldsymbol{\pi}^*_{-i})}}$$

R. Luce. *A Theory of Individual Choice Behavior*, 1957.
R. McKelvey, T. Palfrey. *Quantal Response Equilibria for Normal Form Games*, Games and Economic Behavior, 1994 vol: 10 pp: 6-38.

Quantal Response Equilibrium

For each agent i and each action j:

Assume all expected utilities are observed with some zero-mean **error** ε_{ij}: $\hat{u}_{ij} = \bar{u}_{ij} + \varepsilon_{ij}$
Assume players are **rational**; they will **choose action that maximizes observed expected utility.**
Player *i* will use the action *j* that $\bar{u}_{ij} + \varepsilon_{ij} \geq \bar{u}_{ik} + \varepsilon_{ik}, \forall_{k \in A_i}$.

Logistic Quantal Response Equilibrium, $\varepsilon_{ij} \overset{i.i.d.}{\sim} Gumbel(0, \lambda^{-1})$**,** based on decision theory:

$$\pi^*_{ij}(\bar{u}_i(\boldsymbol{\pi}^*_{-i})) = \frac{e^{\lambda \bar{u}_{ij}(\boldsymbol{\pi}^*_{-i})}}{\sum_{k=1}^{m_i} e^{\lambda \bar{u}_{ik}(\boldsymbol{\pi}^*_{-i})}}$$

Inverse negative temperature

This induces a stochastic policy with full support.

R. Luce. *A Theory of Individual Choice Behavior*, 1957.
R. McKelvey, T. Palfrey. *Quantal Response Equilibria for Normal Form Games*, Games and Economic Behavior, 1994 vol: 10 pp: 6-38.

QRE in Stag Hunt

$$\pi^*_{ij}(\bar{u}_i(\boldsymbol{\pi}^*_{-i})) = \frac{e^{\lambda \bar{u}_{ij}(\boldsymbol{\pi}^*_{-i})}}{\sum_{k=1}^{m_i} e^{\lambda \bar{u}_{ik}(\boldsymbol{\pi}^*_{-i})}}$$

$$p = \pi^*_{1,Stag}$$
$$q = \pi^*_{2,Stag}$$

$$\begin{cases} \pi^*_{1,Stag}(\bar{u}_1(\pi^*_2)) = \dfrac{e^{\lambda \bar{u}_{1,Stag}(\pi^*_2)}}{e^{\lambda \bar{u}_{1,Stag}(\pi^*_2)} + e^{\lambda \bar{u}_{1,Hare}(\pi^*_2)}} \\ \pi^*_{1,Hare}(\bar{u}_1(\pi^*_2)) = 1 - \pi^*_{1,Stag}(\bar{u}_1(\pi^*_2)) \\ \pi^*_{2,Stag}(\bar{u}_2(\pi^*_1)) = \dfrac{e^{\lambda \bar{u}_{2,Stag}(\pi^*_1)}}{e^{\lambda \bar{u}_{2,Stag}(\pi^*_1)} + e^{\lambda \bar{u}_{2,Hare}(\pi^*_1)}} \\ \pi^*_{2,Hare}(\bar{u}_2(\pi^*_1)) = 1 - \pi^*_{2,Stag}(\bar{u}_2(\pi^*_1)) \end{cases}$$

$$\begin{cases} p = \dfrac{e^{10\lambda q}}{e^{10\lambda q} + e^{2\lambda}} = \dfrac{1}{1 + e^{2\lambda - 10\lambda q}} \\ q = \dfrac{e^{10\lambda p}}{e^{10\lambda p} + e^{2\lambda}} = \dfrac{1}{1 + e^{2\lambda - 10\lambda p}} \end{cases}$$

