Mind Games

Peter Dayan

MPI for Biological Cybernetics

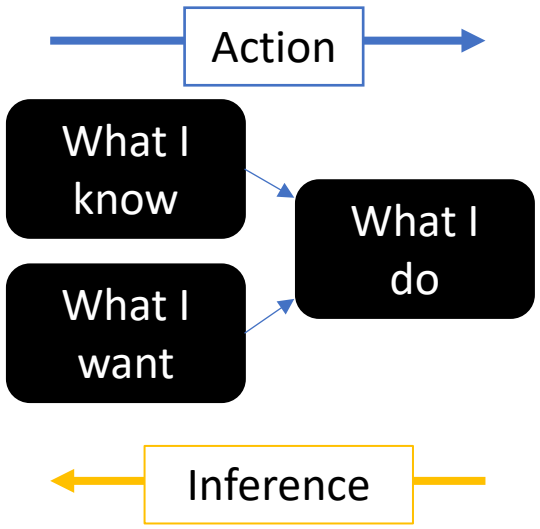
Andreas Hula Read Montague

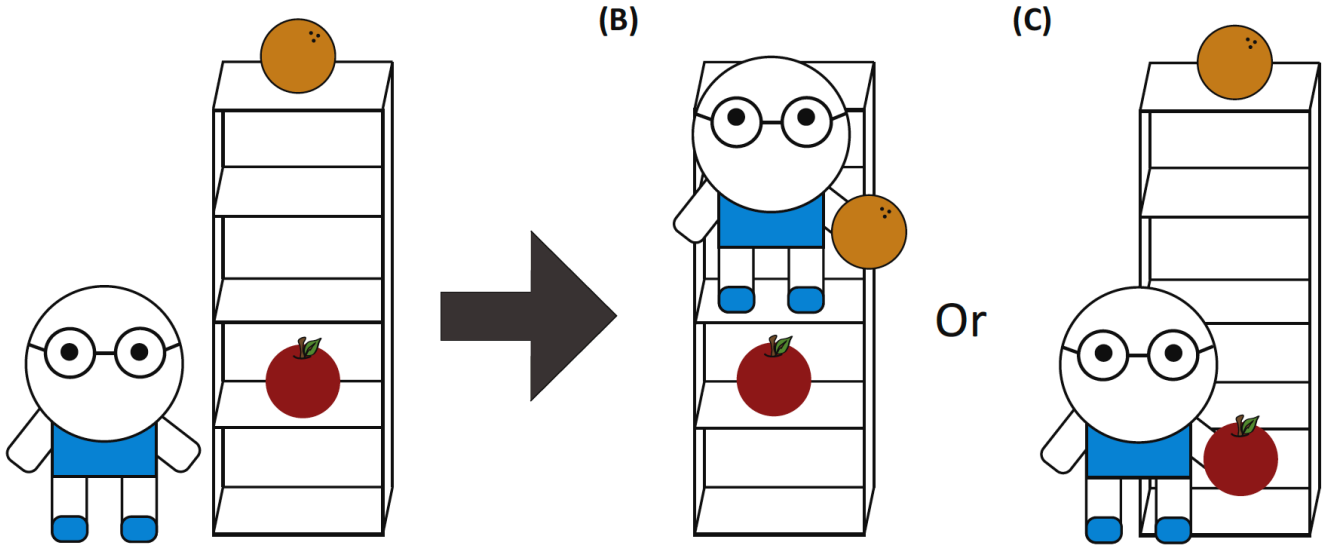
Nitay Alon Joe Barnby Lion Schulz Jeff Rosenchein

Plan

- theory of mind
- the p-beauty game
- recursive theory of mind in shopping
- recursive theory of mind in the ultimatum game
  - dynamic updating
  - defection

ToM as inverse reinforcement learning



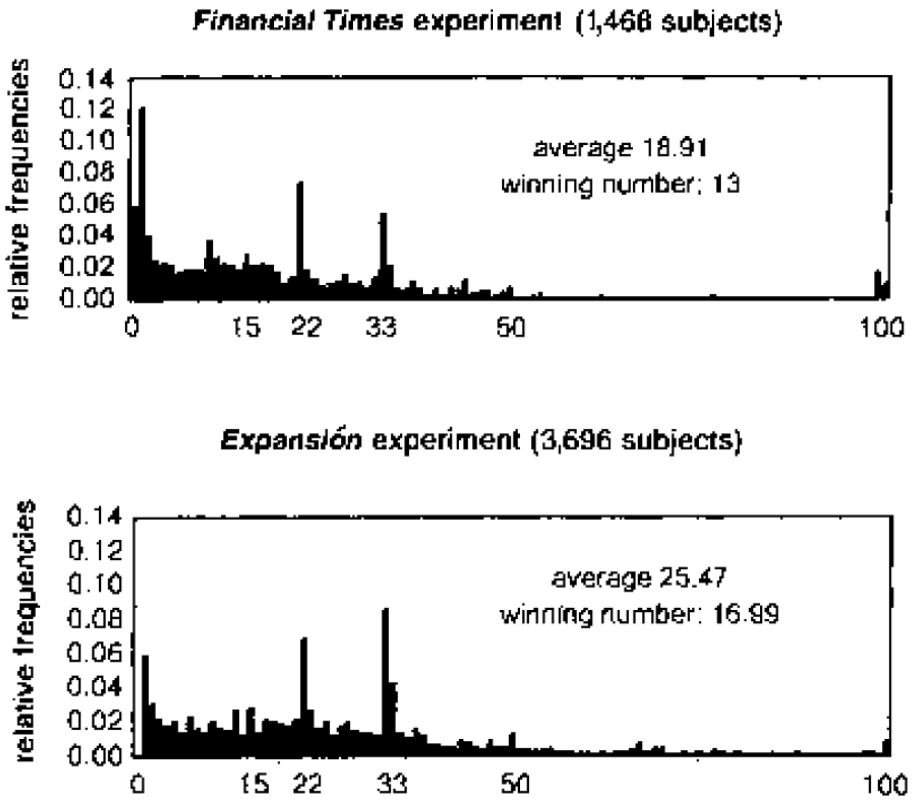


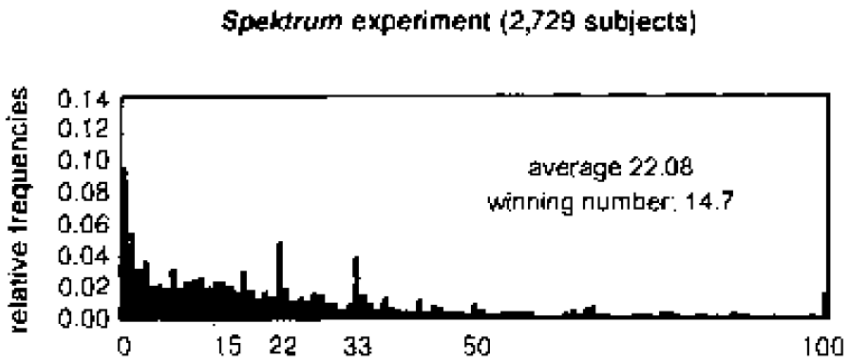
$$P(Rewards|Actions, Costs) \propto P(Action|Costs, Rewards)p(Rewards)$$

Jara-Ettinger et al. (2016), Baker et al. (2017), and many more

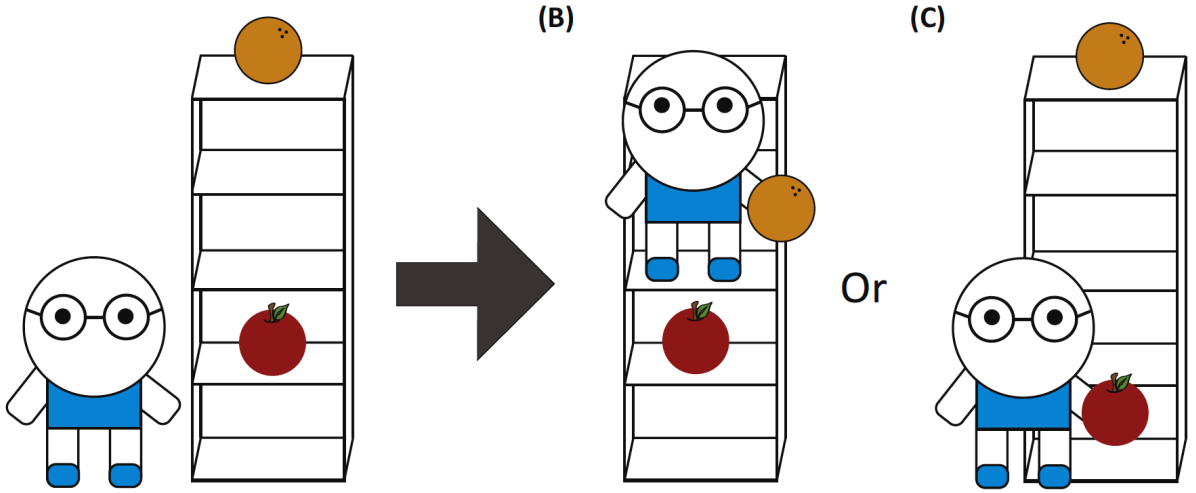
Recursive theory of mind: the p-Beauty Game

- Keynes; Nagel
  - choose number between 1-100
  - person closest to $^{2}/_{3}$rds of the mean wins prize



