

Naive Utility Calculus

Joseph Low

August 2025

Table of Contents

- 1 Introduction
- 2 Related Work
- 3 Experiments 1 and 2
- 4 Naive Utility Calculus
- 5 Experiment 5
- 6 Results
- 7 Discussion

Table of Contents

1 Introduction

2 Related Work

3 Experiments 1 and 2

4 Naive Utility Calculus

5 Experiment 5

6 Results

7 Discussion

How do we make sense of other people's behavior?

How do we make sense of other people's behavior?

- Why did you sleep late last night?

How do we make sense of other people's behavior?

- Why did you sleep late last night?
- Why did you sign up for this course?

How do we make sense of other people's behavior?

- Why did you sleep late last night?
- Why did you sign up for this course?
- Why did you choose to eat out instead of cooking?

Utility-based decision-making

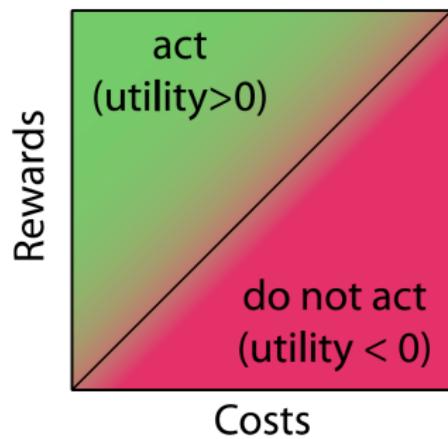
$$\text{Utility} = \text{Rewards} - \text{Costs}$$

- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior

Utility-based decision-making

$$\text{Utility} = \text{Rewards} - \text{Costs}$$

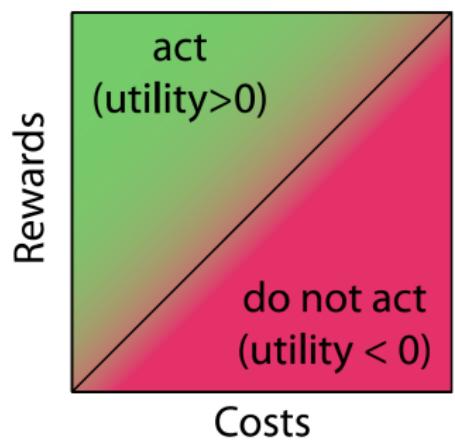
- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior



Utility-based decision-making

$$\text{Utility} = \text{Rewards} - \text{Costs}$$

- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior

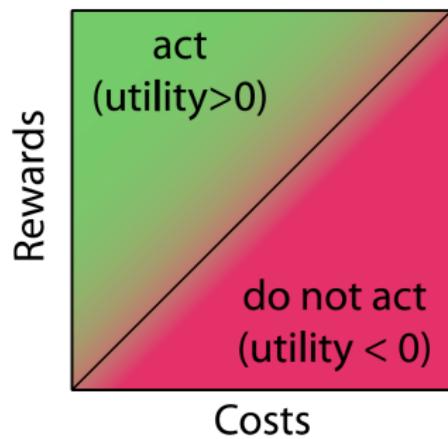


- Pursue high-cost implies reward was even higher

Utility-based decision-making

$$\text{Utility} = \text{Rewards} - \text{Costs}$$

- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior

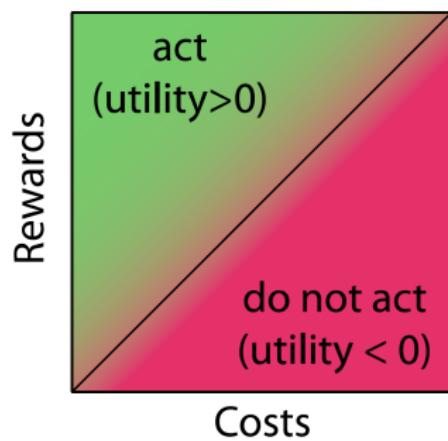


- Pursue high-cost implies reward was even higher
- Pursue low-cost does not give info about reward