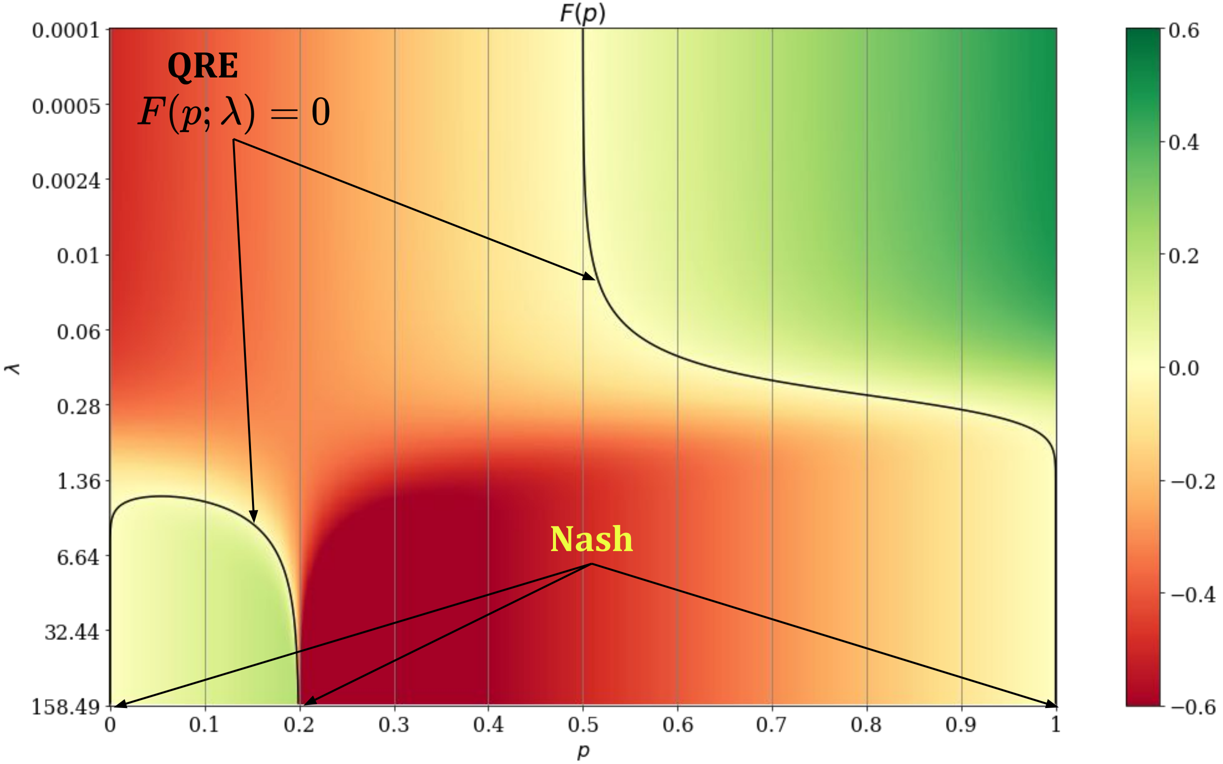
1 \ 2	Stag	Hare
Stag	10,10	0,2
Hare	2,0	2,2

Game is symmetric, $p = q$, therefore $$p = \frac{1}{1 + e^{2\lambda - 10\lambda p}} \Leftrightarrow p - \frac{1}{1 + e^{2\lambda - 10\lambda p}} = 0 \Leftrightarrow F(p; \lambda) = 0$$

Finding the QRE means solving a transcendental equation.

QRE in Stag Hunt

$$\text{QRE} \xrightarrow[\lambda \to \infty]{} \text{NE}$$



Markov Decision Process

Markov Decision Process

$$(S, A, p, u, \gamma)$$

Set of states S
Set of actions A
Probability transition function $p : S \times A \times S \to [0, 1]$
Utility function $u : S \times A \to \mathbb{R}$
Discount factor $\gamma \in (0, 1)$

Single agent in a stochastic environment deciding which actions to take to maximize the expected discounted sum of utilities, a.k.a. the value.

