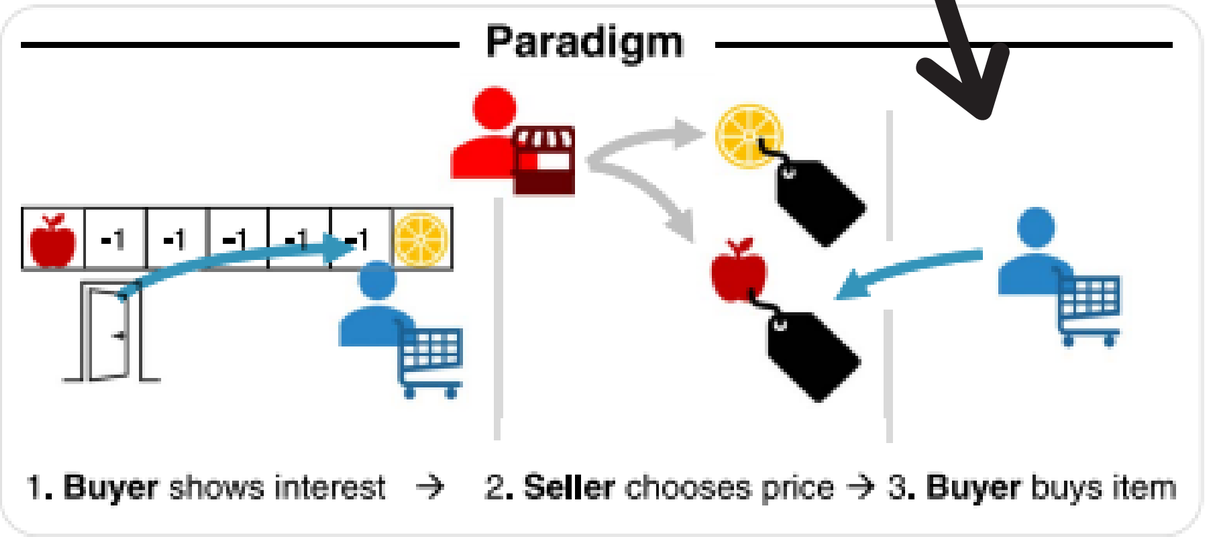
~~Today, We will Learn: How To Read and Manipulate People~~😈😈

Modelling Theory of Mind in Buyer-Seller Interactions

By: Ananya Gandhi
Guide: Prof. Peter Dayan
Indian Institute of Science Education and Research, Bhopal
Universitat Konstanz, Max Plank Institute

Motivation

How do humans anticipate others' thoughts and adjust decisions accordingly?



Paper

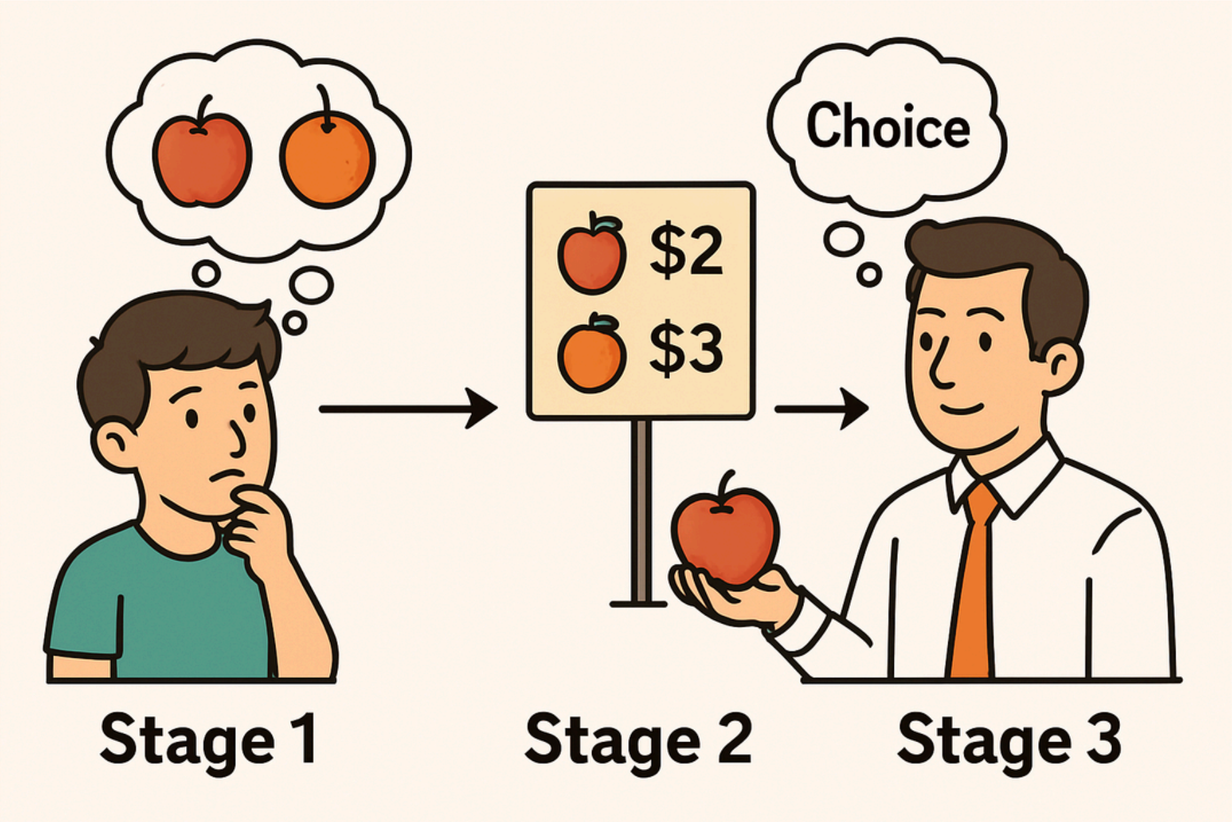
A (Dis-)information Theory of Revealed and Unrevealed Preferences: Emerging Deception and Skepticism via Theory of Mind