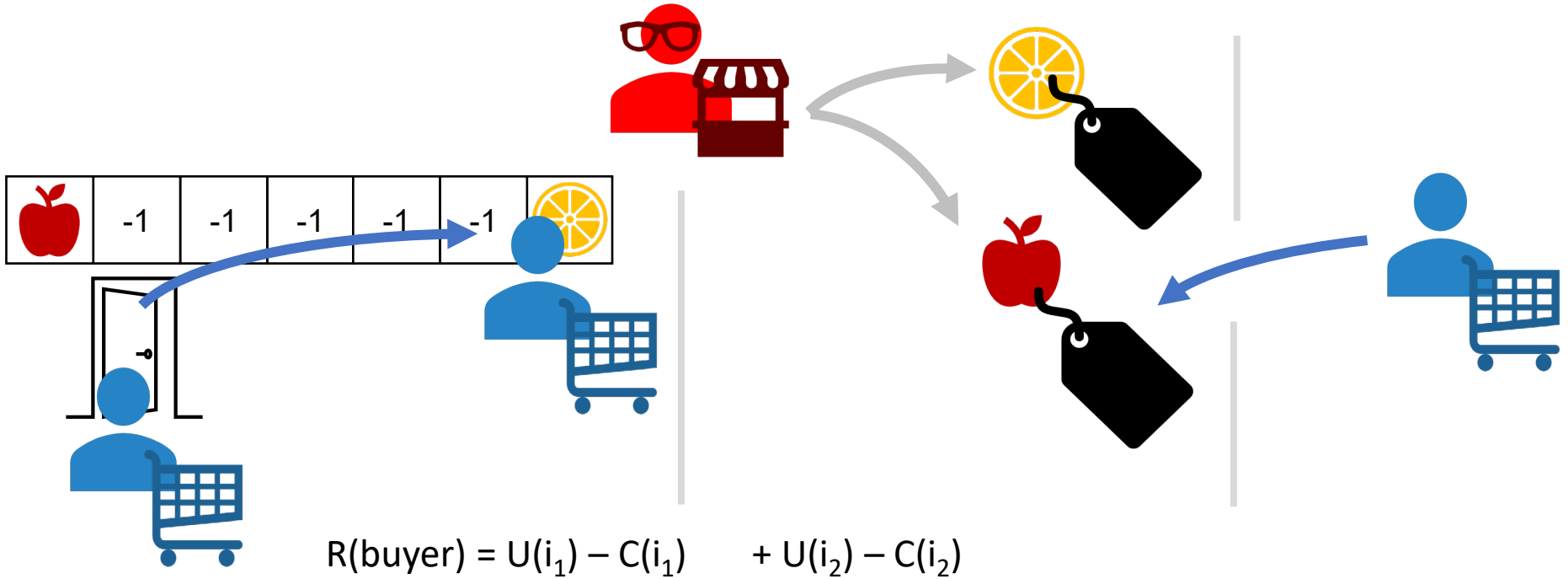


The problem setting - intuition



A three stage disinformation game

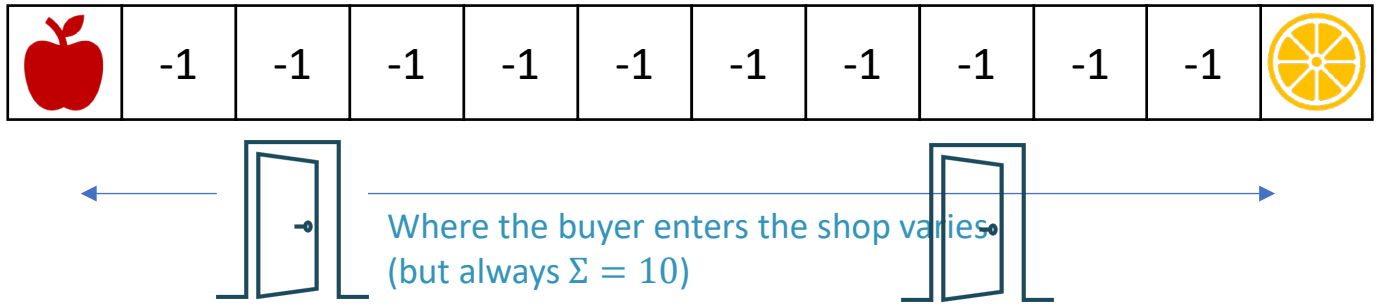
1. **Buyer** shows interest observed by **seller**
2. **Seller** chooses price
3. **Buyer** purchases item based on prices



$R(\text{buyer}) = U(i_1) - C(i_1) \quad + U(i_2) - C(i_2)$

$R(\text{seller}) = C(i_2)$

A three stage disinformation game



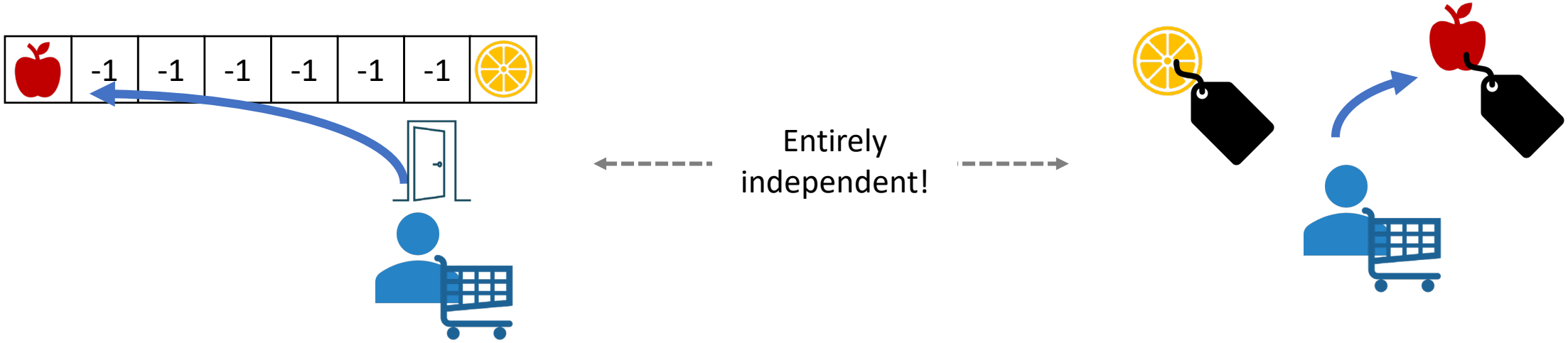
V( )	7	?
V( )	3	?

Flat prior

$\Sigma = 10$ $\Sigma = 10$

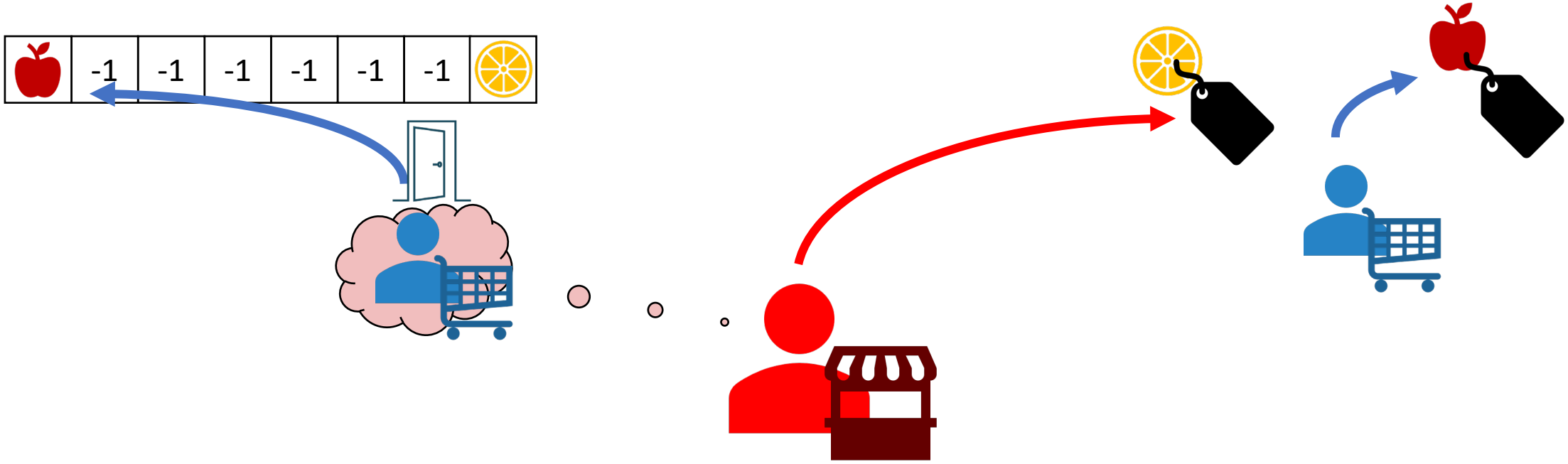
The theory of mind depths

DoM(-1) buyer: The naïve buyer



The theory of mind levels

DoM(0) seller: The inverse reinforcement learning seller
(a.k.a. the Naïve Utility Calculus)