Choose a policy π (which now depends on the state), that maximizes, for every starting state, some value functional:

$$V(s, \pi)$$

Markov Decision Process

EUT infinite-horizon approach - derive a recursive equation for the discounted sum of utilities:

$$
\begin{aligned}
V(s,\pi) &= \mathbb{E}_{s_{t+1}\sim p(\cdot|s_t,\pi(s_t))}\left[\sum_{t=0}^{\infty}\gamma^t u(s_t,\pi(s_t))\middle| s_0 = s\right] \\
&= \mathbb{E}_{s_{t+1}\sim p(\cdot|s_t,\pi(s_t))}\left[u(s_0,\pi(s_0)) + \sum_{t=1}^{\infty}\gamma^t u(s_t,\pi(s_t))\middle| s_0 = s\right] \\
&= u(s,\pi(s)) + \mathbb{E}_{s_{t+1}\sim p(\cdot|s_t,\pi(s_t))}\left[\sum_{t=1}^{\infty}\gamma^t u(s_t,\pi(s_t))\middle| s_0 = s\right] \\
&= u(s,\pi(s)) + \gamma\sum_{s'\in S}\mathbb{E}_{s_{t+1}\sim p(\cdot|s_t,\pi(s_t))}\left[\sum_{t=1}^{\infty}\gamma^{t-1} u(s_t,\pi(s_t))\middle| s_1 = s'\right] p(s'|s_0,\pi(s_0)) \\
&= u(s,\pi(s)) + \gamma\sum_{s'\in S}V(s',\pi)p(s'|s,\pi(s))
\end{aligned}
$$

Markov Decision Process

EUT infinite-horizon approach - derive a recursive equation for the discounted sum of utilities:

Bellman Equation

$$V(s,\pi) = u(s,\pi(s)) + \gamma \sum_{s' \in S} V(s',\pi) p(s'|s,\pi(s))$$

$$V^*(s) = \max_{\pi} \left\{ u(s,\pi(s)) + \gamma \sum_{s' \in S} V(s',\pi) p(s'|s,\pi(s)) \right\}$$

Solution Methods:

- Value Iteration
- Policy Iteration
- Q-Learning
- Many others...

Is there a Bellman equation for CPT-Value?

MDP with CPT-value

Markov Decision Process with CPT-Value

$$P_s^a(\cdot) = P(\cdot | s_t = s, a_t = a)$$

Discount factor

Reference point

Policy

$$V^{\text{CPT}}(s,\pi) = \int_0^{\infty} w^+ \left(\sum_{a \in A(s)} P_s^a (u^+((r(s) + \beta V^{\text{CPT}}(S,\pi) - b)_+) > \epsilon) \pi(a|s) \right) d\epsilon$$

$$- \int_0^{\infty} w^- \left(\sum_{a \in A(s)} P_s^a (u^-((r(s) + \beta V^{\text{CPT}}(S,\pi) - b)_-) > \epsilon) \pi(a|s) \right) d\epsilon$$

Gains

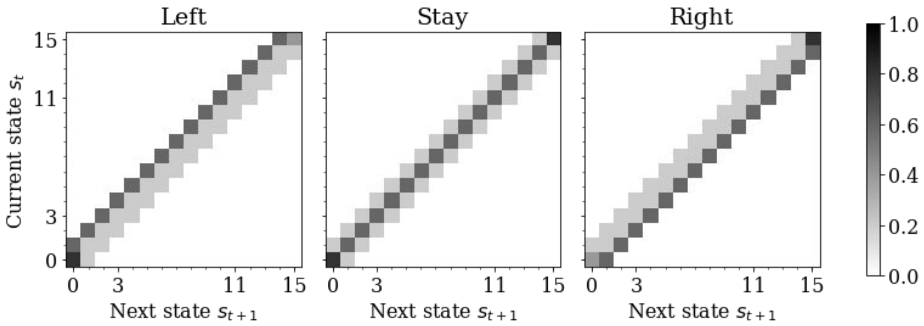
Losses

$$(\cdot)_+ = \max(0,\cdot) \quad (\cdot)_- = -\min(0,\cdot)$$

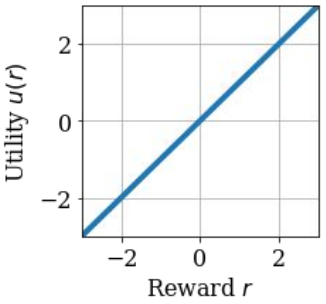
$$V^{\text{CPT}*}(s) = \max_{\pi} V^{\text{CPT}}(s,\pi), \text{ for all s.}$$

A. Ruszczyński. *Risk-averse dynamic programming for markov decision processes*. Mathematical Programming, 125 (2010), pp. 235–261.
K. Lin. *Stochastic Systems with Cumulative Prospect Theory,* PhD Thesis, 2013.