By:- Nitay Alon1 , Lion Schulz , Jeffrey S. Rosenschein , and Peter Dayan

Goal: Implement the Model ToM (Theory of Mind) in buyer-seller exchanges.

Problem Statement

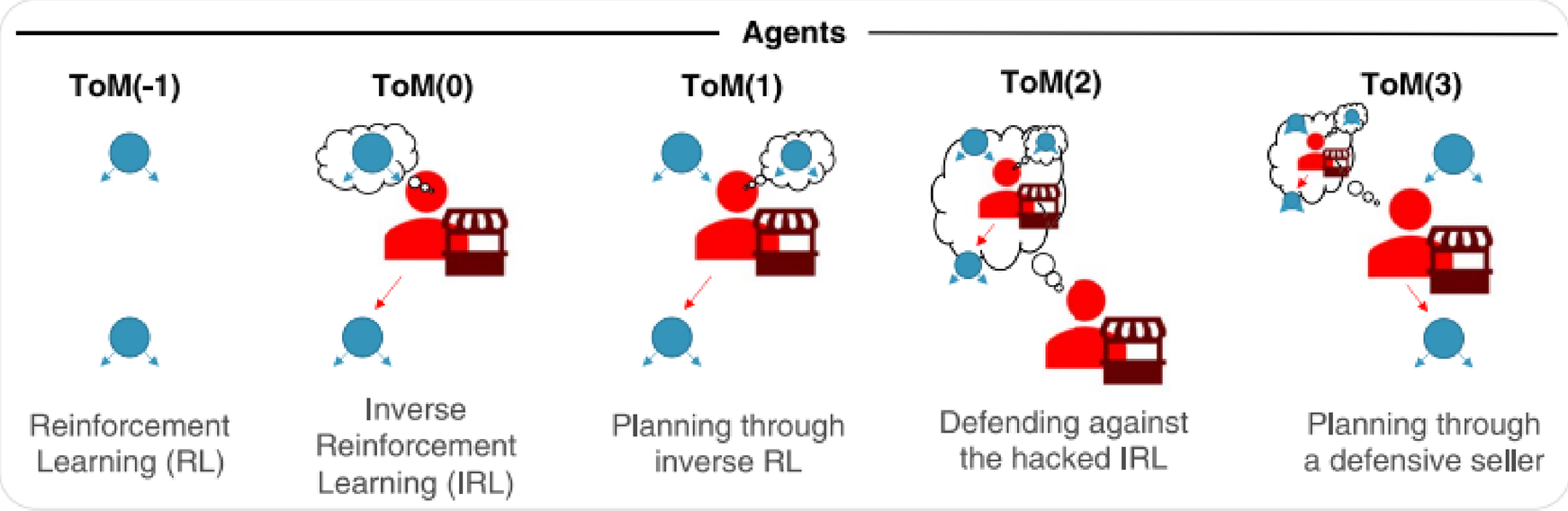
- Two agents:- Buyer and Seller



- Stage 1: Buyer chooses between two items (apple/orange) based on reward and distance
- Stage 2: Seller observes buyer's choice and sets prices for the final purchase
- Stage 3: Buyer makes final purchase decision

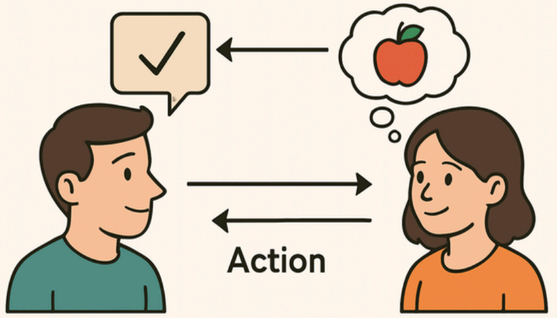
Theory of Mind Levels

ToM Level	Agent	Meaning in Buyer-Seller Context
-1	Buyer	Acts purely on own preferences and costs. Ignores how seller might respond.
0	Seller	Observes buyer's choice and assumes it reflects true preferences (assumes buyer is ToM(-1)).
1	Buyer	Anticipates seller's inference and may act strategically to hide or reveal preferences.
2	Seller	Assumes buyer is strategic (ToM(1)) and tries to outsmart potential deception.



