The Fall of *homo economicus*

The Role of Cognitive Biases and Theory of Mind in Human Coordination

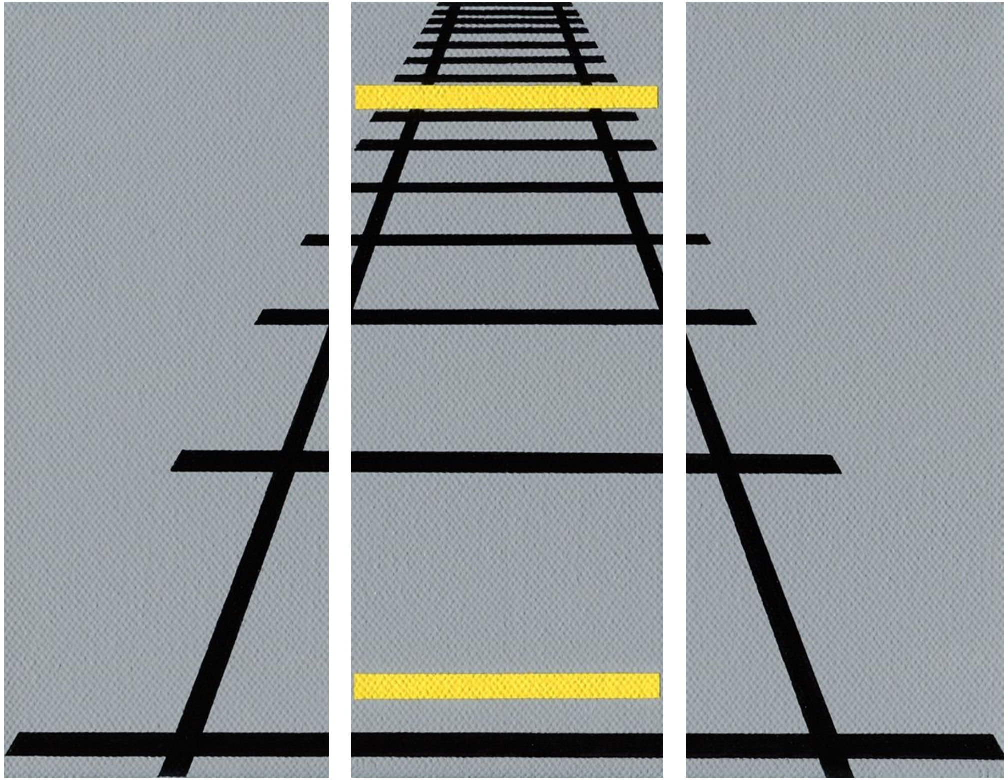
Master Thesis Defense
Instituto Superior Técnico | MMA