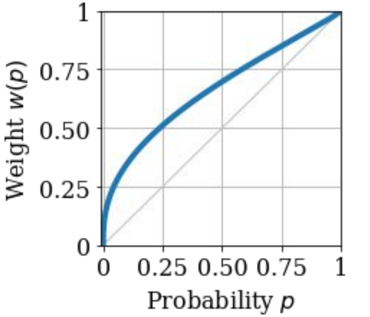
Example

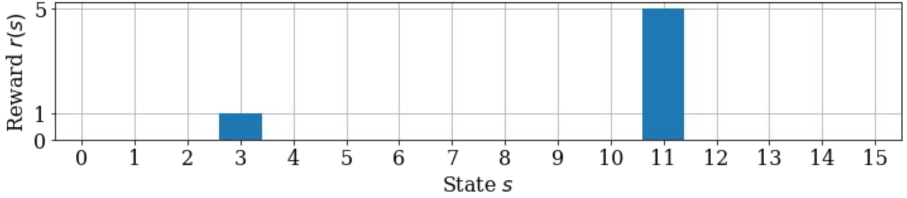


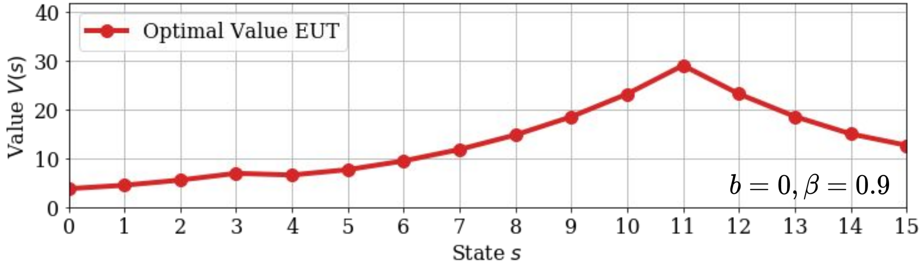
$u(r) = r$



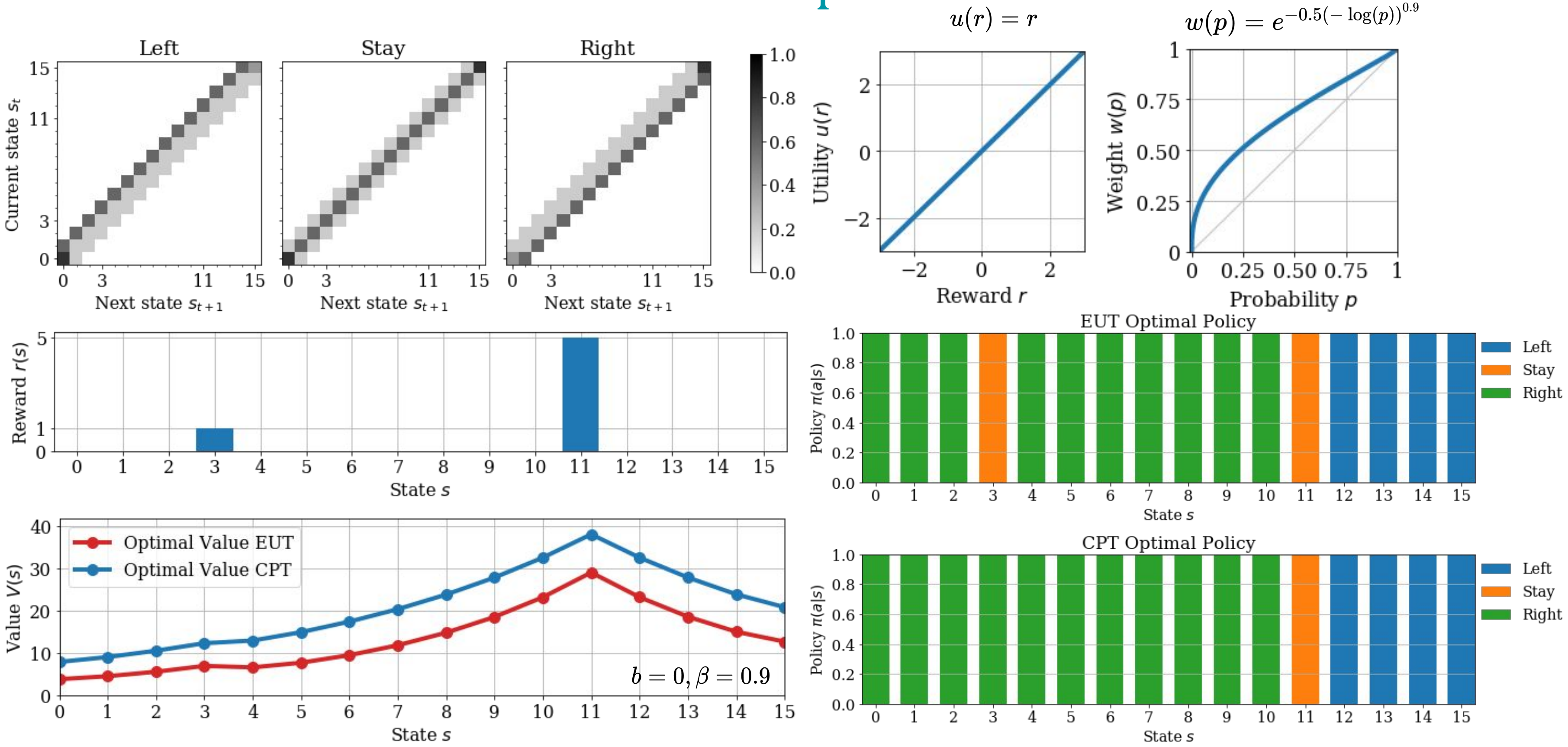
$w(p) = e^{-0.5(-\log(p))^{0.9}}$







Example


Left
Stay
Right
Current state s_t
Next state s_{t+1}
$u(r) = r$
Utility $u(r)$
Reward r
$w(p) = e^{-0.5(-\log(p))^{0.9}}$
Weight $w(p)$
Probability p
Reward $r(s)$
State s
EUT Optimal Policy
Policy $\pi(a|s)$
CPT Optimal Policy
Optimal Value EUT
Optimal Value CPT
Value $V(s)$
$b = 0, \beta = 0.9$

Linearly-solvable MDP

Linearly-Solvable Markov Decision Process

MDPs are too general.
Assume $\gamma = 1$.
Assume a policy is an |S|-dimensional vector, effectively expanding the action space.
Assume the transition probability function can be deformed via some continuous action specified by the policy:

$$p(s'|s,\pi) = p(s'|s)e^{\pi(s')}, \text{ with } \sum_{s'\in S} p(s'|s)e^{\pi(s')} = 1$$

Assume utility can be decomposed into a state utility and a control cost

$$u(s,\pi) = u(s) - D_{KL}\left(p(\cdot|s,\pi)||p(\cdot|s)\right)$$

A.S.A we get a closed-form expression for the dynamics of an agent which moves optimally without explicitly finding the optimal policy:

Optimal policy fully specified by value

$$p(s'|s,V^*) = \frac{p(s'|s)e^{\lambda V^*(s')}}{\sum_{s''\in S} p(s''|s)e^{\lambda V^*(s'')}}$$

Linearly-Solvable Markov Decision Process

The Bellman equation becomes:

$$V^*(s) = u(s) + \sum_{s' \in S} V^*(s') p(s'|s, V^*)$$

Vectorized version:

$$V^* = u + V^* P(V^*)$$

$$P_{s,s'}(V^*) = p(s'|s, V^*)$$

Find the optimal value via Robbins-Monro algorithm [Robbins,1951]:

$$V_{t+1} \leftarrow u + V_t P(V_t)$$

H. Robbins and S. Monro. *A stochastic approximation method.* The annals of mathematical statistics, 1951. pp. 400-407.

LMDP and QRE

Why LMDP?