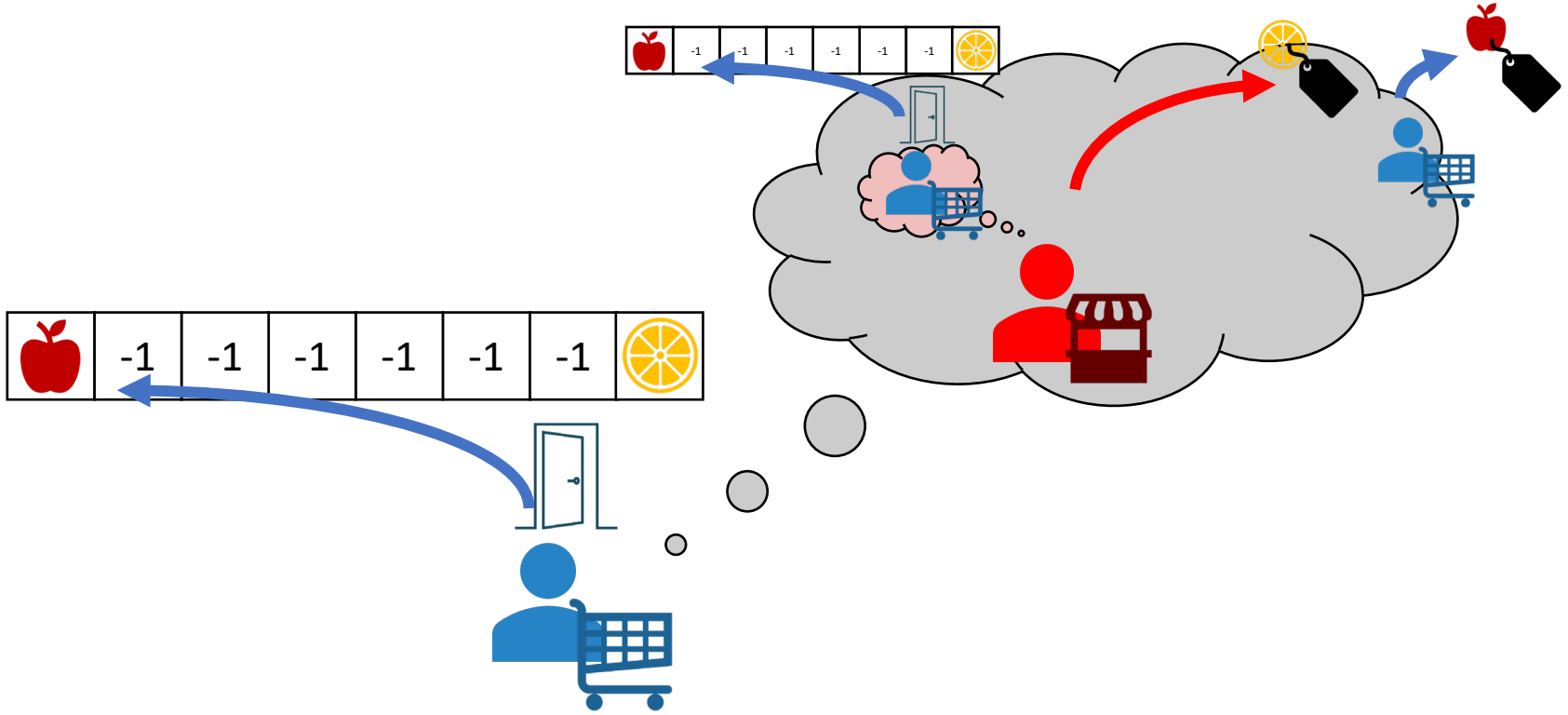


$P(Value|Action, Cost)$

The theory of mind levels

DoM(1) buyer: Hacking the inverse reinforcement learner

Observer awareness

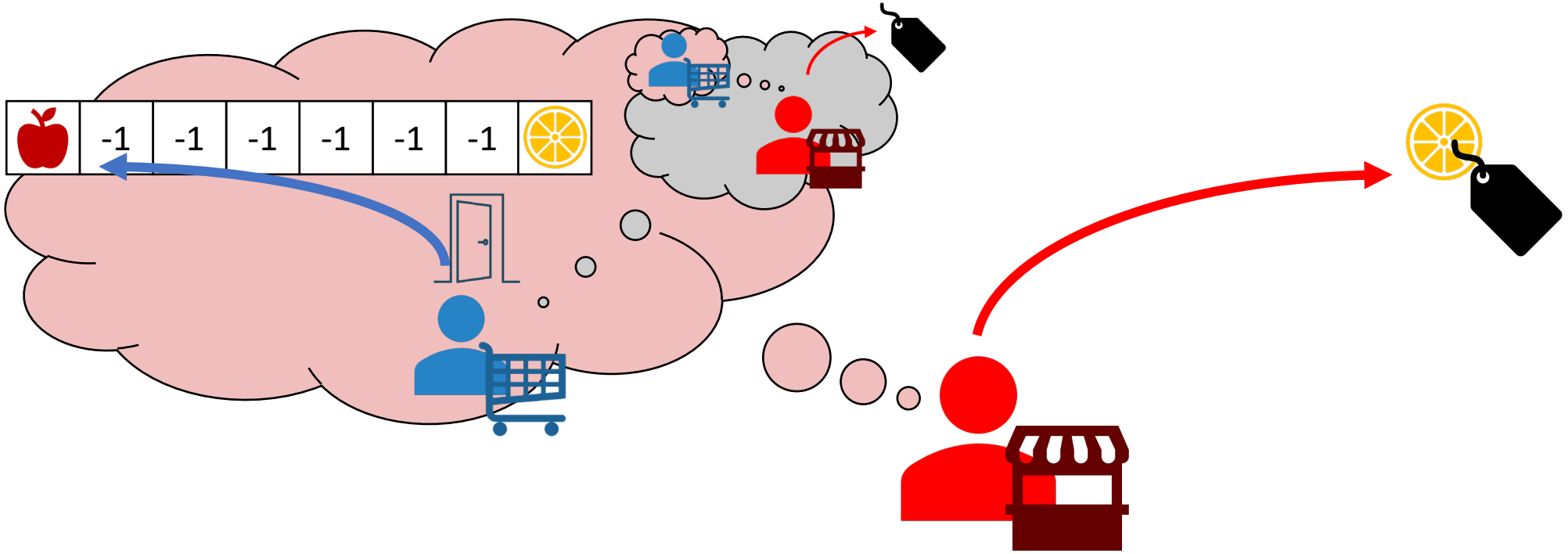


Optimizing $P(Value|Action, Cost)$ to its advantage

The theory of mind levels

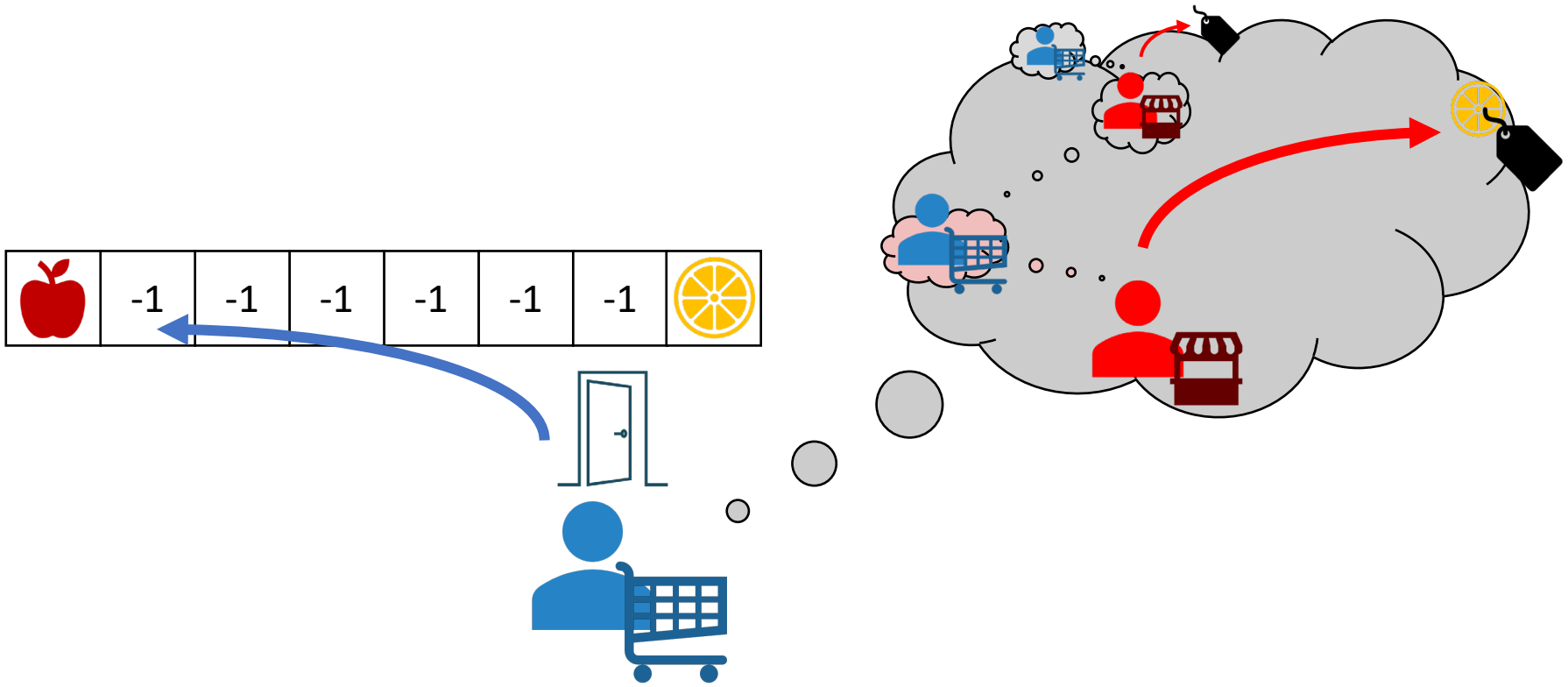
DoM(2) seller: Defending against the hack

A skeptical utility calculus



The theory of mind levels

DoM(3) buyer: Planning through the skeptical utility calculus

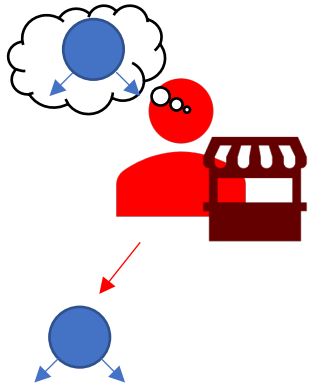


Recap: Theory of mind levels

DoM(-1)
Buyer

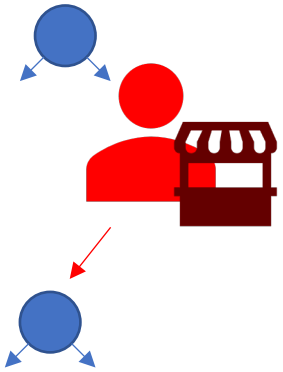
Reinforcement Learning (RL)

DoM(0)
Seller



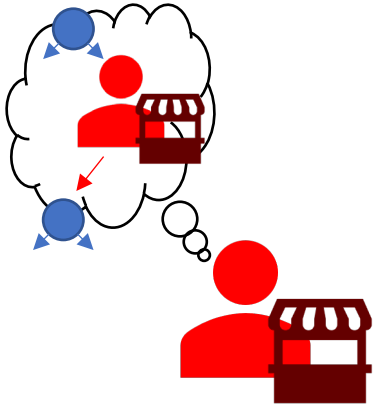
Inverse Reinforcement Learning (IRL)

DoM(1)
Buyer



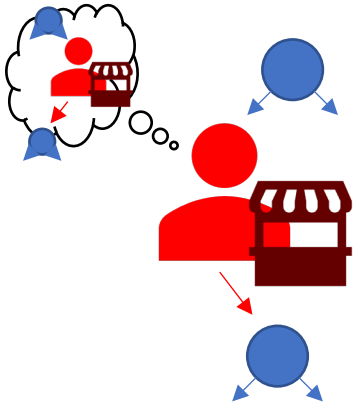
Hacking inverse RL

DoM(2)
Seller



Defending against the hacked IRL

DoM(3)
Buyer

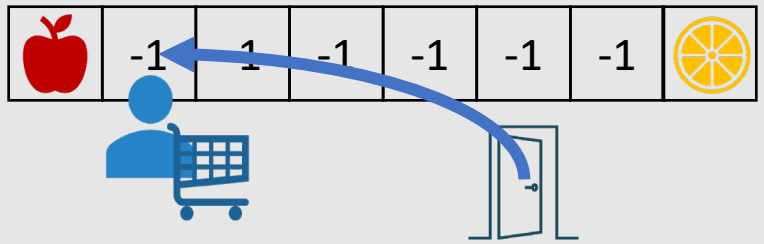


Planning through a defensive seller

Gmytrasiewicz & Doshi, 2005

Behaviors of interest

1. **Buyer** picks an item



2. **Seller** chooses a price
based on buyer's action

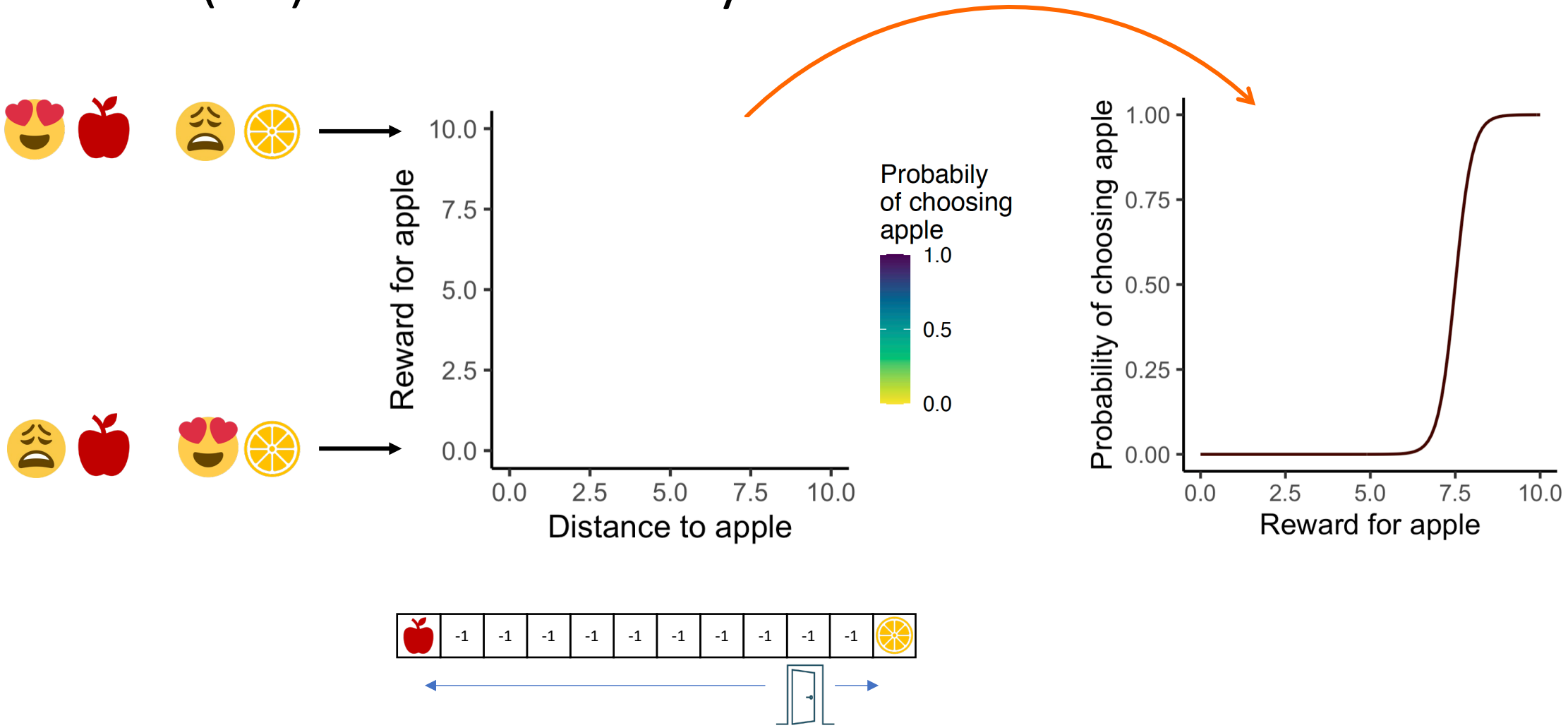
3. **Buyer** buys one of the options
based on seller's price