Advisors
Prof. Francisco Santos
Prof. Conceição Amado

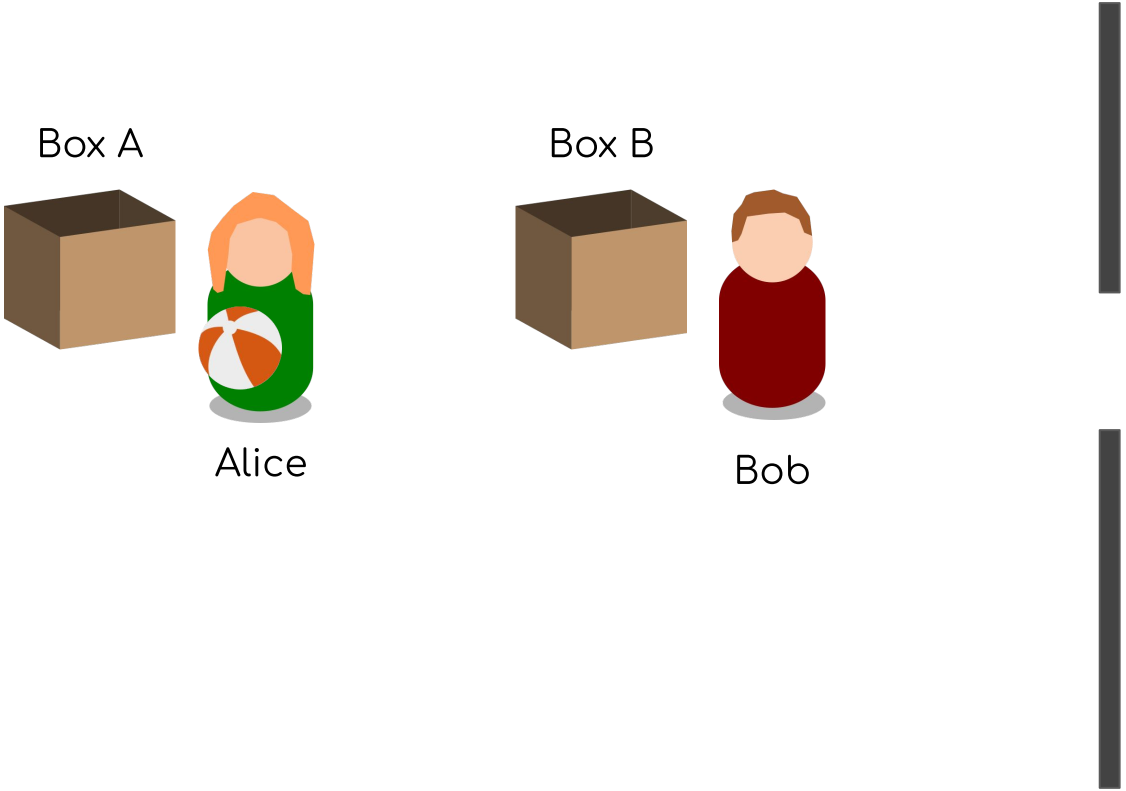
Candidate
Pedro Ferreira

Committee
Prof. António Pires
Prof. Francisco Santos
Prof. Alberto Sardinha

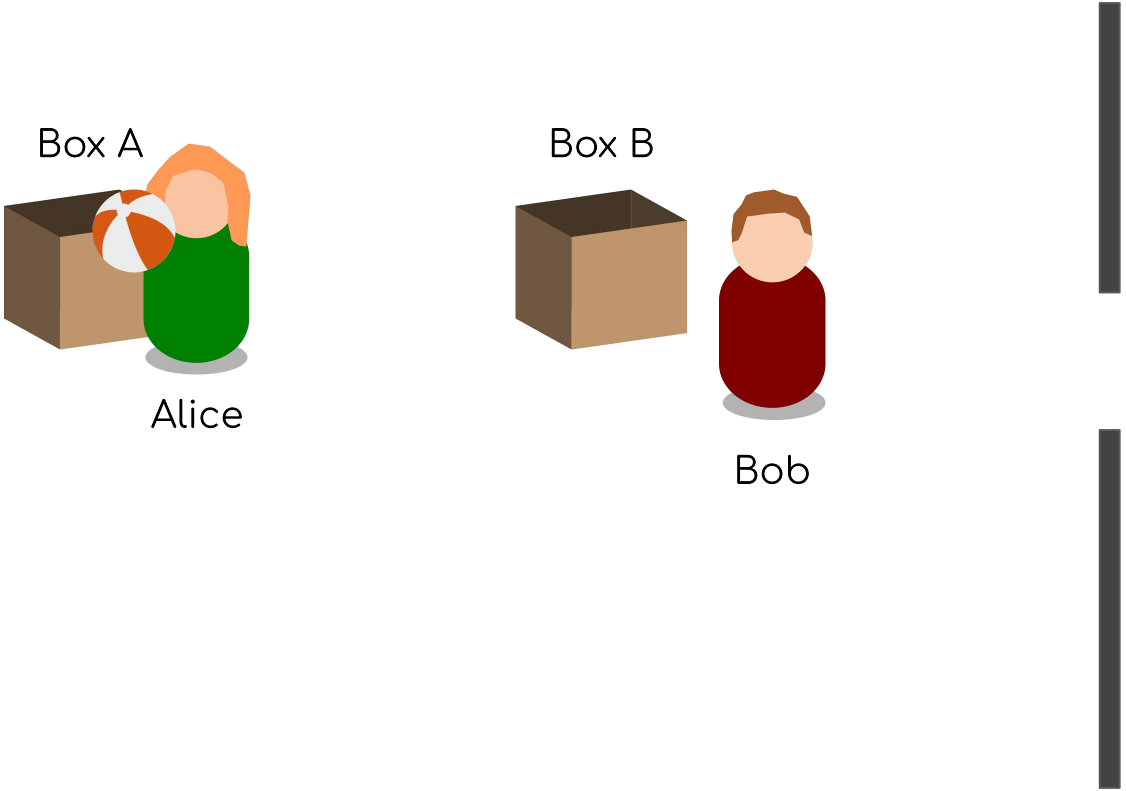
Introduction



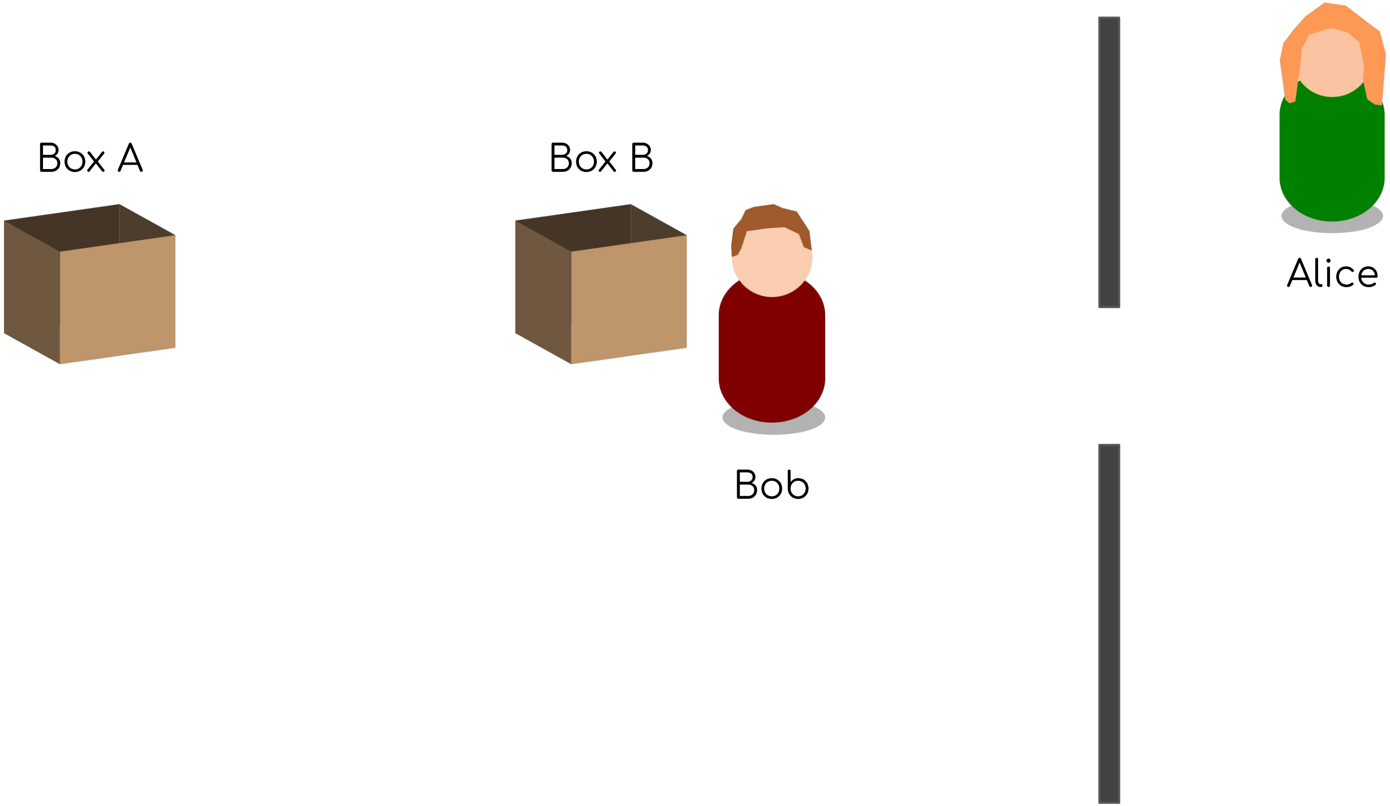
Introduction



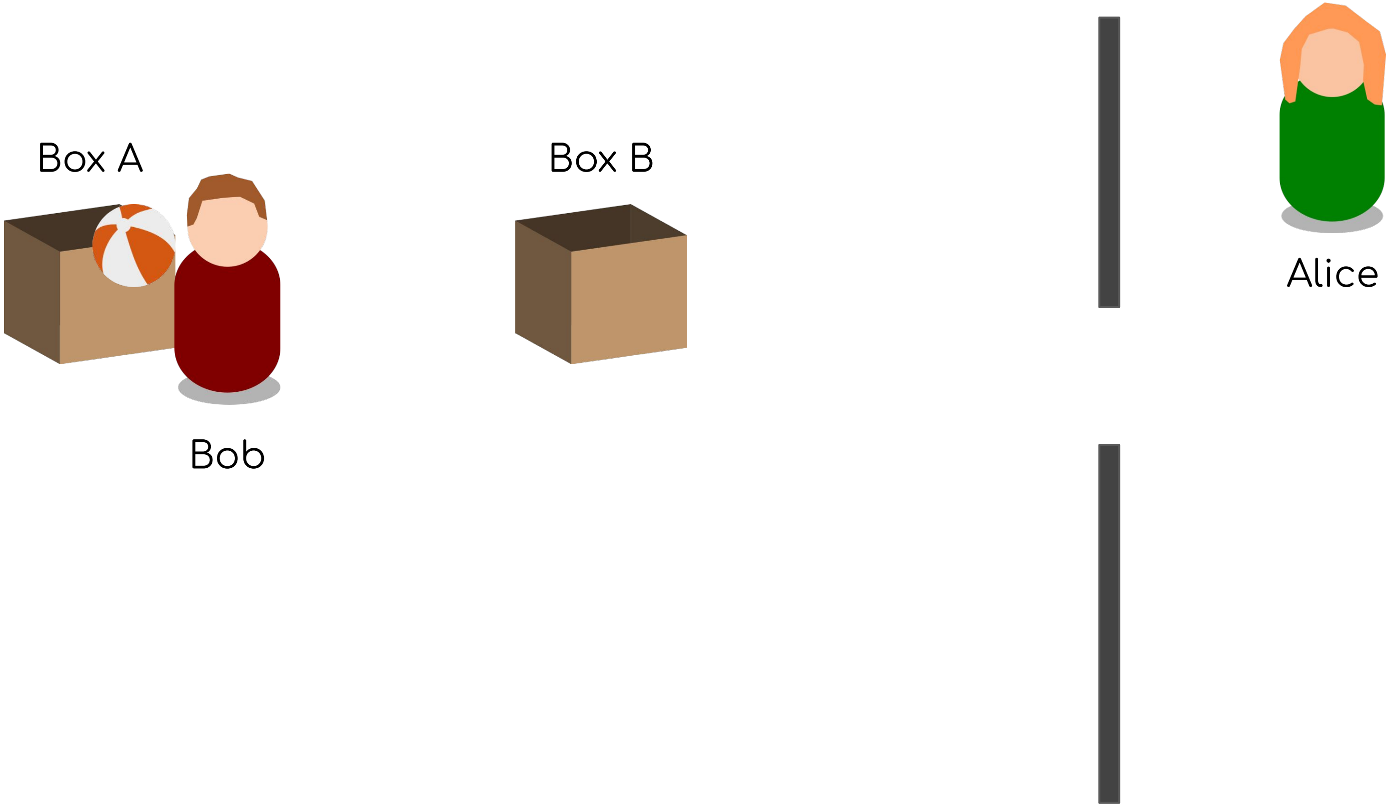
Introduction



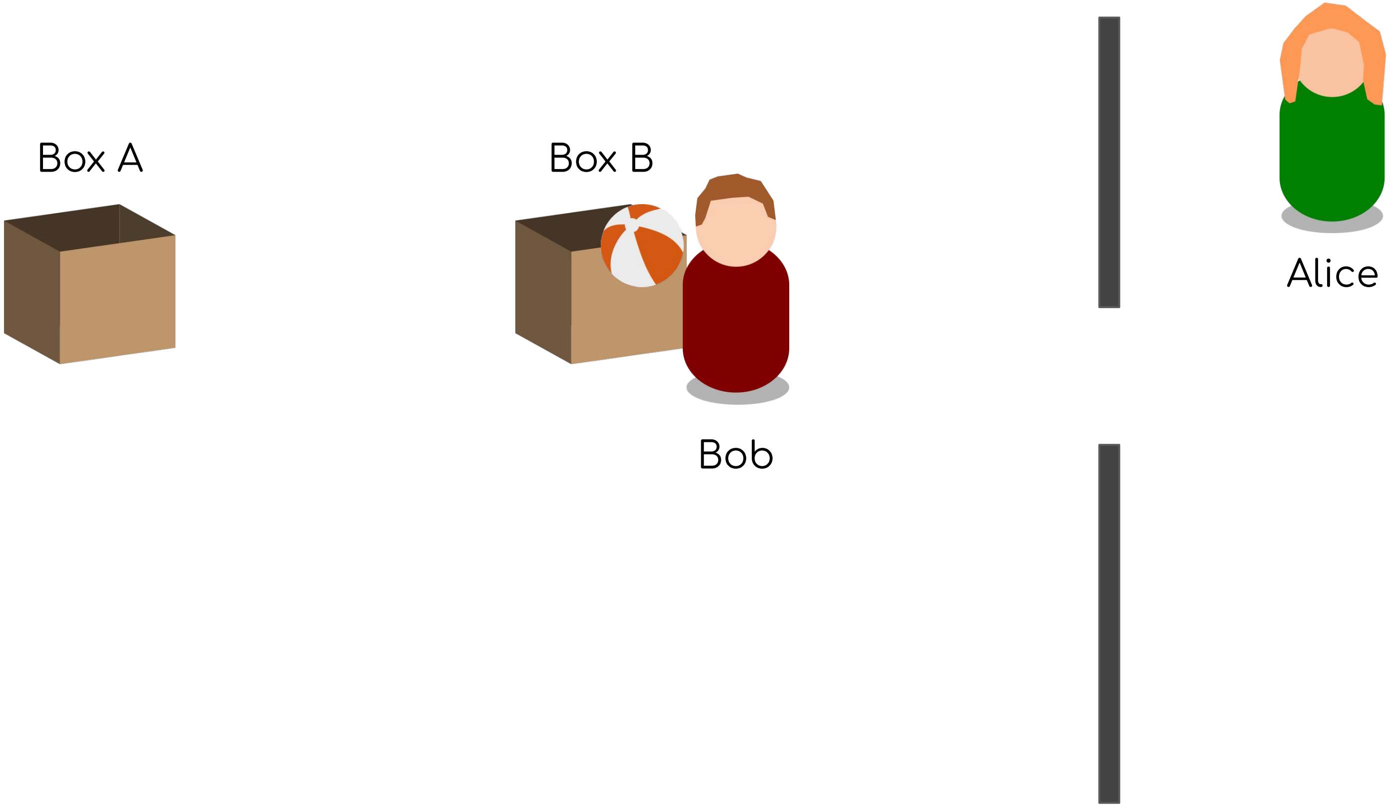
Introduction



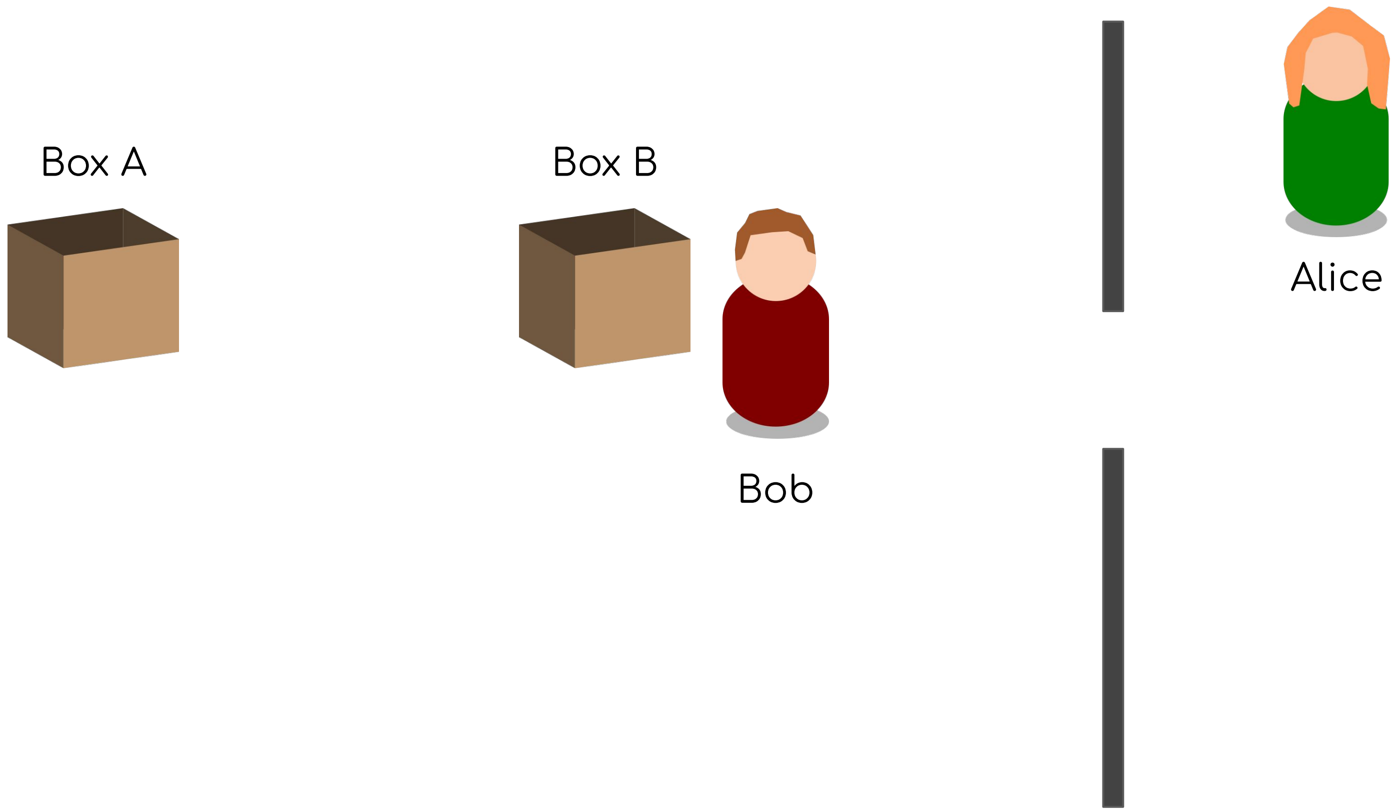
Introduction



Introduction

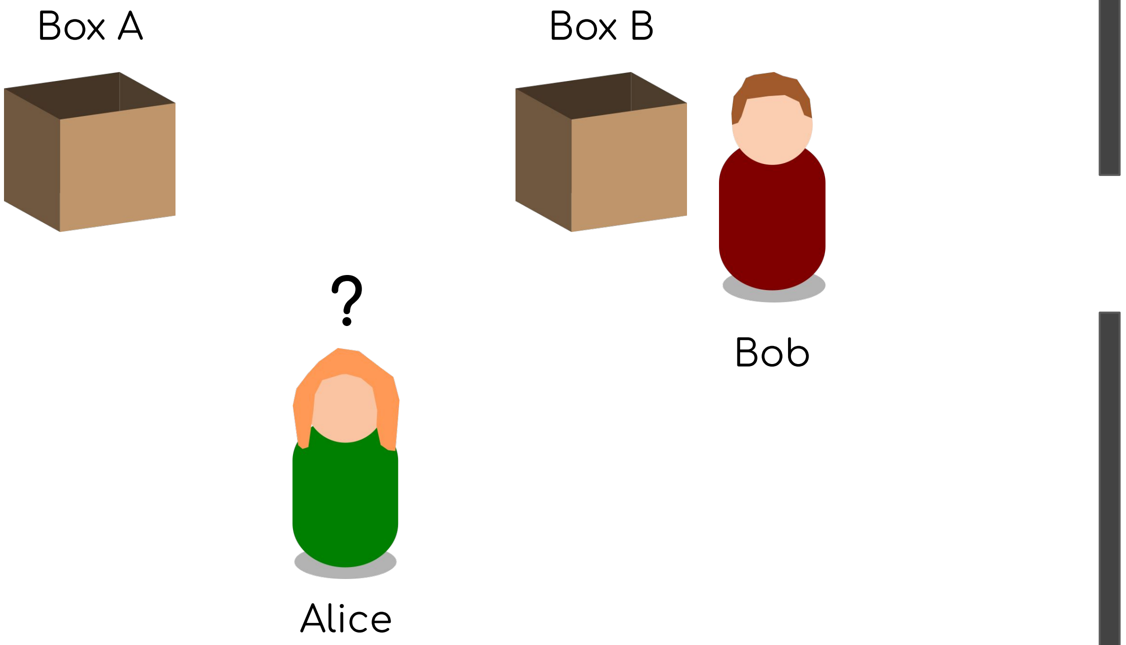


Introduction



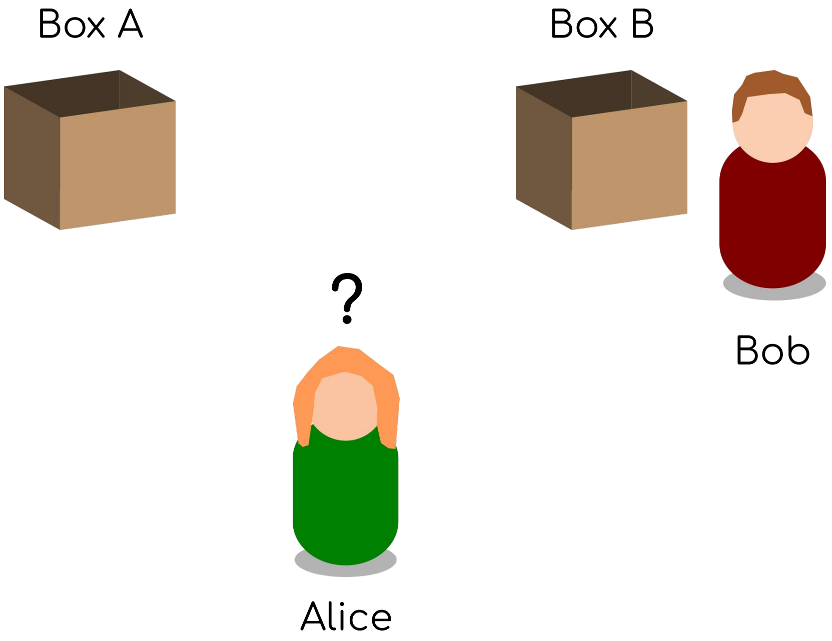
Introduction

Where will Alice look for the ball?



Introduction

Where will Alice look for the ball?



False-Belief Task

Tests for the existence of a theory of mind.

Theory of Mind

Ability to create internal models of others.

Problem Statement & Outline