- Easier to solve.
- Similar interpretation as the QRE.

QRE

$$\pi^*_{ij}(\bar{u}_i(\boldsymbol{\pi}^*_{-i})) = \frac{e^{\lambda \bar{u}_{ij}(\boldsymbol{\pi}^*_{-i})}}{\sum_{k=1}^{m_i} e^{\lambda \bar{u}_{ik}(\boldsymbol{\pi}^*_{-i})}}$$

LMDP

$$p(s'|s, V^*) = \frac{p(s'|s)e^{\lambda V^*(s')}}{\sum_{s'' \in S} p(s''|s)e^{\lambda V^*(s'')}}$$

Game Theory of Mind Model

Model

Framework: LMDP (and all its assumptions)

$$N = \{1, 2\}$$
$$S = S_1 \times S_2, S_1 = S_2 = \{1, \ldots, 16\}$$

$$V_1^* = u_1 + V_1^* P(V_1^*, V_2^*)$$
$$V_2^* = u_2 + V_2^* P(V_1^*, V_2^*)$$

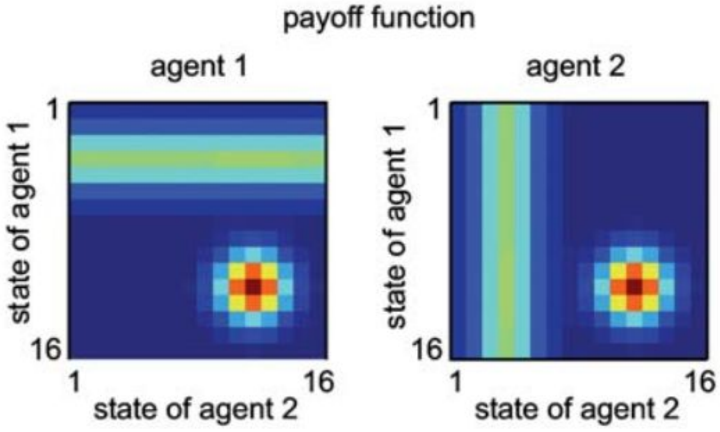
Sequential and independent moves

$$P(V_1, V_2) = P_2(V_2) P_1(V_1)$$

$$[P_i(V_i)]_{s,s'} = p_i(s'|s, V_i) = \frac{\Pi_i(s'|s) e^{\lambda V_i(s')}}{\sum_{s'' \in S} \Pi_i(s''|s) e^{\lambda V_i(s'')}}$$

$$\Pi_1 = I \otimes P_1$$
$$\Pi_2 = P_2 \otimes I$$

Uncontrolled Transition Probabilities
1 cannot change state of 2 and vice-versa.



Model

$$V_1^* = u_1 + V_1^* P(V_1^*, V_2^*)$$
$$V_2^* = u_2 + V_2^* P(V_1^*, V_2^*)$$

Hunter 1 wishes to find his optimal value, but cannot, since he does not know the value function of Hunter 2.

How to solve this?

Start by assume the other hunter moves randomly. Then assume best responses against that.

$$V_1^{(1)} = u_1 + V_1^{(1)} P(V_1^{(1)}, 0) \qquad V_2^{(1)} = u_2 + V_2^{(1)} P(0, V_2^{(1)})$$

$$V_1^{(2)} = u_1 + V_1^{(2)} P(V_1^{(2)}, V_2^{(1)}) \qquad V_2^{(2)} = u_2 + V_2^{(2)} P(V_1^{(1)}, V_2^{(2)})$$

$$\vdots \qquad \vdots$$

$$V_1^{(k)} = u_1 + V_1^{(k)} P(V_1^{(k)}, V_2^{(k-1)}) \qquad V_2^{(k)} = u_2 + V_2^{(k)} P(V_1^{(k-1)}, V_2^{(k)})$$

Game Theory of Mind

[Yoshida, 2008] shows cooperation increases with recursion level *k*, i.e. the Hare solution is preferable for low sophistication levels and Stag becomes the equilibrium for higher sophistication levels.