Utility-based decision-making

$$\text{Utility} = \text{Rewards} - \text{Costs}$$

- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior

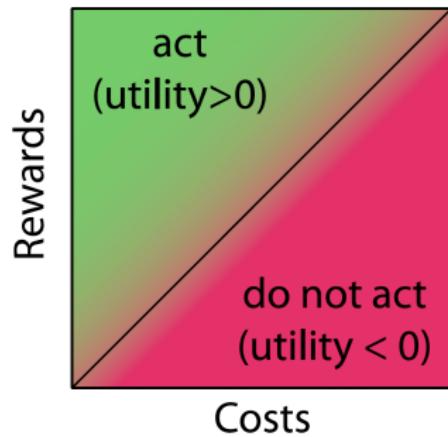


- Pursue high-cost implies reward was even higher
- Pursue low-cost does not give info about reward
- Forego high-cost does not give info about reward

Utility-based decision-making

$$\text{Utility} = \text{Rewards} - \text{Costs}$$

- There is empirical support that humans intuitively use utility-based reasoning to make sense of other people's behavior



- Pursue high-cost implies reward was even higher
- Pursue low-cost does not give info about reward
- Forego high-cost does not give info about reward
- Forego low-cost implies reward was even lower

Naive Utility Calculus

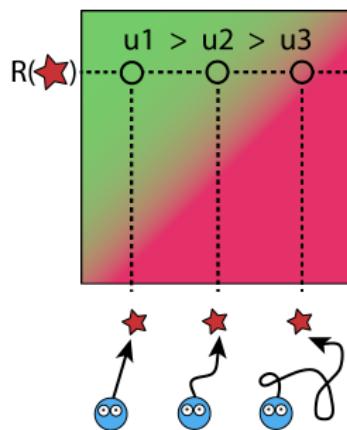
$$U(p, o) = R(o) - C(p)$$

- $U(p, o)$: utility expected from acting according to plan p to reach outcome o
- $R(o)$: subjective reward the agent expects from outcome o
- $C(p)$: subjective cost of executing plan p

Caveats

- **Descriptive, not normative:** This is not about how people *should* make decisions (economic utility theory)
- **How we actually operate:** This describes how we *intuitively* make sense of other people's behavior
- People don't explicitly compute utilities when they act - this is the cognitive framework we use to understand others

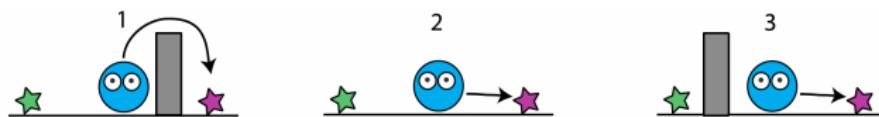
Utility and Efficiency



- More efficient paths are less costly and therefore produce higher utilities
- When agents act, they will fulfill their goals as efficiently as possible to maximize utility

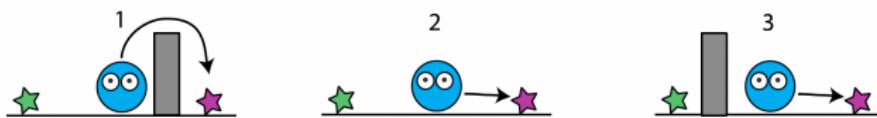
Graded Preference Inference

Does the agent prefer the green or purple star? To what extent?



Graded Preference Inference

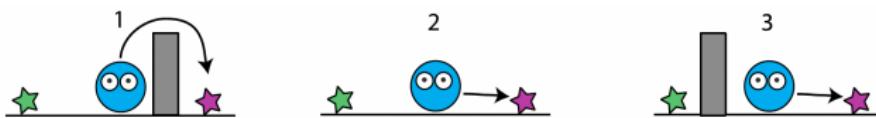
Does the agent prefer the green or purple star? To what extent?