> Current research paradigm uses *homo economicus* as a representation of human agents.
>
> *Homo economicus* uses Expected Utility Theory, which fails to describe human behavior [3,4,5,6].

We study the effects of cognitive biases and theory of mind on the emergence of coordination.

Ingredients

Theory of Value: Cumulative Prospect Theory
+
Theory of Mind: Level-k Bounded Rationality
+
Coordination Game: Stag Hunt

Experiments

Experiment I: Normal-form Game
+
Experiment II: Markov Game

Related Work

Prospect Theory is better at describing behavior than Expected Utility Theory.

- Kahneman, Tversky, 1979. *Prospect Theory: An Analysis of Decision Under Risk.*
- Fiegenbaum, 1990. *Prospect theory and the risk-return association.*
- Cxvi et al., 2001. *Prospect Theory and Asset Prices.*
- List, 2004. *Neoclassical theory versus prospect theory: Evidence from the marketplace.*
- Vis and Van Kersbergen, 2007. *Why and how do political actors pursue risky reforms?*
- Abdellaoui, Bleichrodt, Kammoun, 2013. *Do financial professionals behave according to prospect theory? An experimental study.*

Description of normal-form games with CPT value.

- Metzger, Rieger, 2019, *Non-cooperative games with prospect theory players and dominated strategies.*

Theory of mind helps coordination among agents using EUT in a sequential 2-player Stag Hunt.

- Yoshida, Dolan & Friston, 2008, *Game Theory of Mind.*

Definition and solution of Markov Decision Processes with CPT value function.

- Lin, Marcus, 2013, *Dynamic Programming with Non-Convex Risk-Sensitive Measures.*

Experiment I - Theory

Game: Agents + Actions + Information Structure + Reward Structure

Policy: Probability distribution over the action space.