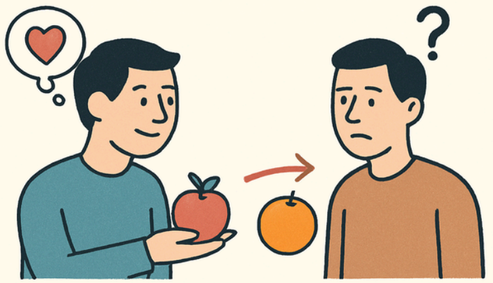
Information-Theoretic Metrics

- **Mutual Information (I(r, a1)):** How much does the buyer's action reveal about their preferences?
- **Skepticism :-KL Divergence (DKL(p(r|a1) || p(r))):** How much does the seller's belief change after observing the buyer?
- **Deception:- Policy KL Divergence:** How different is the actual buyer's policy from what the seller expects?



MUTUAL INFORMATION

DECEPTION



Buyer hides true preference
by choosing unexpectedly

SKEPTICISM

Seller remains unsure about
buyer's true preferences

Simulation SetUp

Parameter Sweeps

- Vary key parameters to explore model behaviour:
  - distance_apple (0 to 10): How far the apple is from the buyer
  - reward_apple (0 to 10): How much the buyer values the apple
- For each value, set distance_orange = 10 - distance_apple and reward_orange = 10 - reward_apple (to keep total constant)

Agent Sophistication

- Set Theory of Mind (ToM) levels:
  - Buyer ToM (k_b): e.g., 3 (strategic)
  - Seller ToM (k_s): e.g., 2 (strategic)

Randomisation / Robustness

- Add noise to preferences and distances to test robustness of results

Outputs Collected

- Probability buyer chooses apple or orange
- Seller's optimal price and expected utility
- Information-theoretic metrics (mutual information, KL divergence)

Visualisation

- Line, violin, scatter, bar, and heat map plots to show how outcomes change with parameters

Main Equation Used:

1. Soft-max Policy (Buyer's Choice Probability)

$P(a) = \exp(\beta \cdot Q(a)) / \Sigma_{a'} \exp(\beta \cdot Q(a'))$

- $P(a)$: Probability of choosing action a (e.g., apple or orange)
- $Q(a)$: Utility of action a
- β: Inverse temperature (higher = more deterministic, lower = more random)

2. Buyer Utility

- Stage 1 (before price):

$Q(a) = \text{reward}_a - \text{distance}_a$

- Stage 3 (after price):

$Q(a) = \text{reward}_a - \text{price}_a$

3. Seller's Inference (Bayesian Update)

$p(r \mid a_1) \propto p(a_1 \mid r) \cdot p(r)$

- $p(r \mid a_1)$: Posterior over buyer preferences after observing action a_1
- $p(a_1 \mid r)$: Likelihood of action given preference
- $p(r)$: Prior over preferences

4. Mutual Information

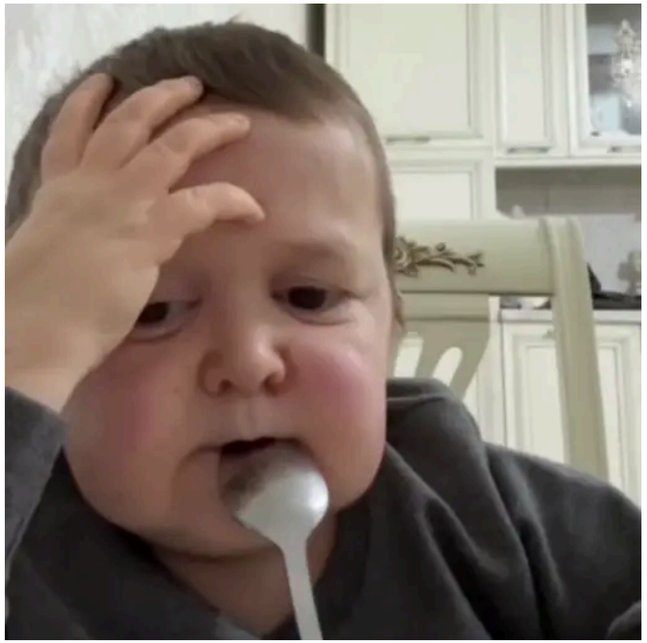
$I(r; a_1) = \Sigma_{(r,\, a1)}\, p(r, a_1) \cdot \log_2 [p(r, a_1) / (p(r) \cdot p(a_1))]$

5. KL Divergence (Seller Skepticism)

$D_{kl}(p(r \mid a_1) \,||\, p(r)) = \Sigma_r\, p(r \mid a_1) \cdot \log_2 [p(r \mid a_1) / p(r)]$

Test Results:

BUT!