- **Fig.1: Pursue high-cost:** Agent pays high cost (long detour) to get purple star → strong preference inference

Graded Preference Inference

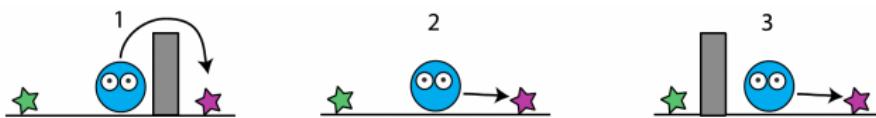
Does the agent prefer the green or purple star? To what extent?



- **Fig.1: Pursue high-cost:** Agent pays high cost (long detour) to get purple star → strong preference inference
- **Fig.2: Pursue low-cost:** Agent pays low cost (short path) to get purple star → weak preference inference

Graded Preference Inference

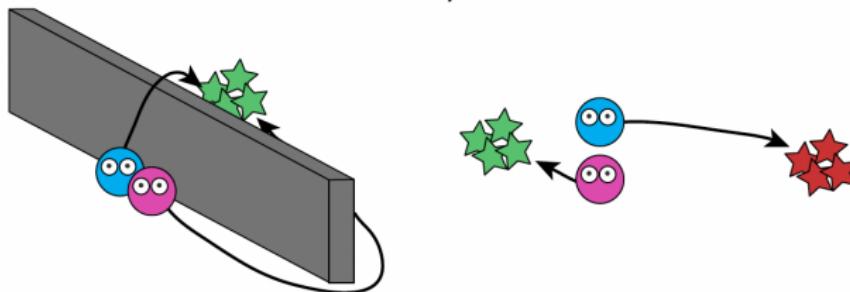
Does the agent prefer the green or purple star? To what extent?



- **Fig.1: Pursue high-cost:** Agent pays high cost (long detour) to get purple star → strong preference inference
- **Fig.2: Pursue low-cost:** Agent pays low cost (short path) to get purple star → weak preference inference
- **Fig.3: Forego high-cost:** Agent foregoes high cost (doesn't climb wall) to get green star, chooses purple instead → no preference inference

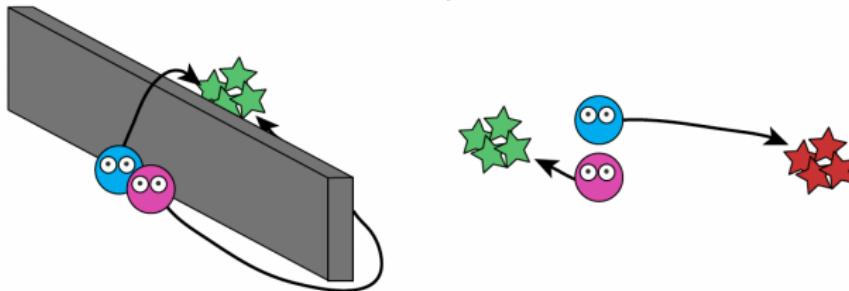
Costs and Rewards Vary Across Agents

What do the agent's actions imply about their preferences?



Costs and Rewards Vary Across Agents

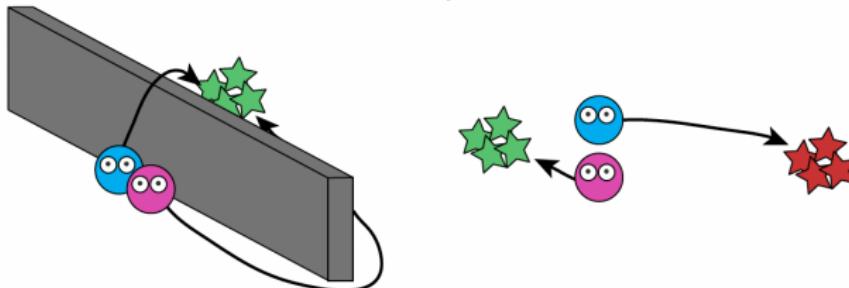
What do the agent's actions imply about their preferences?