Recap: Theory of mind levels

DoM(-1)
Buyer

Reinforcement Learning (RL)

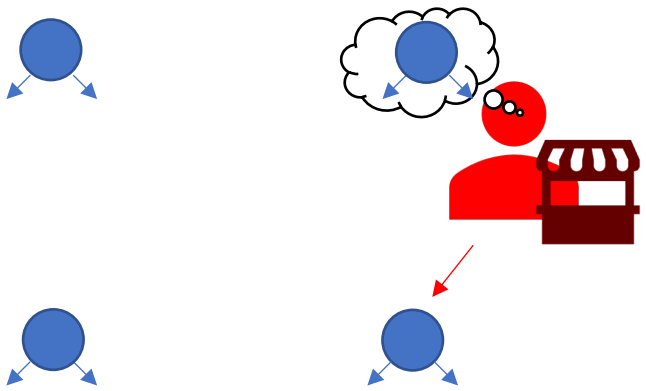
DoM(-1): “Naïve” Buyer



Recap: Theory of mind levels

DoM(-1)
Buyer

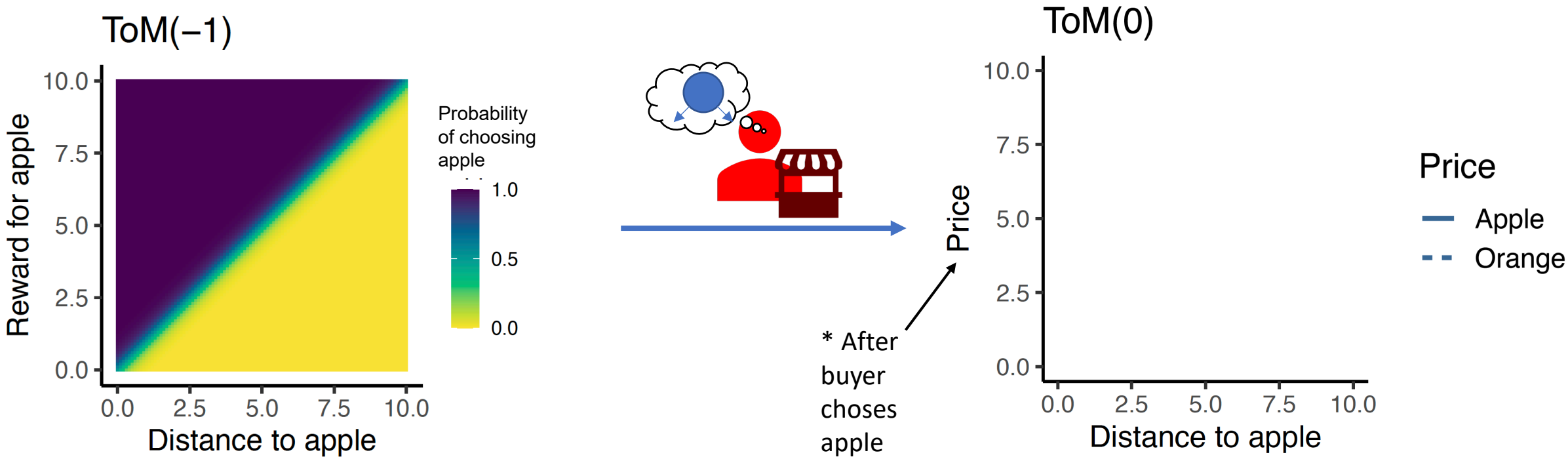
DoM(0)
Seller



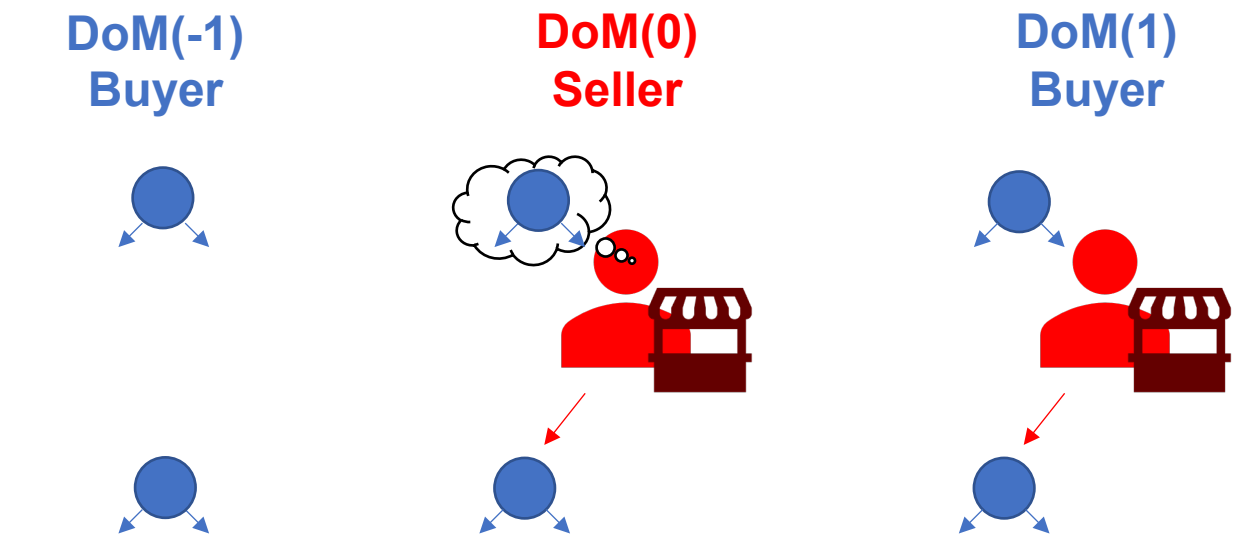
Reinforcement Learning (RL)

Inverse Reinforcement Learning (IRL)

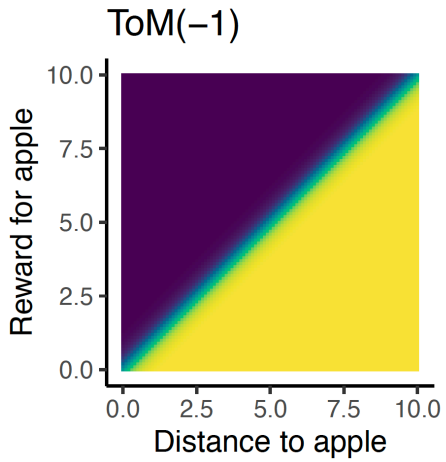
DoM(0): “Naïve Utility Calculus” Seller

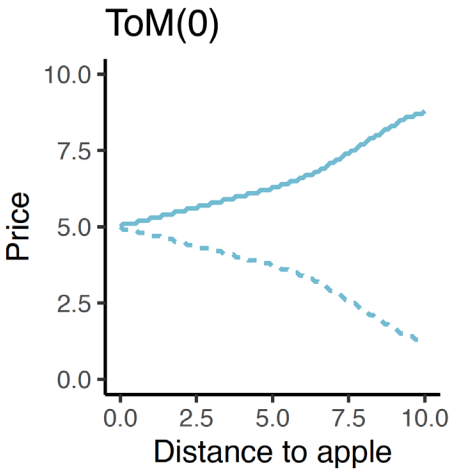


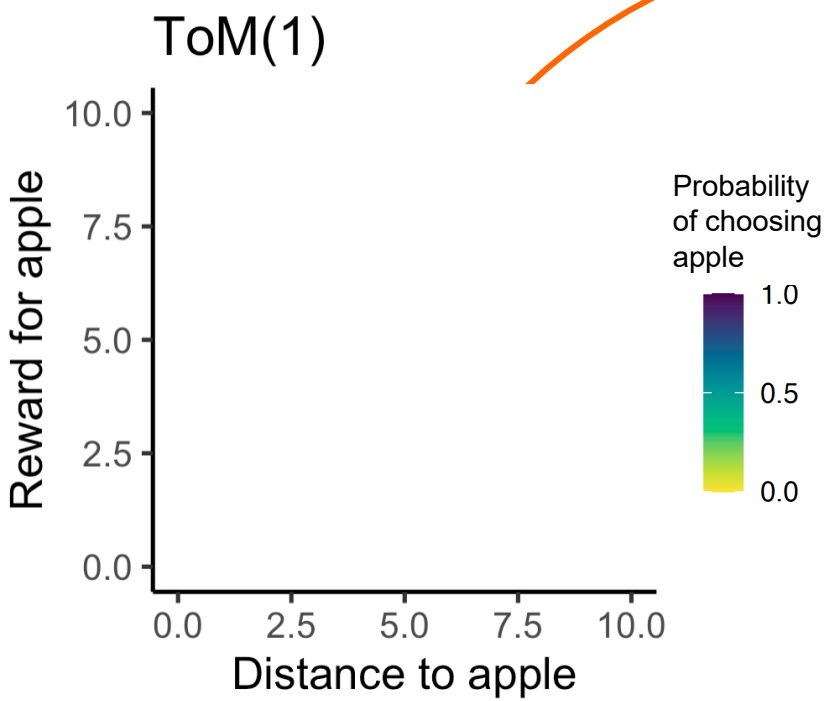
Recap: Theory of mind levels

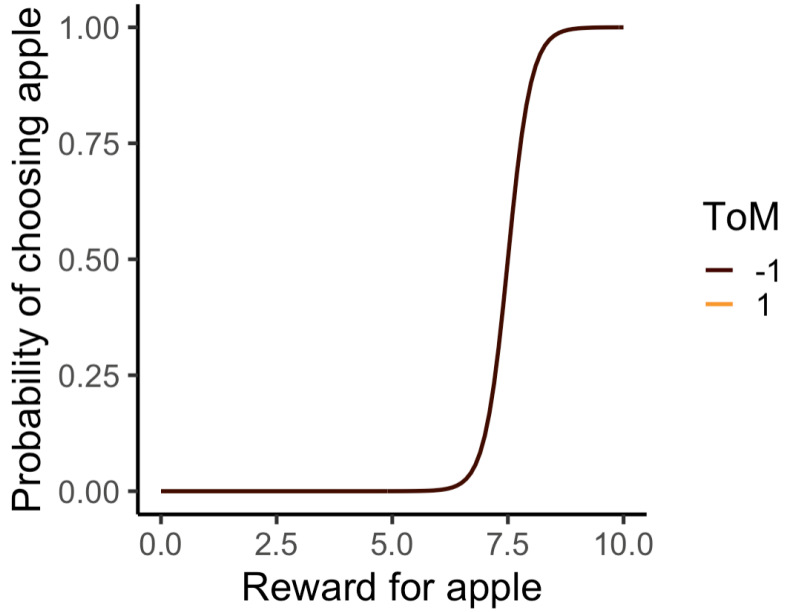


DoM(1): Hacking the “Naïve Utility Calculus”