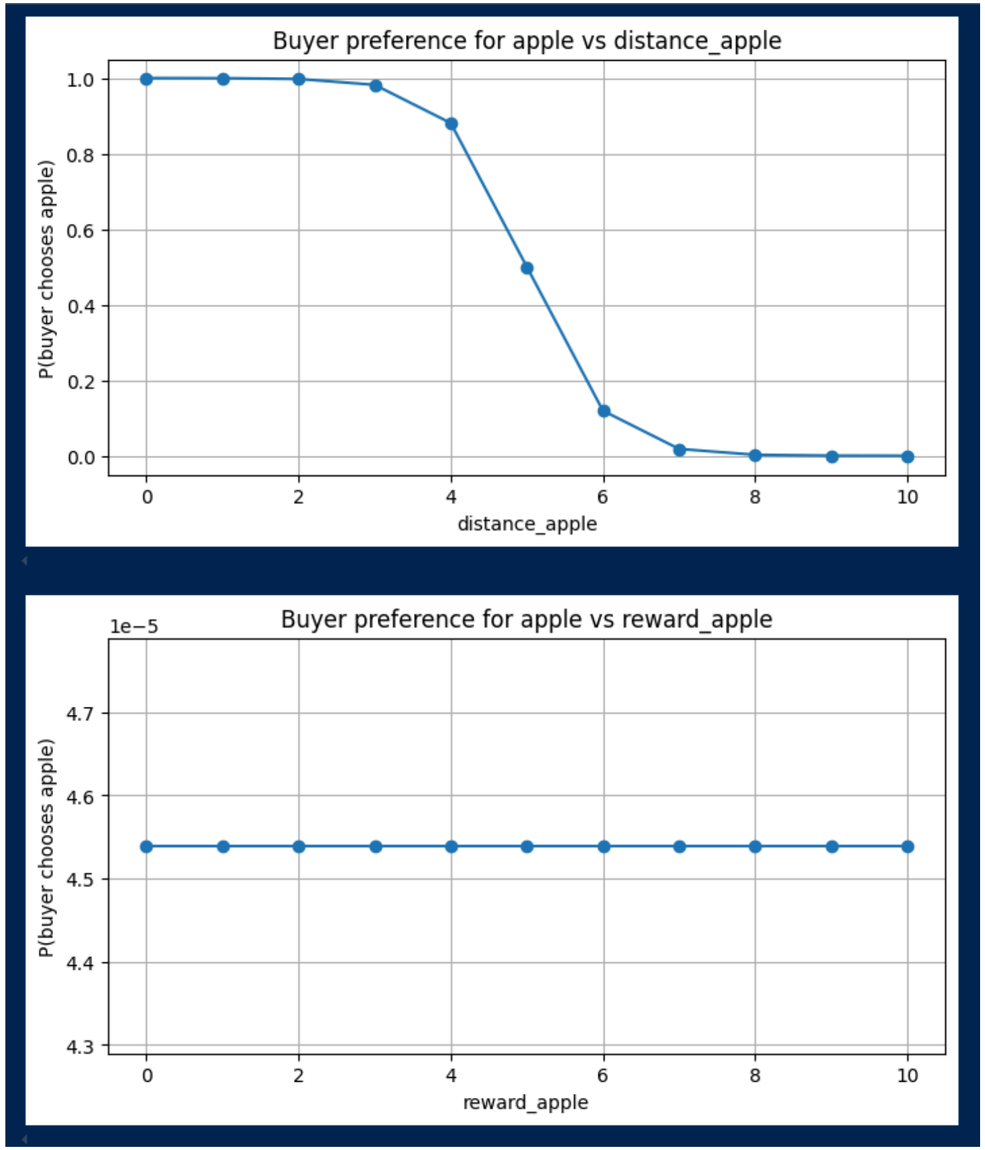
Before we move ahead....

At each buyer/seller alternation, we branch for each item (apple, orange) and for each possible preference. The number of recursive calls is roughly proportional to 2^(k_b + k_s) times the number of preferences at each level.

Practical usage:

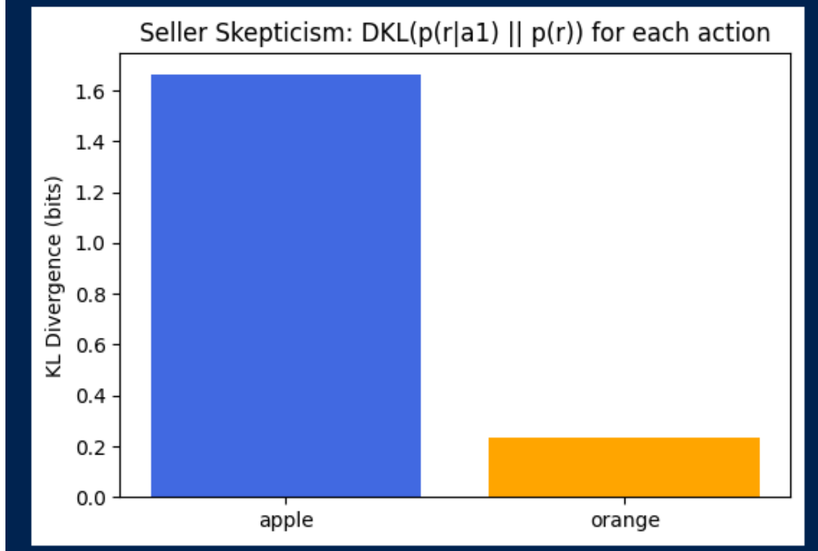
k_b = 1, k_s = 2: Runs in seconds.
k_b = 3, k_s = 2: Runs in seconds to minutes. ← For now, We stick to This.
k_b = 3, k_s = 4: May take minutes to hours.
k_b = 5, k_s = 4: Likely infeasible on a laptop; could take hours or days.