Joint Policy: Vector of policies of all agents.

Normal-form Game

- N agents choose policies simultaneously.
- N agents receive a reward based on the chosen joint policy.

Solving a normal-form game: Finding the Nash Equilibria.

Nash Equilibrium: A joint policy from which no agent is better off by changing its individual policy.

Experiment I - Theory

Expected Utility Theory

(EUT)

Let $R(a)$ be a random variable representing the reward when choosing action a, $\text{support}\{R(a)\} = \{r_1, \ldots, r_n\}$, and $P(R(a) = r_i) = p_i$.
Let $u : \mathbb{R} \to \mathbb{R}$ be a utility function that transforms actual rewards into perceived utility.

$$V^{EUT}(a) = \sum_{i=1}^{n} u(r_i) p_i$$

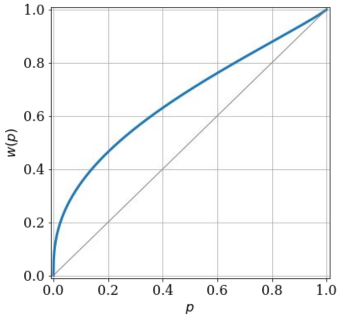
Experiment I - Theory

Cumulative Prospect Theory

(CPT) [3]

Let $R(a)$ be a random variable representing the reward when choosing action a,
$\text{support}\{R(a)\} = \{r_1, \ldots, r_n\}$, and $P(R(a) = r_i) = p_i$.
Let $u^+ : \mathbb{R} \to \mathbb{R}$ and $u^- : \mathbb{R} \to \mathbb{R}$ be the utility functions for **gains and losses**, respectively.
Let $w : [0,1] \to [0,1]$ be a probability weighting function that **transforms probability** into perceived probability.

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



$$(\underbrace{[r_1, p_1], \ldots, [r_k, p_k]}_{\text{Losses}}, \uparrow_{\text{Reference Point}} \underbrace{[r_{k+1}, p_{k+1}], \ldots, [r_n, p_n]}_{\text{Gains}})$$

$$V^{CPT}(a) = \sum_{i=k+1}^{n} u^+(r_i)[w(P(R(a) \geq r_i)) - w(P(R(a) > r_i))]$$
$$+ \sum_{i=0}^{k} u^-(r_i)[w(P(R(a) \geq r_i)) - w(P(R(a) > r_i))]$$

Experiment I - Model

Stag Hunt

- Two hunters go on a stag hunt.
- During the hunt, two hares are spotted.
- A decision is presented to both hunters:
  - To keep hunting the stag, or
  - To hunt a hare.

Hunting a Stag: High payoff, but risky (requires both hunters).
Hunting a Hare: Low payoff, but safe (can be hunted solo).

Represents a dilemma between the safety of a low payoff outcome and the risk of a high payoff outcome [7]

Experiment I - Results

Let $\pi_1 = (p_1, 1 - p_1)$ and $\pi_2 = (p_2, 1 - p_2)$ be the policies of agent 1 and 2, resp..

Game is symmetrical. Hence $p_1 = p_2 = p$.

Using the original weighting function $w(p) = e^{-0.5(-\log(p))^{0.9}}$, we obtain:

		2	
		Stag	**Hare**
1	**Stag**	**5,5**	**0,1**
	Hare	**1,0**	**1,1**

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

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

$p_{EUT} = 0.2$

$p_{CPT} = 0.028$

Coordination increases

Experiment II - Theory

Markov Decision Process

- State Space S
- Action Space A
- Transition Probability Function $p : S \times A \times S \to [0,1]$
- Reward Function $r : S \times A \to \mathbb{R}$
- Discount Factor $\beta \in (0,1)$

Solution of a MDP
Find a policy $\pi : S \times A \to [0,1]$ that maximizes a value functional $V(s,\pi)$, over the state space S.

Optimal Value

$$V^*(s) = \max_{\pi} V(s,\pi), \forall s \in S$$

Optimal Policy

$$\pi^* = \operatorname*{argmax}_{\pi} V(s,\pi), \forall s \in S$$

Experiment II - Theory

EUT Infinite-horizon Value Functional

$$V^{\mathrm{EUT}}(s,\pi) = \mathbb{E}_{s_{t+1}\sim p(\cdot|s_t,\pi(s_t))}\left[\sum_{t=0}^{\infty}\beta^t r(s_t,\pi(s_t))\middle| s_0 = s\right]$$, solved using **Dynamic Programming** [8].

$$V^{\mathrm{EUT}}(s,\pi) = r(s,\pi(s)) + \beta\sum_{s'\in S} V^{\mathrm{EUT}}(s',\pi)p(s'|s,\pi(s))$$

Experiment II - Theory

Markov Game [10]

- State Spaces S_i
- Action Spaces A_i
- Transition Probability Functions $p_i : \boldsymbol{S} \times A_i \times \boldsymbol{S} \to [0,1]$
- Reward Function $r : \boldsymbol{S} \times \boldsymbol{A} \times \boldsymbol{S} \to \mathbb{R}^N$
- Discount Factors $\beta_i \in (0,1)$

MG = MDP + Agents

Solution of a Markov Game
Each agent i finds a policy $\pi_i : \boldsymbol{S} \times A_i \to [0,1]$ that maximizes his value functional $V_i(\boldsymbol{s}, \pi_i, \boldsymbol{\pi}_{-i})$, over the joint state space $\boldsymbol{S} = S_1 \times \ldots \times S_N$.

Optimal Value

$$V_i^*(\boldsymbol{s}, \boldsymbol{\pi}_{-i}) = \max_{\pi} V_i(\boldsymbol{s}, \pi, \boldsymbol{\pi}_{-i}), \forall \boldsymbol{s} \in \boldsymbol{S}$$

Optimal Policy

$$\pi_i^*(\boldsymbol{\pi}_{-i}) = \underset{\pi}{\mathrm{argmax}} V_i(\boldsymbol{s}, \pi, \boldsymbol{\pi}_{-i}), \forall \boldsymbol{s} \in \boldsymbol{S}$$

Experiment II - Theory

How does each agent calculate Value?

Given the joint policy of others,
MG is the same as a MDP.

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

How to "know" the joint policy of others?

with
$$\begin{cases} P_{i,s,+}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon) = \sum_{\boldsymbol{a}_{-i} \in A_{-i}(s)} P_s^{a_i,\boldsymbol{a}_{-i}} \left(u_i^+ \left((r_i(s) + \beta_i V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(S) - b_i)_+\right) > \epsilon\right) \boldsymbol{\pi}_{-i}(\boldsymbol{a}_{-i}|s) \\ P_{i,s,-}^{a_i,\boldsymbol{\pi}_{-i}}(\epsilon) = \sum_{\boldsymbol{a}_{-i} \in A_{-i}(s)} P_s^{a_i,\boldsymbol{a}_{-i}} \left(u_i^- \left((r_i(s) + \beta_i V_i^{\pi_i,\boldsymbol{\pi}_{-i}}(S) - b_i)_-\right) > \epsilon\right) \boldsymbol{\pi}_{-i}(\boldsymbol{a}_{-i}|s) \end{cases}$$

Experiment II - Theory

Level-k Bounded Rationality [11]