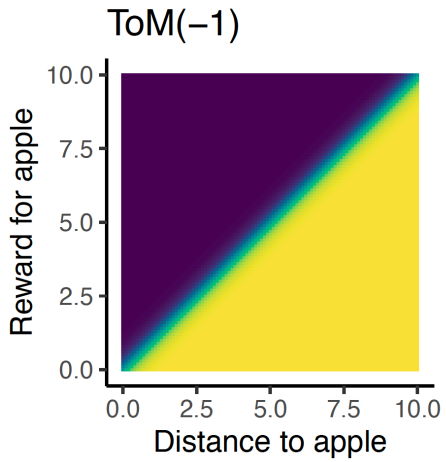


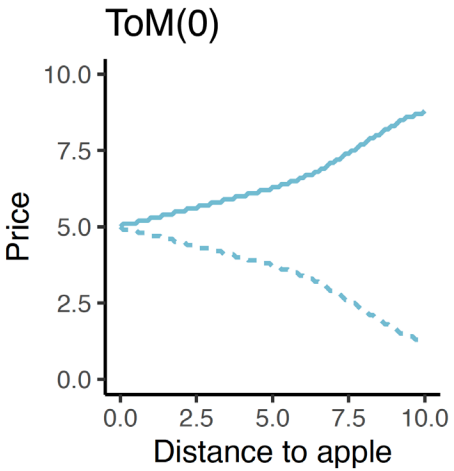


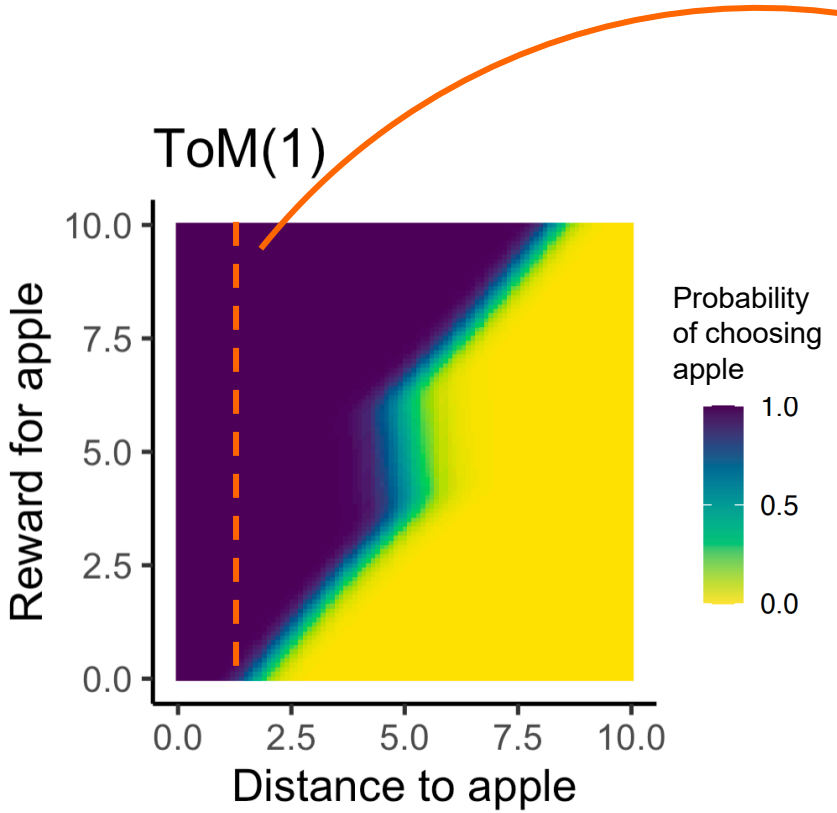


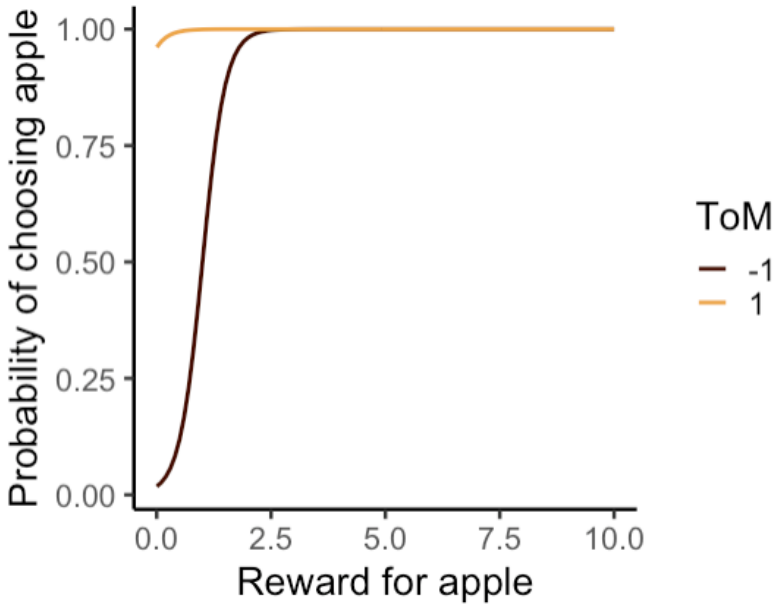


DoM(1): Hacking the “Naïve Utility Calculus”







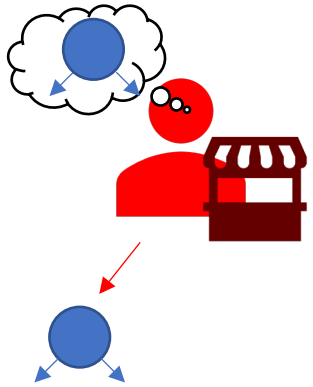


Recap: Theory of mind levels

DoM(-1)
Buyer

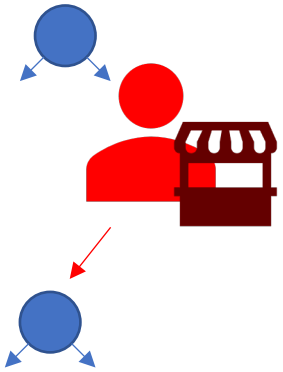
Reinforcement Learning (RL)

DoM(0)
Seller



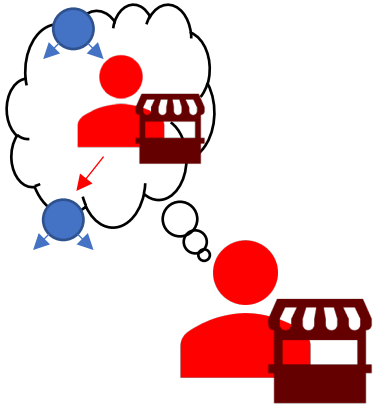
Inverse Reinforcement Learning (IRL)

DoM(1)
Buyer



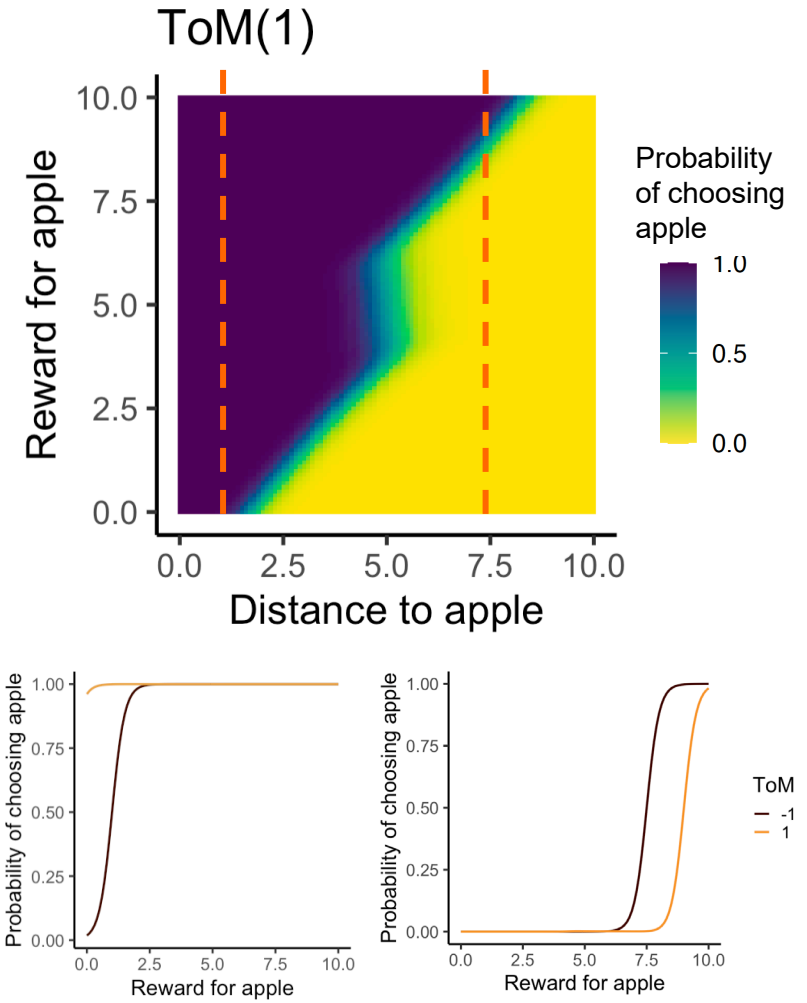
Hacking inverse RL

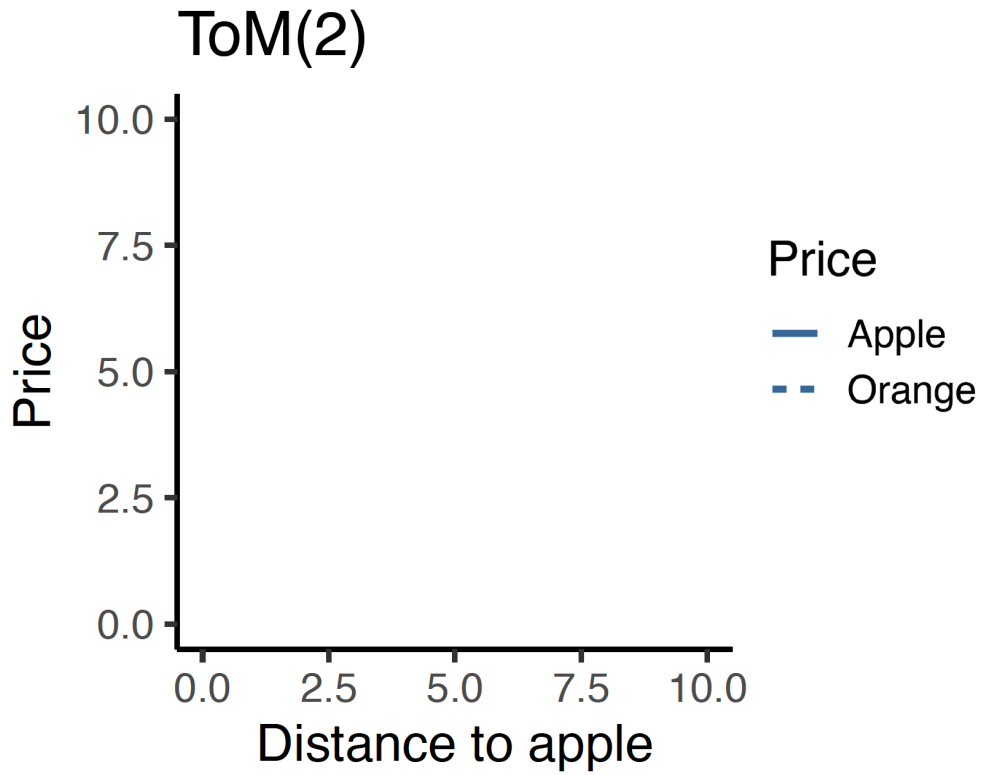
DoM(2)
Seller



Defending against the hacked IRL

DoM(2): The “Skeptical Utility Calculus”