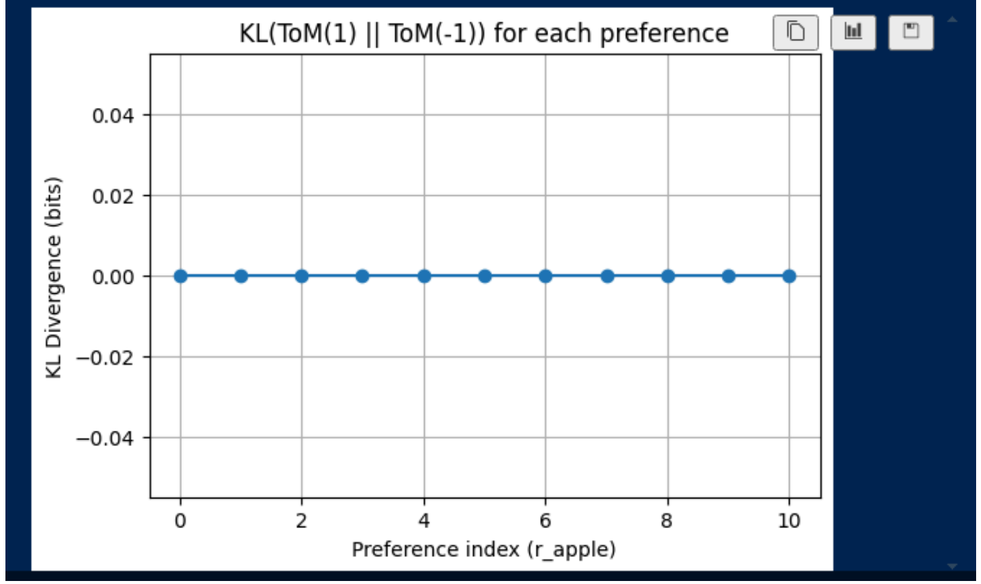
ToM(Buyer) = 3
ToM(Seller) = 2



The top plot shows that buyer preference for apple decreases sharply as the distance to apple increases — reflecting rational cost-sensitive behavior. In contrast, the bottom plot reveals that varying the reward for apple (with fixed distance) barely affects buyer choice when:
Beta = 1
Reward(apple) = 7
Distance(apple) = 3

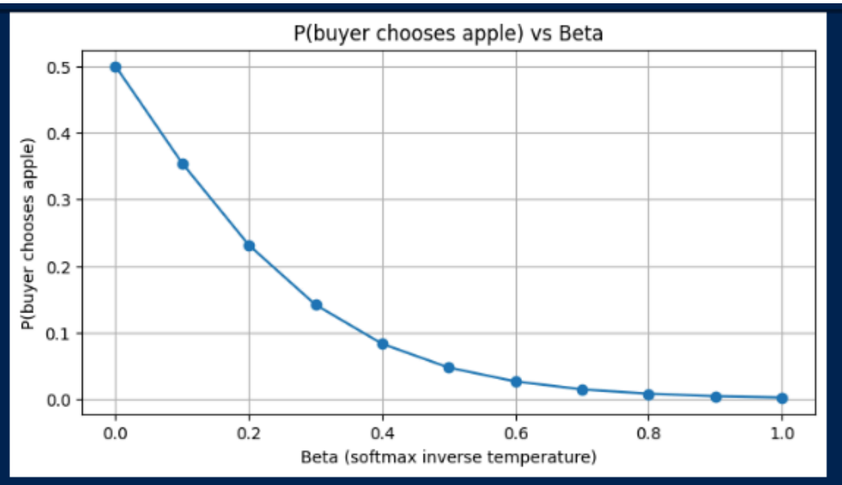
```
Mutual Information I(r, a1): 0.5606 bits
KL Divergence DKL(p(r|a1) || p(r)) for each action:
  apple: 1.6632 bits
  orange: 0.2358 bits
Mean KL divergence between ToM(1) and ToM(-1) buyer policies: 0.0000 bits
```

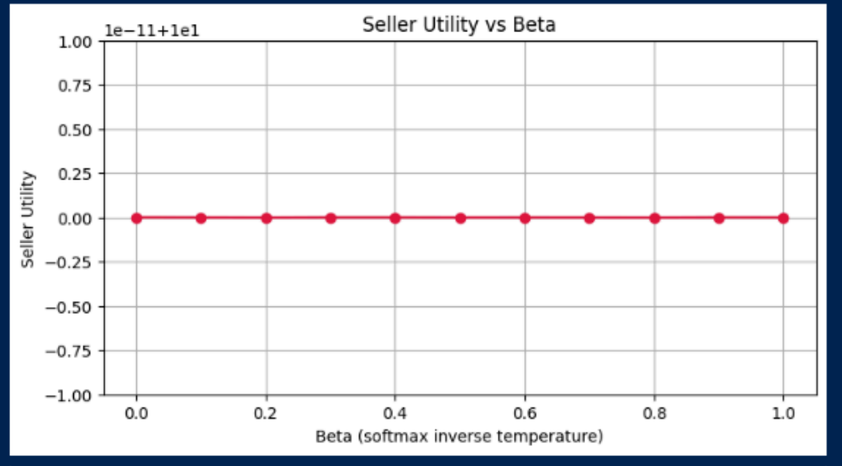




ToM(Buyer) = 3
ToM(Seller) = 2

Top panel shows how much the seller distrusts buyer actions for different items. The seller is very skeptical about "apple" signals (high bar) but trusts "orange" signals more (low bar). Bottom panel shows that buyer strategies remain stable over time (flat line near zero). The mutual information score (0.5606 bits) shows buyers partially hide their true preferences, but some information still leaks through their actions.