Inference of the sophistication levels of the hunters can be made and Yoshida et. al. provide a method to manage and update the beliefs over the recursion levels based on past observations.

$$
\begin{aligned}
k_i^j &= \operatorname*{argmax}_{k_j \in \mathcal{K}_i} \{Pr(k_j | \boldsymbol{h}_T, \boldsymbol{k}_i)\} \\
&= \operatorname*{argmax}_{k_j \in \mathcal{K}_i} \{Pr(\boldsymbol{h}_T | k_j, \boldsymbol{k}_i) Pr(k_j)\} \\
&= \operatorname*{argmax}_{k_j \in \mathcal{K}_i} \left\{ Pr(s_0) \prod_{t=1}^{T} Pr\left(s_t | s_{t-1}, k_j, k_i(t)\right) Pr(k_j) \right\}
\end{aligned}
$$

Game Theory of Mind

A model inspired by:

- Expected utility theory-based control model (**Linearly-solvable MDP**)
- Quantal response equilibrium
- **Level-K cognitive model**

$$\pi_2^{(0)} \qquad \pi_1^{(0)}$$

$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(0)}}(s) = \pi_1^{(1)}(s) \qquad \pi_2^{(1)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(0)},\pi_2}(s)$$

$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(1)}}(s) = \pi_1^{(2)}(s) \qquad \pi_2^{(2)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(1)},\pi_2}(s)$$

$$\vdots \qquad \vdots$$

$$\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1,\pi_2^{(k-1)}}(s) = \pi_1^{(k)}(s) \qquad \pi_2^{(k)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(k-1)},\pi_2}(s)$$

Advantages

- Fits human data of humans.
- Can bound recursion levels to study boundedly rational agents.
- More plausible behavior than Nash equilibria refinements.

Disadvantages

- Recursion level is not the same across games.
- Recursion level changes as people learn to play or are more aware of the rules.

W. Yoshida, R. Dolan, Karl J. Friston. *Game theory of mind.* PLoS computational biology 4.12, 2008: e1000254.
Stahl, D. O. (1993). Evolution of Smartn Players. Games and Economic Behavior, 5(4), 604-617.