In a 2 agent scenario

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

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

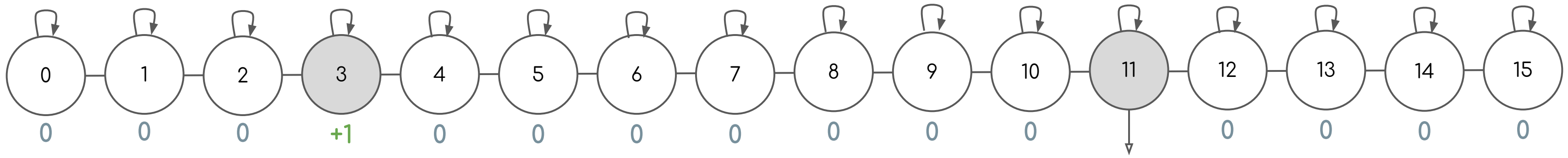
$$
\begin{array}{rcl}
& \pi_2^{(0)} \quad \pi_1^{(0)} & \\
& \downarrow \qquad \downarrow & \\
\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1, \pi_2^{(0)}}(s) = \pi_1^{(1)}(s) & & \pi_2^{(1)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(0)}, \pi_2}(s) \\
\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1, \pi_2^{(1)}}(s) = \pi_1^{(2)}(s) & & \pi_2^{(2)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(1)}, \pi_2}(s) \\
\vdots & & \vdots \\
\underset{\pi_1}{\operatorname{argmax}} V_1^{\pi_1, \pi_2^{(k-1)}}(s) = \pi_1^{(k)}(s) & & \pi_2^{(k)}(s) = \underset{\pi_2}{\operatorname{argmax}} V_2^{\pi_1^{(k-1)}, \pi_2}(s)
\end{array}
$$

Experiment II - Model

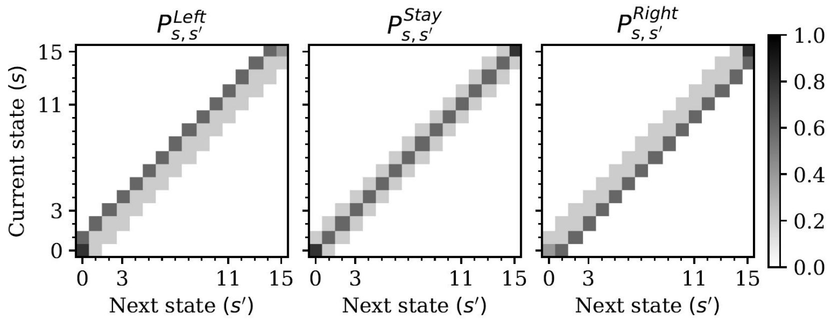
Markov Stag Hunt

Hares

Stags



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



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

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

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

Stereotype policies are assumed uniform for both agents.

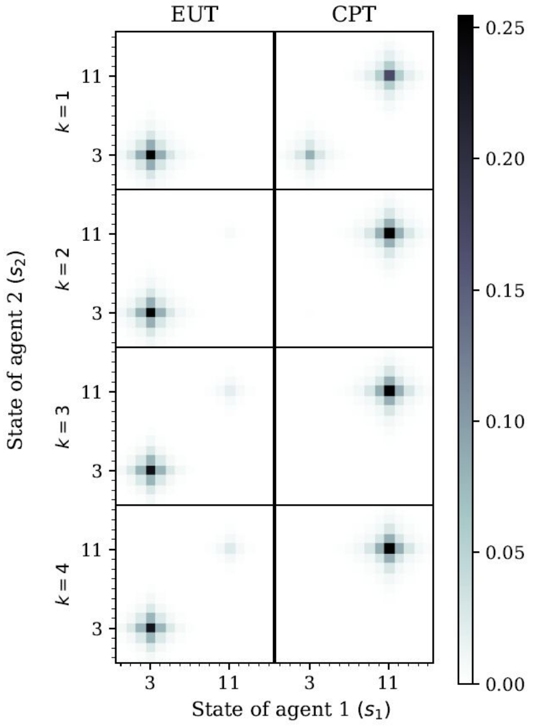
Experiment II - Results

Conditioned on the joint policy, a Markov Game becomes a Markov Chain.

The resulting Markov Chain is irreducible and aperiodic.

Markov Chain admits a stationary distribution.

Stationary Distribution of Agents



EUT-agents prefer hares.

CPT-agents prefer stags.

Preference for stags increases with sophistication level.

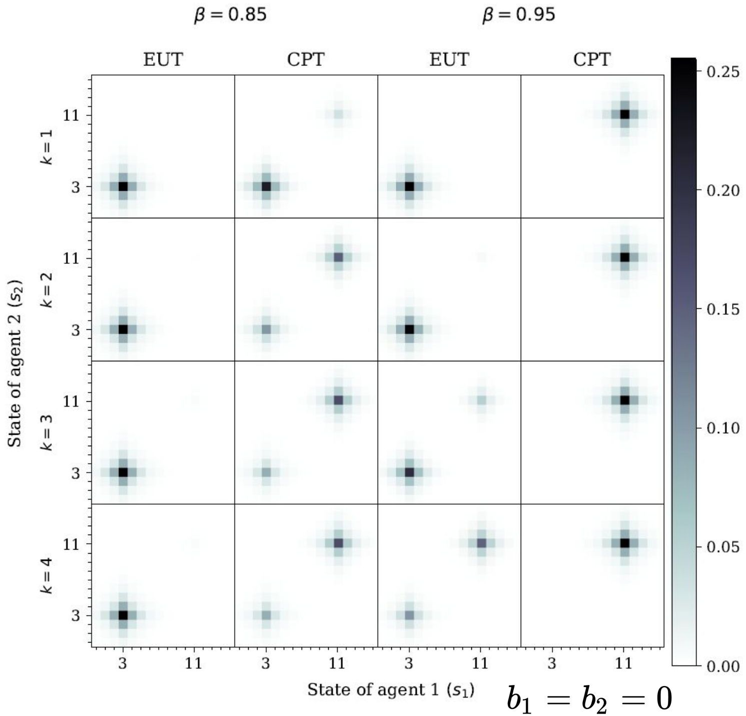
Sophistication levels higher than 3 do not change outcome.

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

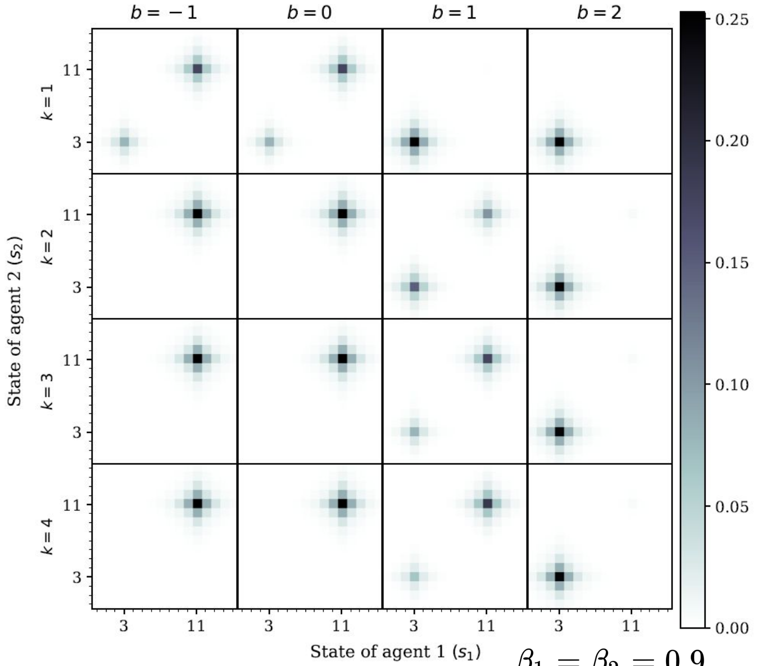
$$u(r) = r \quad w(p) = e^{-0.5(-\log(p))^{0.9}}$$