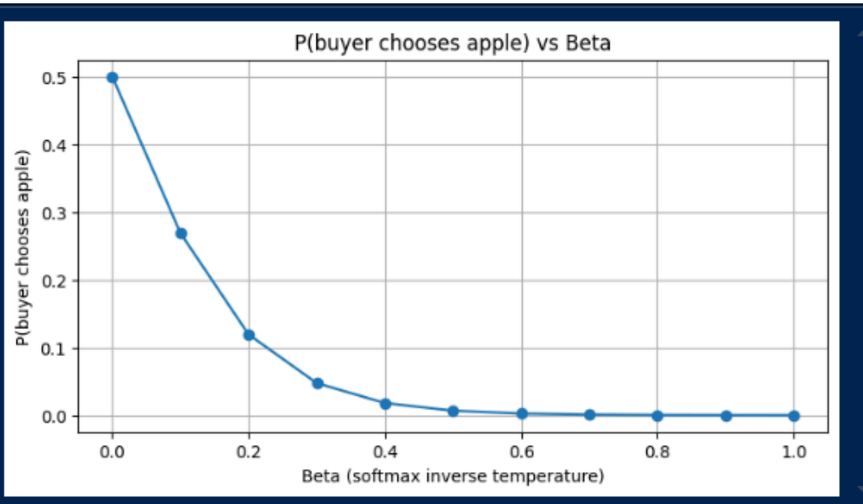
Theory of Mind levels

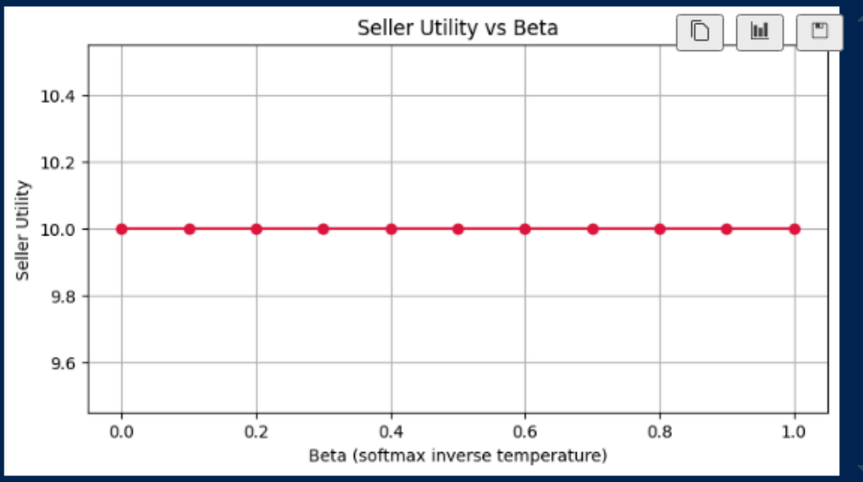
k_b = 3 # Theory of Mind level for buyer k_s = 2 # Theory of Mind level for seller

reward_apple = 6.0 # Reward for item i reward_orange = 10 - reward_apple # Reward for item j

distance_apple = 8.0 # Distance for item i distance_orange = 10 - distance_apple # Distance for item j

Steepening of the Curve : Probability of Buyer choosing item Apple as action 1 varying across different values of Beta, when reward_apple is increased from 6.0 to 7.0 and distance_apple is decreased from 8.0 to 2.0