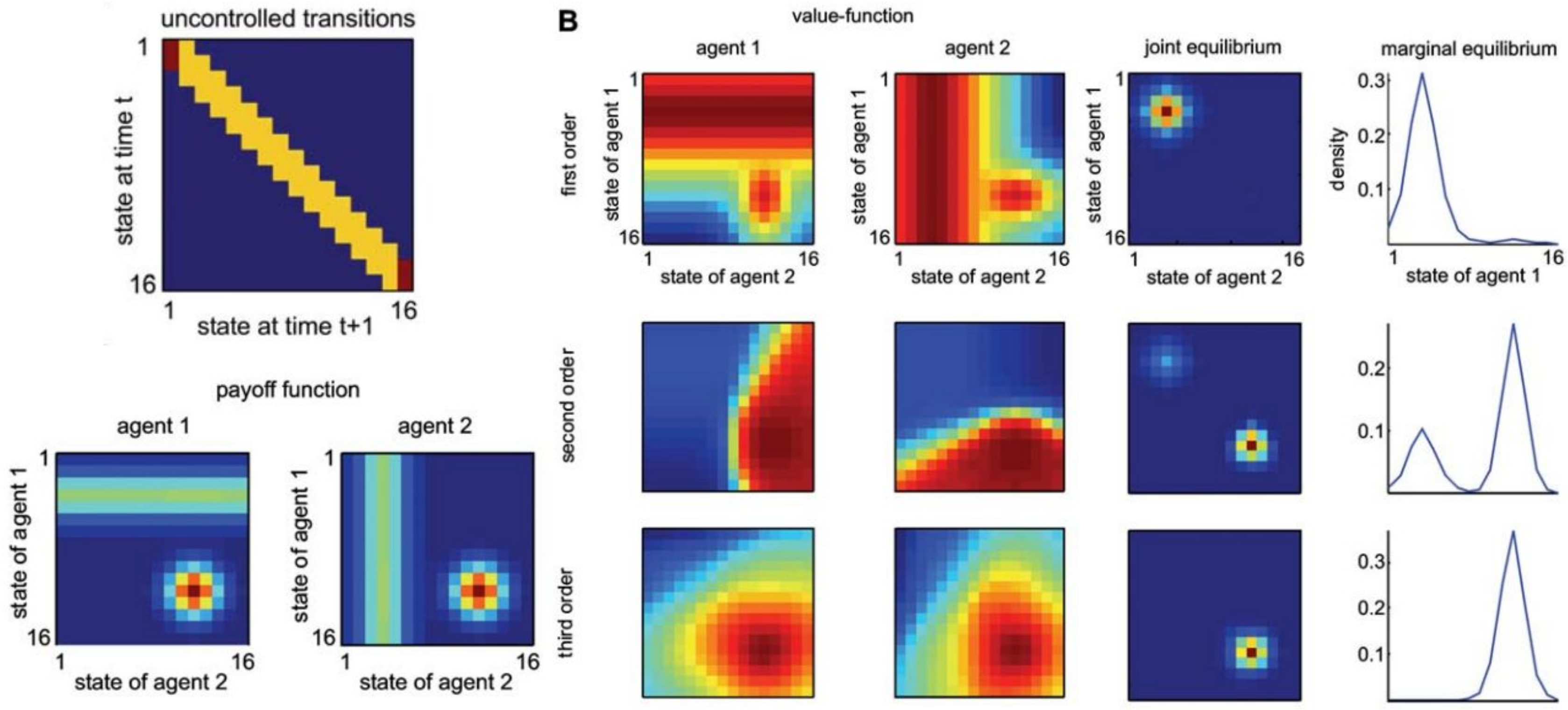
Game Theory of Mind

- Two hunters move repeatedly in sequence (1,2,1,...).
- 16 areas to hunt in.
- Hunters start at some random location.
- Each day, they hunt the area they are in, and can move to adjacent areas after or stay in the same area.
- Moving is stochastic (humans err).
- Two prey areas:
  - Hares: Fixed in area 4, can be hunted solo. Yields small meat.
  - Stags: Fixed in area 12, can only be hunted as a group. Yields big meat per hunter.
- Prey does not disappear after being hunted, they're infinite.
- Hunters know all of the above and are rational.

How should each hunter move each day so as to maximize his expected hunted meat?

W. Yoshida, R. Dolan, Karl J. Friston. *Game theory of mind.* PLoS computational biology 4.12, 2008: e1000254.

Game Theory of Mind



Maximizing the CPT-value

Maximizing CPT-Value

$$V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(s) = \int_0^\infty w_i^+\left(\sum_{a_i\in A_i(s)} P_{s,+}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon)\pi_i(a_i|s)\right)d\epsilon - \int_0^\infty w_i^-\left(\sum_{a_i\in A_i(s)} P_{s,-}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon)\pi_i(a_i|s)\right)d\epsilon$$

State space is **discrete**.
This means survival function is **piecewise constant**.

$$\{\epsilon_k^+ : \forall k > 0, \epsilon_k^+ \text{ is an ordered atom of } P_{a,+}^{a_i,\boldsymbol{\pi}_{-i}}, \epsilon_0^+ = 0\}_{k=0}^{K^+}$$
$$\{\epsilon_k^- : \forall k > 0, \epsilon_k^- \text{ is an ordered atom of } P_{a,-}^{a_i,\boldsymbol{\pi}_{-i}}, \epsilon_0^- = 0\}_{k=0}^{K^-}$$

$$V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(s) = \sum_{k=1}^{K^+} w_i^+\left(\sum_{a_i\in A_i(s)} P_{s,+}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon_k)\pi_i(a_i|s)\right)(\epsilon_k - \epsilon_{k-1}) - \sum_{k=1}^{K^-} w_i^-\left(\sum_{a_i\in A_i(s)} P_{s,-}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon_k)\pi_i(a_i|s)\right)(\epsilon_k - \epsilon_{k-1})$$

Atoms

Maximize the sum of nonlinear functions, instead of improper integral, over a simplex.

This work used scipy's implementation of SLSQP (sequential least squares quadratic programming), with (0,1) bounds and constrained the sum to unity.