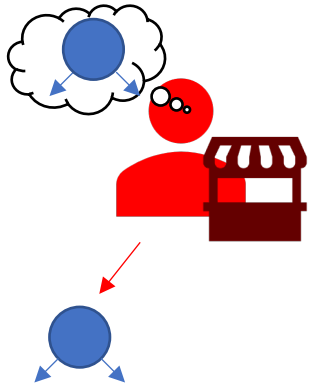


Recap: Theory of mind levels

DoM(-1)
Buyer

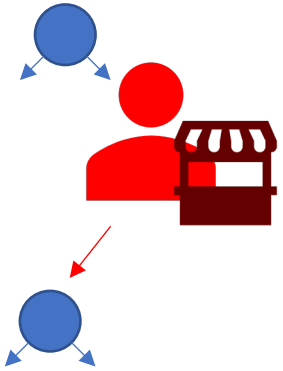
Reinforcement Learning (RL)

DoM(0)
Seller



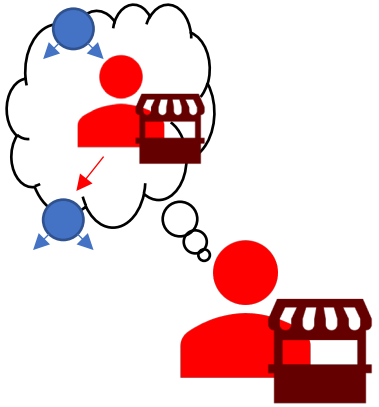
Inverse Reinforcement Learning (IRL)

DoM(1)
Buyer



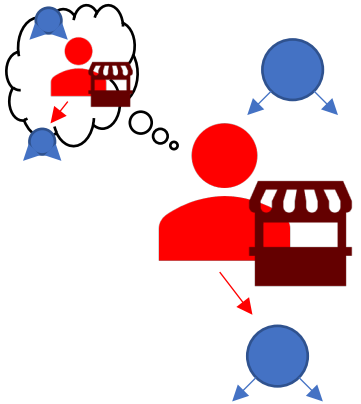
Hacking inverse RL

DoM(2)
Seller



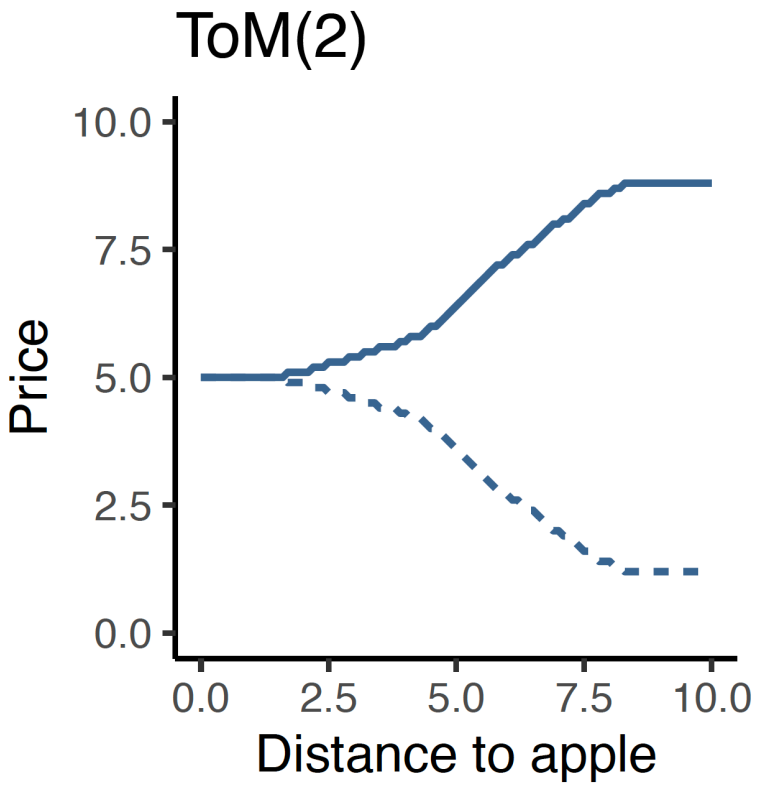
Defending against the hacked IRL

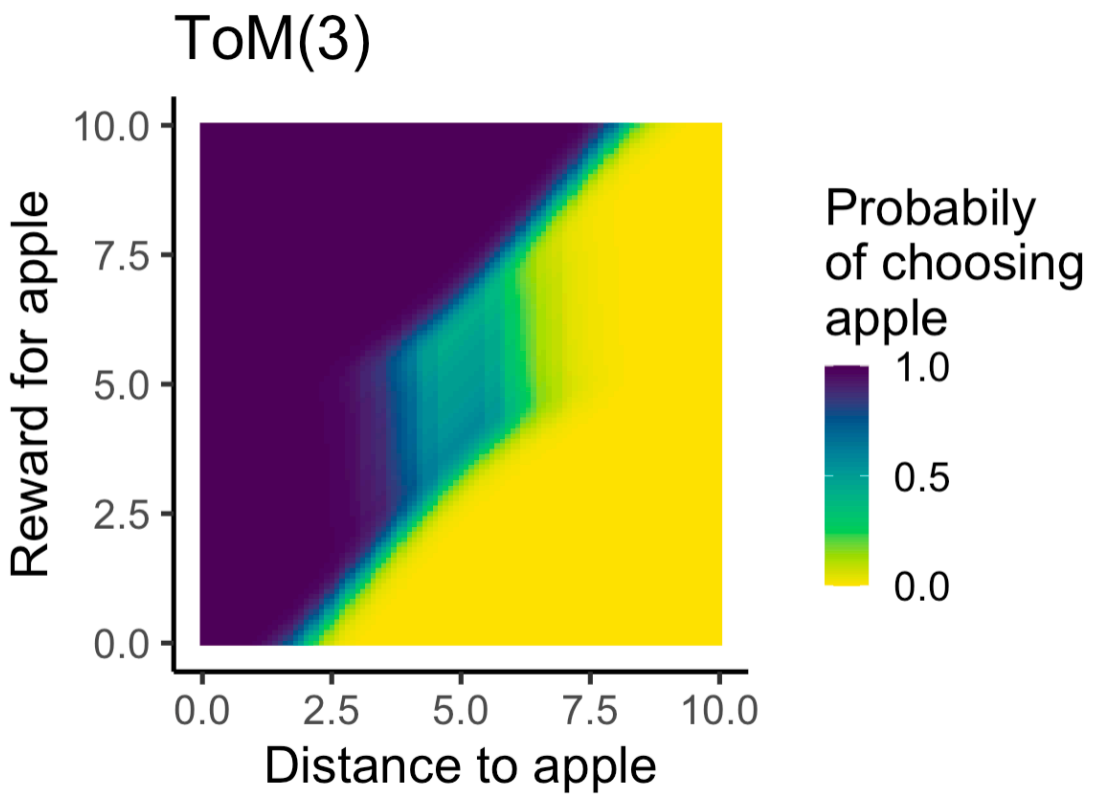
DoM(3)
Buyer



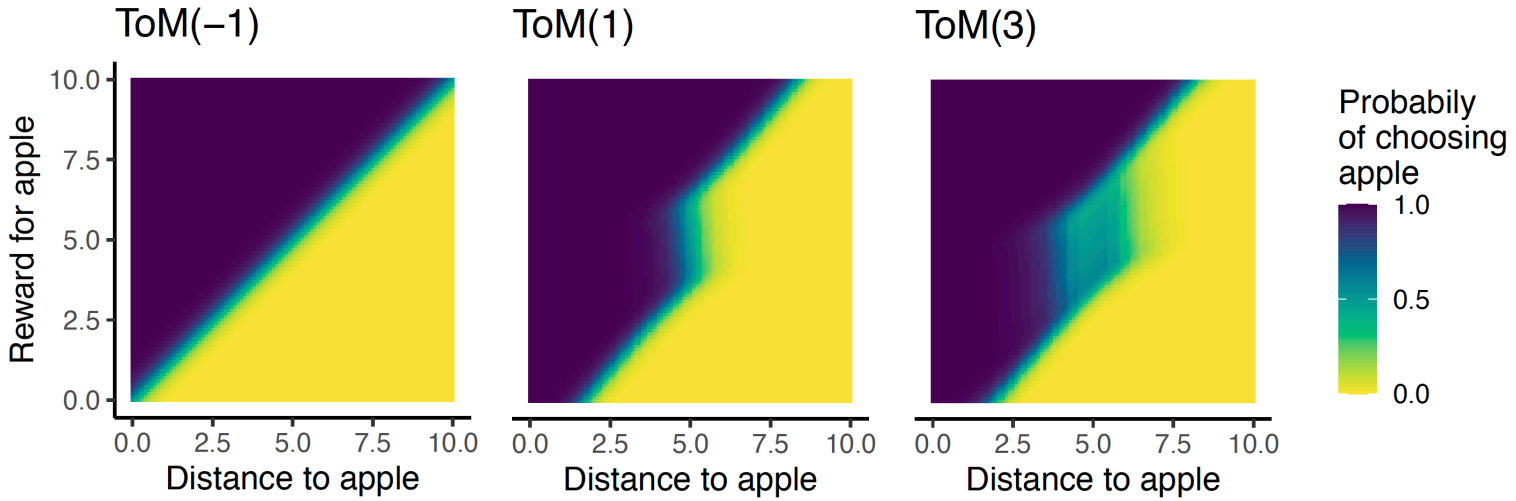
Planning through a defensive seller

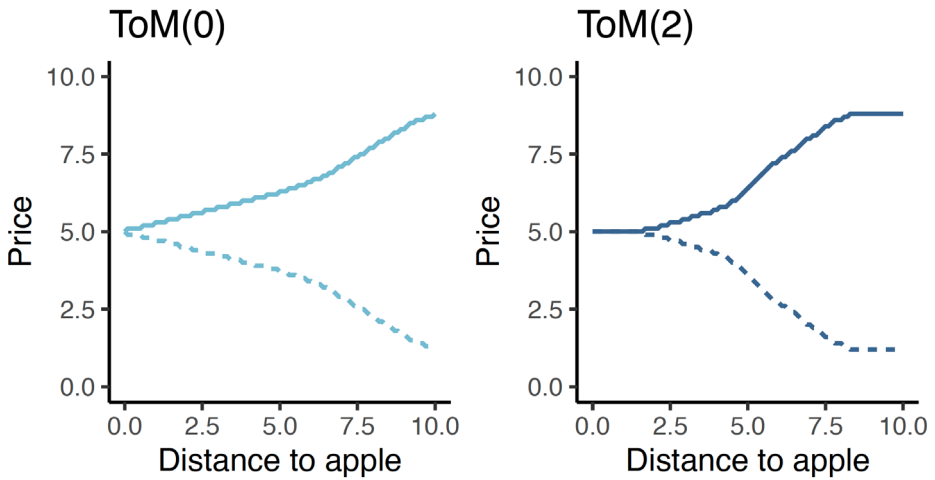
DoM(3): The “Cornered Hacker”