Experiment II - Results

Stationary Distribution of Agents



$b_1 = b_2 = 0$



$\beta_1 = \beta_2 = 0.9$

High discount factor increases coordination

High reference point decreases coordination

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

Conclusion

Main Contributions:

- Novel framework to study human interaction (Markov Game + CPT + Level-k).
- Equipping agents with CPT helps coordination.
- Increasingly sophisticated policies in the context of bounded rationality helps coordination.
- Higher sophistication levels than 3 do not change outcome.
- Preference of long-term over short-term rewards increases coordination.
- Optimistic perception of rewards increases coordination.

Future Work:

- Revisiting social conflict problems from a different perspective.
  - Climate Change Agreements as Public Goods Games
  - Diffusion of Responsibility Problems
- Optimization of value algorithm
- Efficient level-k theory of mind for N>2 agents.
- Experimental validation.

Humans are good at coordination may stem from the fact that we are cognitively biased to do so. Machine agents ought to be built to incorporate the cognitive biases of humans.

References

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

[2] A. Tversky and D. Kahneman, “The Framing of Decisions and the Psychology of Choice,” Science, vol. 211, no. 4481, pp. 453–458, 1981.

[3] A. Tversky and D. Kahneman, “Advances in Prospect Theory: Cumulative Representation of Uncertainty,” Tech. Rep., 1992.

[4] D. Kahneman and A. Tversky, “Prospect Theory: An Analysis of Decision Under Risk,” Econometrica, vol. 47, no. 2, pp. 263–291, 1979.

[5] M. Allais, “The Foundations of a Positive Theory of Choice Involving Risk and a Criticism of the Postulates and Axioms of the American School (1952),” Expected utility hypotheses and the Allais paradox, pp. 27–145, 1979.

[6] D. Ellsberg, “Risk, ambiguity and the Savage Axioms,” The quarterly journal of economics, pp. 643–669, 1961.

[7] B. Skyrms, The stag hunt and the evolution of social structure. Cambridge University Press, 2004.

[8] R. A. Howard, Dynamic Programming and Markov Processes, 1960.

[9] K. Lin and S. I. Marcus, “Dynamic Programming with Non-Convex Risk-Sensitive Measures,” in American Control Conference. Washington, DC, USA: IEEE, 2013, pp. 6778–6783.

[10] L. S. Shapley, “Stochastic Games,” in Proceedings of the National Academy of Sciences, 1953, pp. 1095–1100.

[11] D. O. Stahl, “Evolution of smartn players,” Games and Economic Behavior, vol. 5, pp. 604–617, 1993.

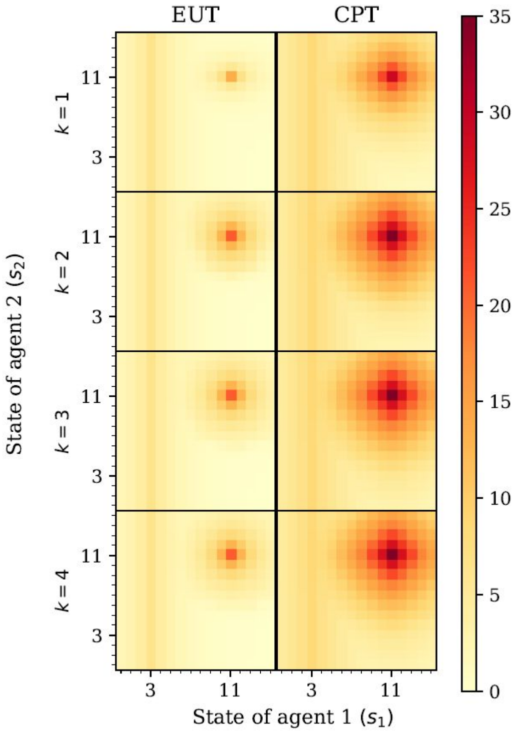
Thank You!

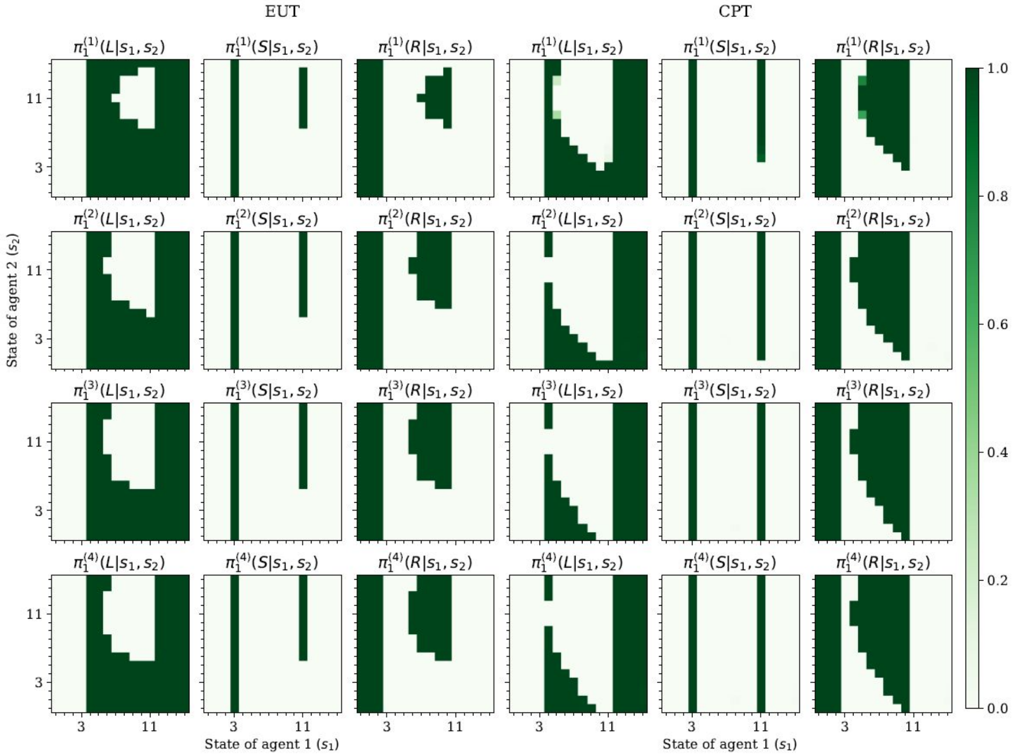
Experiment II - Value & Policies

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

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

Value of agent 1





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.

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,

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

EUT vs PT vs CPT Example

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

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

$$V^{EUT} \approx 2.85$$

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

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

$$V^{CPT} = \sum_{k=1}^{6} u(k)[w(P(R \geq k)) - w(P(R > k))]$$
$$= u(1)(w(1)-w(5/6)) + u(2)(w(5/6)-w(4/6)) + u(3)(w(4/6)-w(3/6))$$
$$+ u(4)(w(3/6)-w(2/6)) + u(5)(w(2/6)-w(1/6)) + u(6)(w(1/6)-w(0))$$

$$V^{CPT} \approx 3.48$$

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

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	**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	**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	**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**.