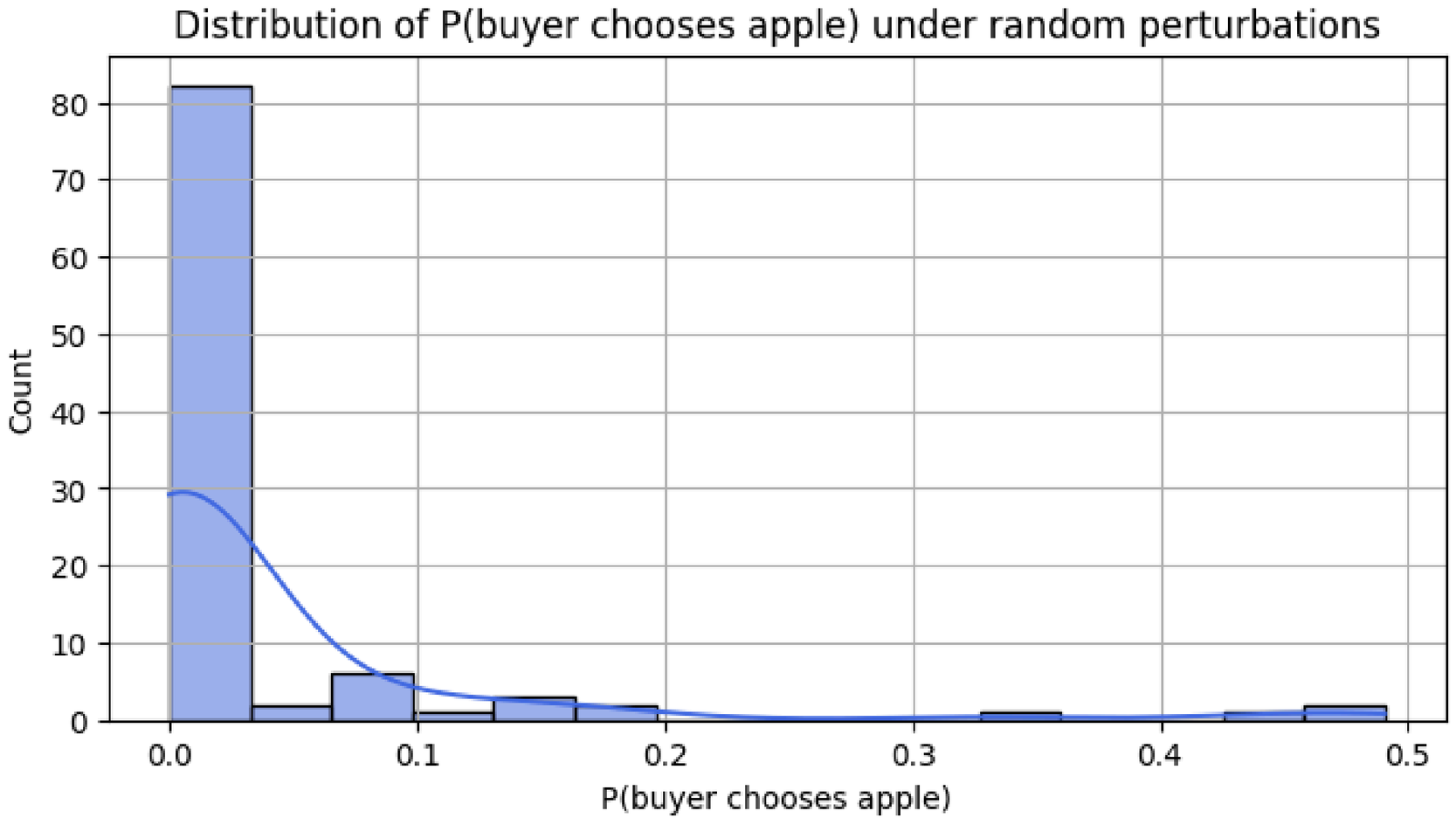


Generate + Code + Markdown

```
# Theory of Mind levels
k_b = 3  # Theory of Mind level for buyer
k_s = 2  # Theory of Mind level for seller

reward_apple = 7.0  # Reward for item i
reward_orange = 10 - reward_apple  # Reward for item j

distance_apple = 2.0  # Distance for item i
```