Intermediate summary

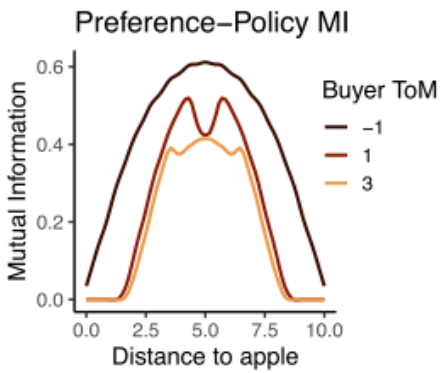




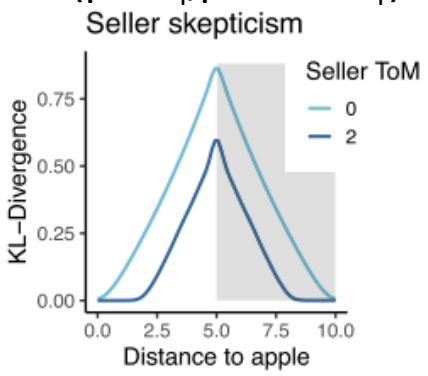
Intermediate summary

- **theory of mind and pure reward maximization** give rise to
  - *observed* agents (partially) **hiding their preferences**
  - *observing* agents selectively **ignoring or reinterpreting** the signals sent by the observed agents
  - belief updating in the face of opponent
- information theoretic view

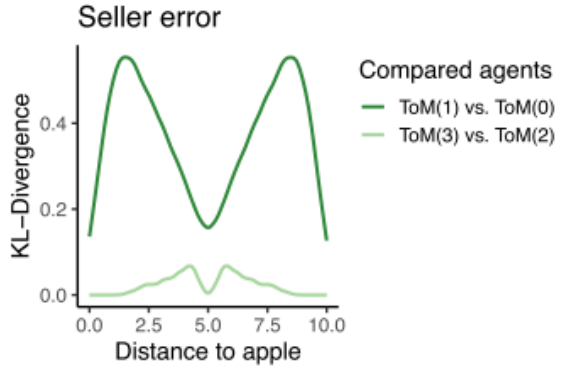
$I(r, a_1)$



KL(prior$_r$,posterior$_r$)



KL(buyer$_{k+1}$,seller$_k$)



Plan

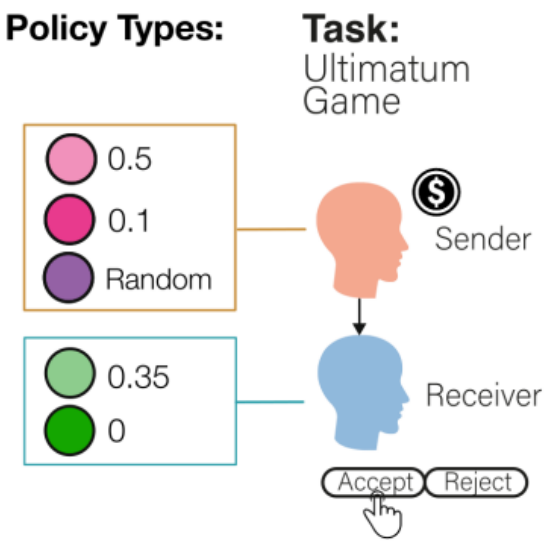
- theory of mind
- the p-beauty game
- recursive theory of mind in shopping
- recursive theory of mind in the ultimatum game
  - dynamic updating
  - defection

The Ultimatum Game

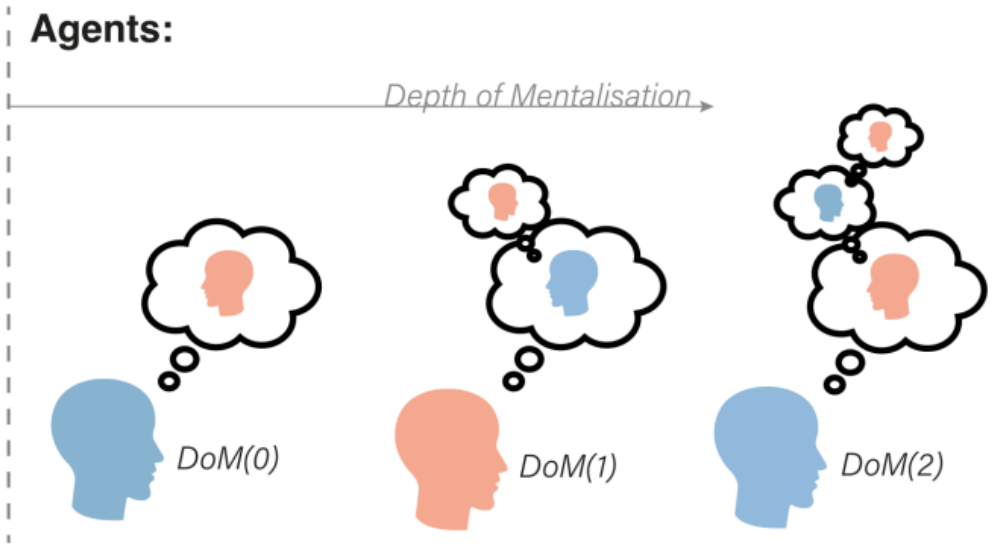
Oosterbeek et al, 2004

Country	*N* (1)	Mean offer (2)	Mean reject (3)	IDV (4)	PDI (5)	AUTH (6)	TRUST (7)	COMP (8)	GDP pc (9)	GINI index (10)
Austria	1	39.21	16.10	55	11	−0.05	0.32	6.78	12955	23.1
Bolivia	1	37.00	0.00						1721	42.0
Chile	1	34.00	6.70	23	63	1.10	0.23	5.94	4890	56.5
Ecuador	2	34.50	7.50	8	78				2830	46.6
France	3	40.24	30.78	71	68	−0.15	0.23	5.97	13918	32.7
Germany	1	36.70	9.52	67	35	−1.30	0.38	6.75	11666	30.0
Honduras	1	45.70	23.05						1385	53.7
Indonesia	4	46.63	14.63	14	78				2102	36.5
Israel	5	41.71	17.73	54	13				9843	35.5
Japan	3	44.73	19.27	46	54	−1.58	0.42	5.52	15105	24.9
Yugoslavia	1	44.33	26.67	27	76	−0.65	0.30	7.07	4548	31.9
Kenya	1	44.00	4.00	27	64				914	57.5
Mongolia	2	35.50	5.00						1842	33.2
Netherlands	2	42.25	9.24	80	38	−0.55	0.56	5.60	13281	31.5
Papua New-Guinea	2	40.50	33.50						1606	50.9
Paraguay	1	51.00	0.00						2178	59.1
Peru	1	26.00	4.80	16	64	1.75	0.05	6.54	2092	46.2
Romania	2	36.95	23.50				0.16	7.32	2043	28.2
Slovakia	3	43.17	12.67			−0.55	0.23	6.97	4095	19.5
Spain	1	26.66	29.17	51	57	0.60	0.34	5.70	9802	38.5
Sweden	1	35.23	18.18	71	31	−1.35	0.66	6.78	13986	25.0
Tanzania	4	37.50	19.25	27	64				534	38.2
UK	2	34.33	23.38	89	35	0.10	0.44	6.19	12724	32.6
US East	22	40.54	17.15	91	40	1.11	0.50	6.70	17945	40.1
US West	6	42.64	9.41	91	40	1.11	0.50	6.70	17945	40.1
Zimbabwe	2	43.00	8.50						1162	56.8

The Ultimatum Game



12 rounds



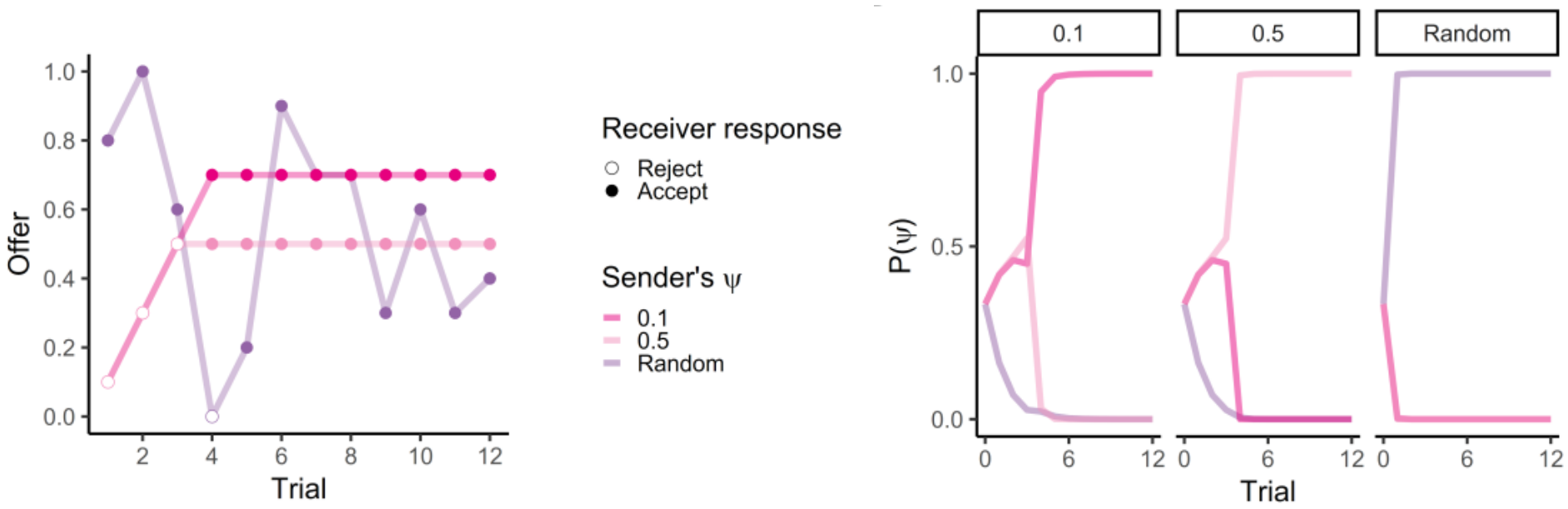
$$u_S^t(\eta_S, a_S^t, a_R^t) = (1 - a_S^t - \eta_S) * a_R^t$$

$$u_R^t(\eta_R, a_S^t, a_R^t) = (a_S^t - \eta_R) * a_R^t$$

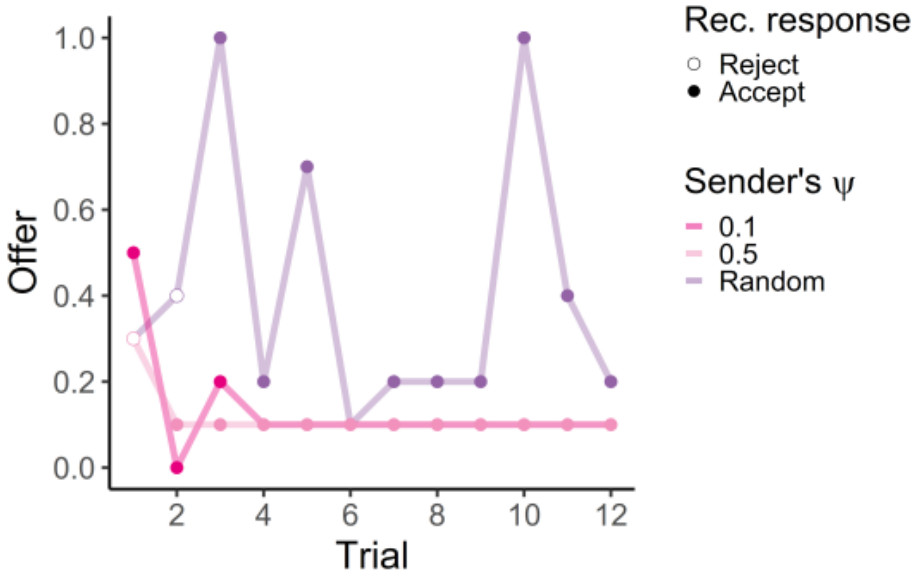
Sender

- DoM(-1):
  - random: just acts randomly
  - threshold: myopic use of lower and upper bounds
    - lower: increase from rejection
    - upper: decrease from acceptance
    - softmax policy within bounds
- DoM(1):
  - pretends to be random