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

		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:

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

		2	
		Stag	**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**.

		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	**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**.

		2	
		Stag	**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	$\boxed{\underline{10},\underline{10}}$	0,2
	Hare	2,0	$\boxed{\underline{2},\underline{2}}$

What is the Nash Equilibrium here?

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

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

statistical reaction function
(or quantal response function)

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

Joint p.d.f of player *i*'s preference shocks

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

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

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

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

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

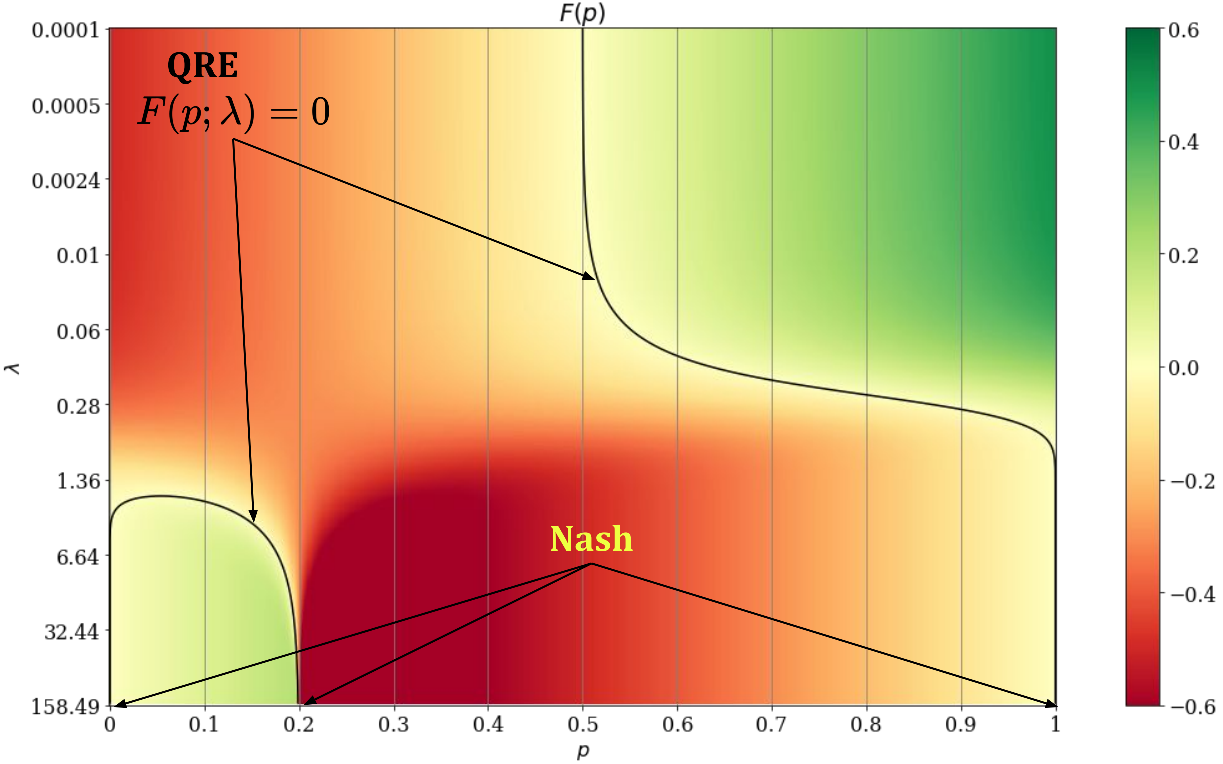
1 \ 2	Stag	Hare
Stag	**10,10**	**0,2**
Hare	**2,0**	**2,2**

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

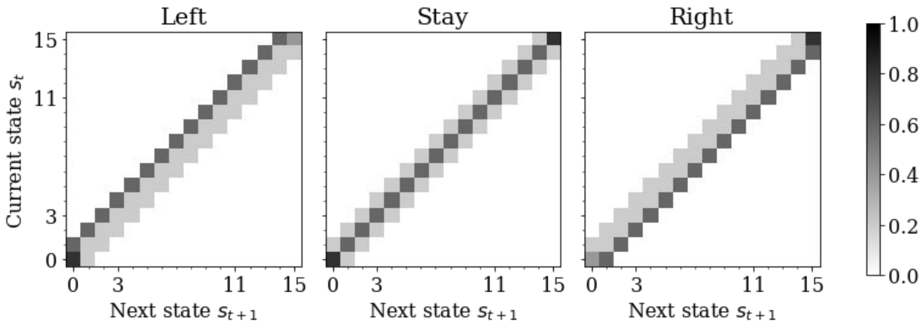
Finding the QRE means solving a transcendental equation.

QRE in Stag Hunt

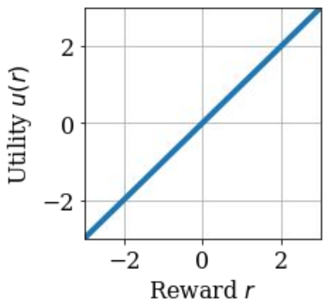
$$\text{QRE} \xrightarrow[\lambda\to\infty]{} \text{NE}$$



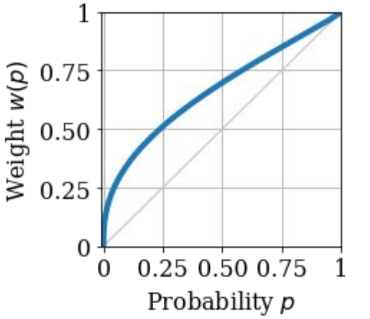
MDP with CPT Example

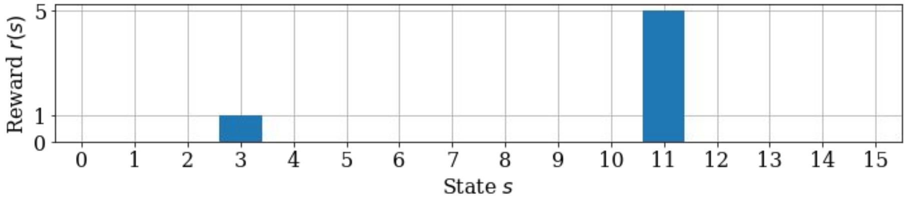


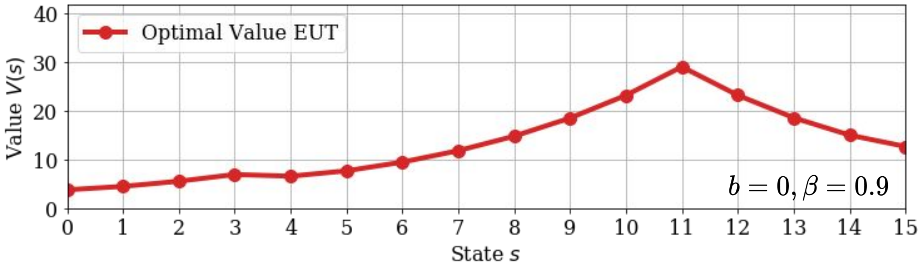
$u(r) = r$



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







MDP with CPT Example