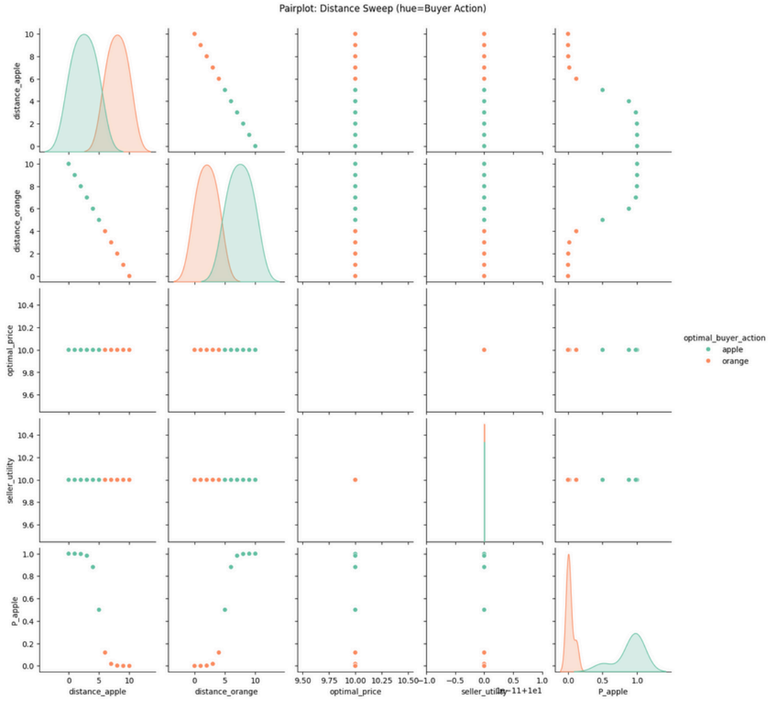


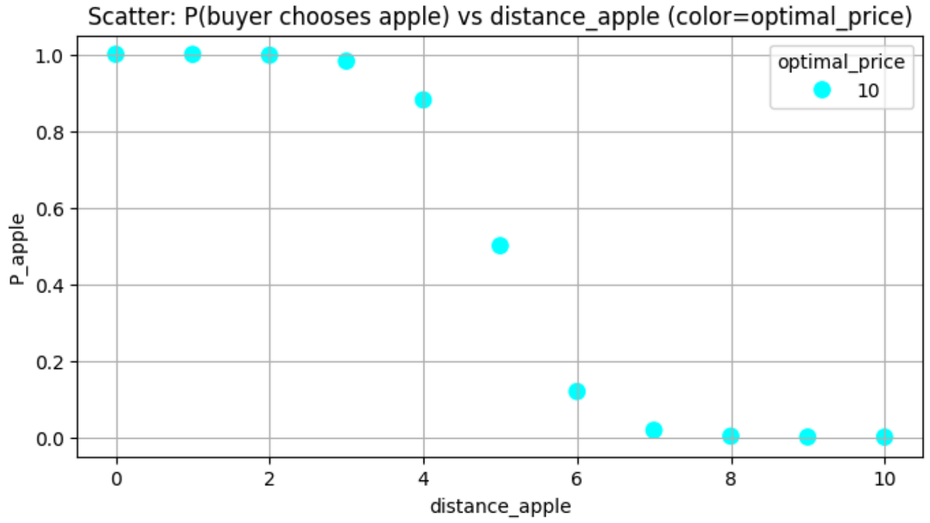
ToM(Buyer) = 3
ToM(Seller) = 2

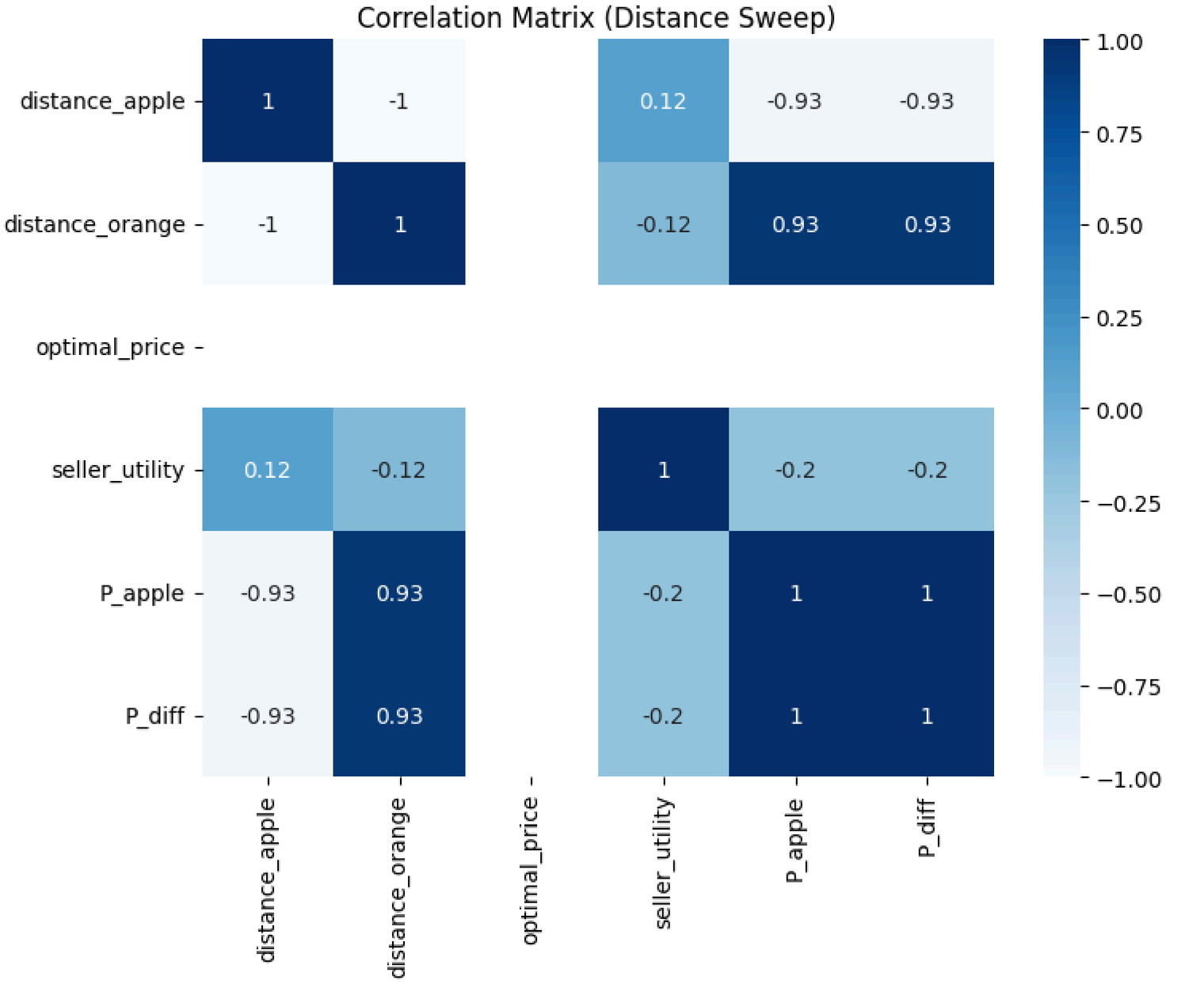
initial reward(apple) = 6
initial distance (apple) = 8

Most buyer choice probabilities for apple are near zero because the baseline strongly favours orange via higher reward or lower distance. Since rewards and distances sum to fixed totals, small perturbations rarely make apple more attractive. The buyer's nonlinear choice model then overwhelmingly selects orange.

Distribution of buyer's probability of choosing apple under random perturbations. Most of the time, the buyer strongly prefers orange, showing the model's sensitivity to underlying preferences and costs


Pairplot: Distance Sweep (hue=Buyer Action)
distance_apple
distance_orange
optimal_price
seller_utility
P_apple
optimal_buyer_action
apple
orange


Scatter: P(buyer chooses apple) vs distance_apple (color=optimal_price)
optimal_price
10
P_apple
distance_apple


Correlation Matrix (Distance Sweep)
distance_apple
distance_orange
optimal_price
seller_utility
P_apple
P_diff
1
-1
0.12
-0.93
-0.93
-1
1
-0.12
0.93
0.93
0.12
-0.12
1
-0.2
-0.2
-0.93
0.93
-0.2
1
1
-0.93
0.93
-0.2
1
1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Acknowledgements

- Ahmed, Christophe, Daniela
- Prof. Peter Dayan
- Everyone Here!