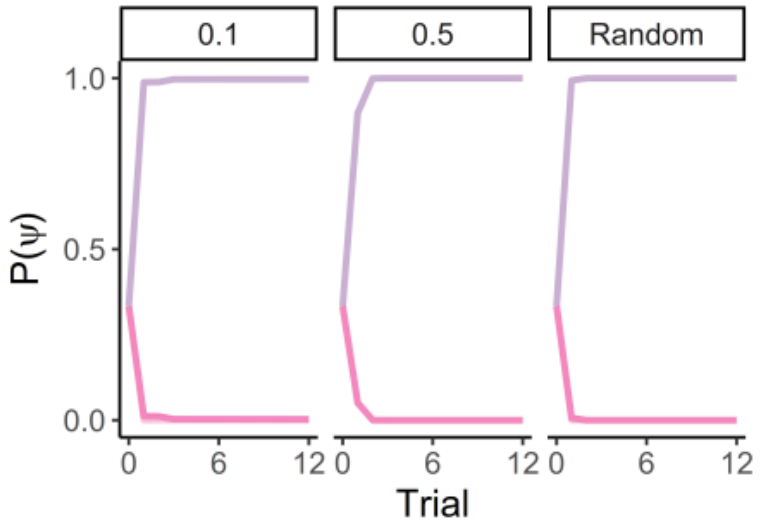
DoM(-1) vs DoM(0)

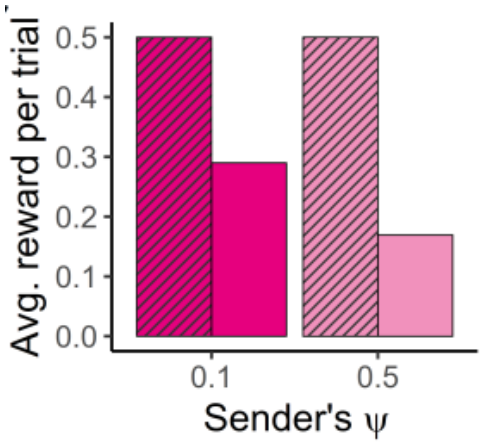

Offer
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
Trial
Receiver response
Reject
Accept
Sender's ψ
0.1
0.5
Random
0.1
0.5
Random
P(ψ)
1.0
0.5
0.0
0
6
12
Trial

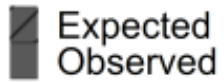
DoM(1) vs DoM(0)


Offer
1.0
0.8
0.6
0.4
0.2
0.0
2
4
6
8
10
12
Trial

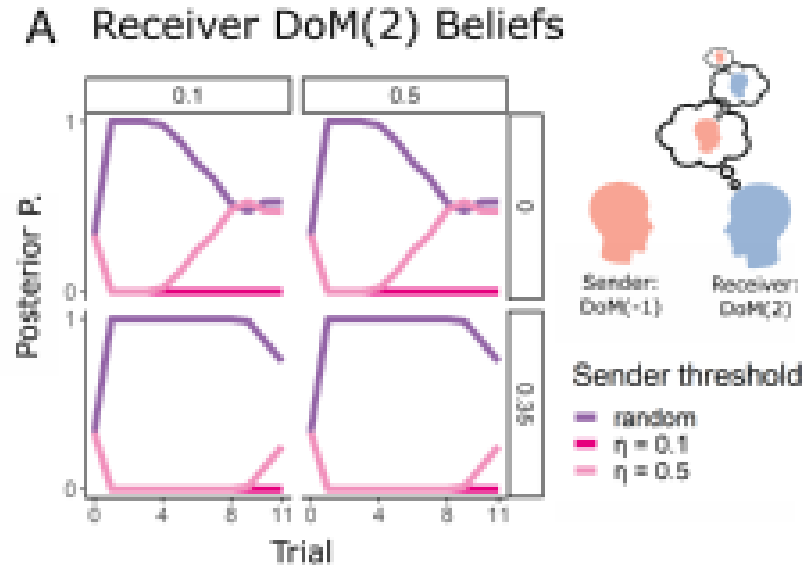
Rec. response
Reject
Accept
Sender's ψ
0.1
0.5
Random


0.1
0.5
Random
P(ψ)
1.0
0.5
0.0
0
6
12
Trial


Avg. reward per trial
0.5
0.4
0.3
0.2
0.1
0.0
0.1
0.5
Sender's ψ


Expected
Observed

DoM(2) vs DoM(-1) : ‘Paranoia’

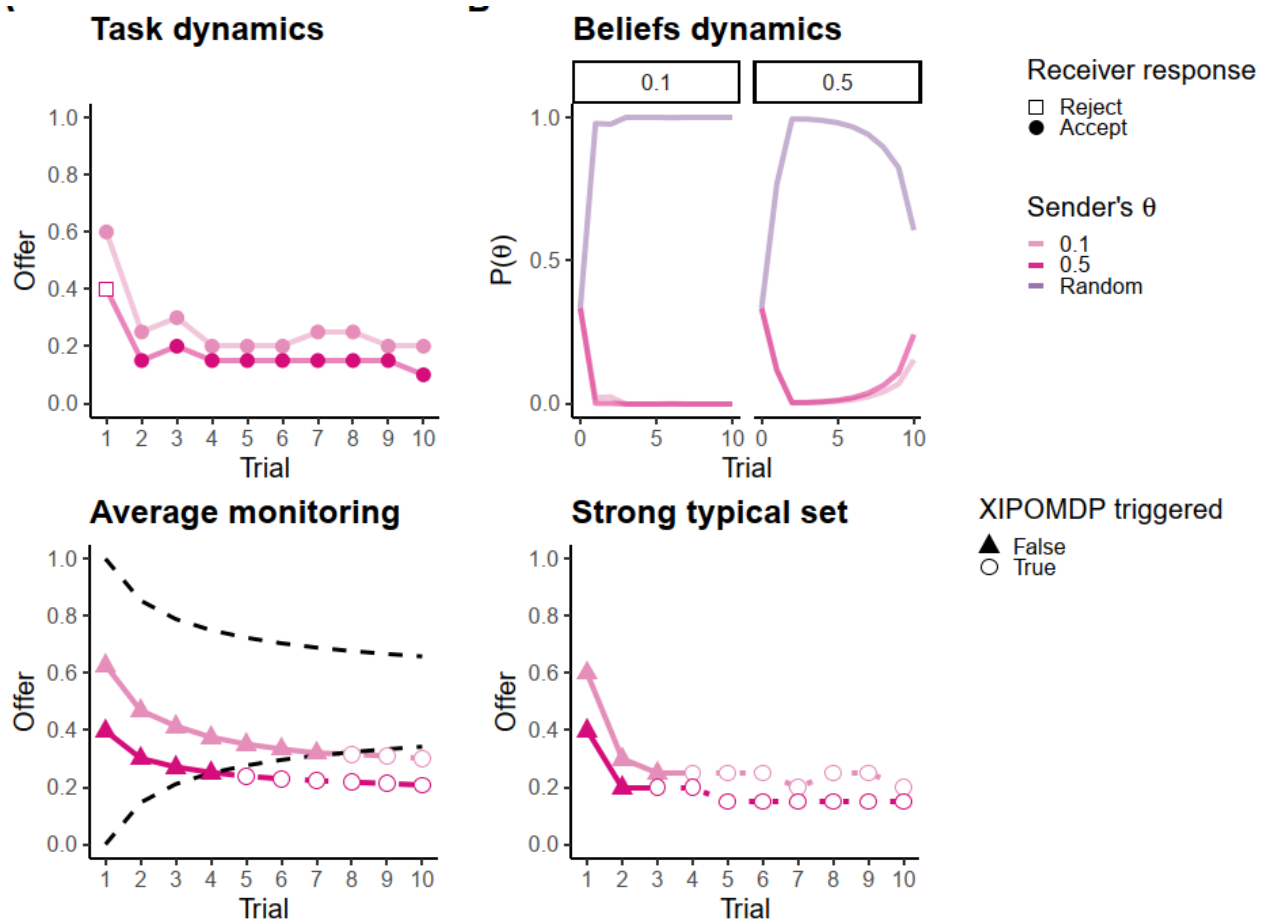


- DoM(1) masquerades as random
  - so true randomness is hard to infer – implying cost for the receiver
  - DoM(-1) threshold behaviour is atypical for DoM(1) random-pretenders
    - so infer that it is actually random
    - and so lose out!

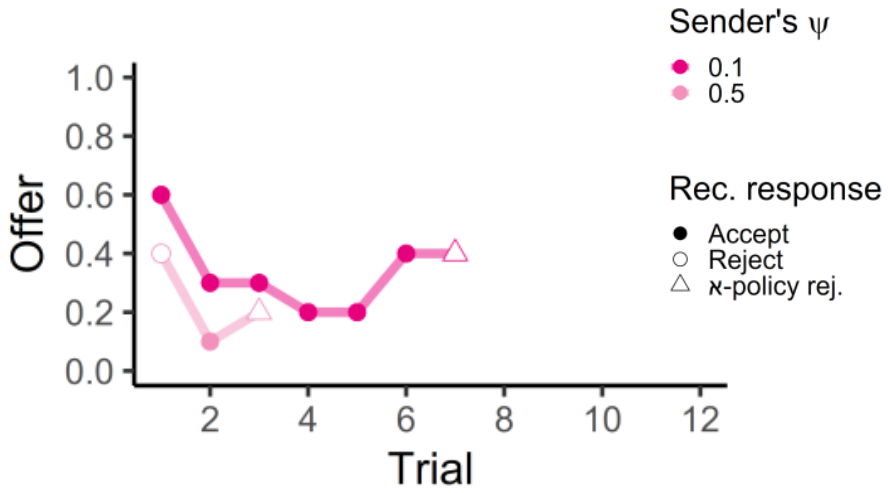
Self-Protection

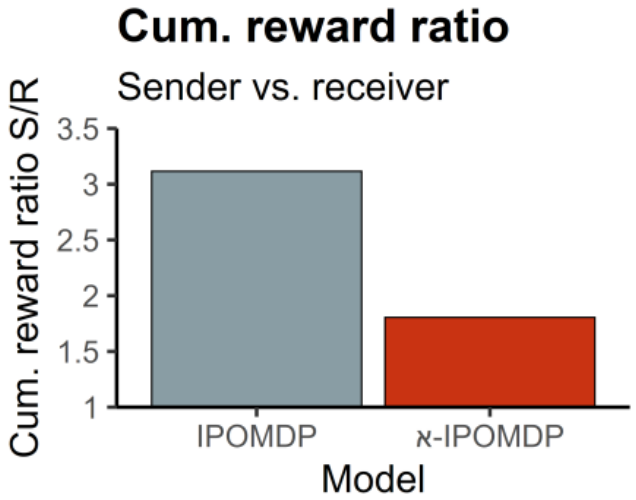
- how can DoM(0) protect against DoM(1)?

non-stationarity



Credible Threats





Summary

- from inverse RL to deception, protection, threats
  - dynamic belief updating; non-stationarity; mixed motives
- depth of mentalizing
- critical dependence on DoM(-1) strategy
- difficulty of being too smart
- difficulty of being insufficiently smart
  - need to detect manipulation (non-likelihood based)
  - need an א-policy
- personality disorders: irritation