- **Costs vary:** Blue agent can climb walls easily, pink agent cannot → different action costs

Costs and Rewards Vary Across Agents

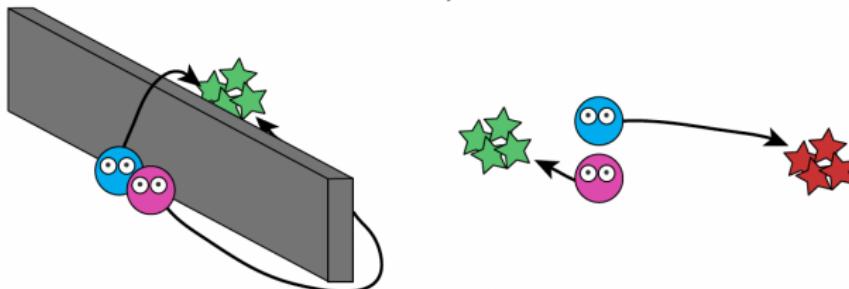
What do the agent's actions imply about their preferences?



- **Costs vary:** Blue agent can climb walls easily, pink agent cannot → different action costs
- **Rewards vary:** Blue agent prefers red stars, pink agent prefers green stars → different goal values

Costs and Rewards Vary Across Agents

What do the agent's actions imply about their preferences?



- **Costs vary:** Blue agent can climb walls easily, pink agent cannot → different action costs
- **Rewards vary:** Blue agent prefers red stars, pink agent prefers green stars → different goal values
- **Same action, different inference:** Identical behavior can imply different preferences based on agent capabilities and values

Table of Contents

1 Introduction

2 Related Work

3 Experiments 1 and 2

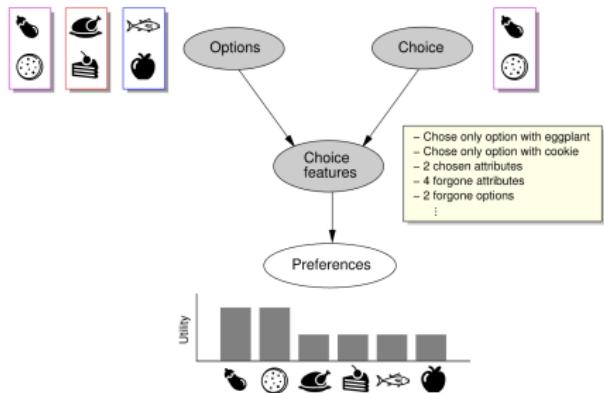
4 Naive Utility Calculus

5 Experiment 5

6 Results

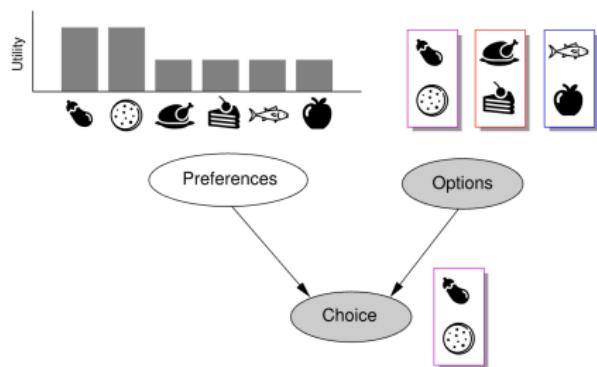
7 Discussion

Feature-Based Approach



- **Method:** Look at simple features of the choice and apply rules
- **Example:** Alice chooses {eggplant sandwich} over {turkey, tuna, ham}
- **Feature:** "She chose the only option with eggplant"
- **Rule:** "When someone picks the unique option, they probably like that thing"
- **Conclusion:** "Alice likes eggplant"

Inverse Decision-Making Approach



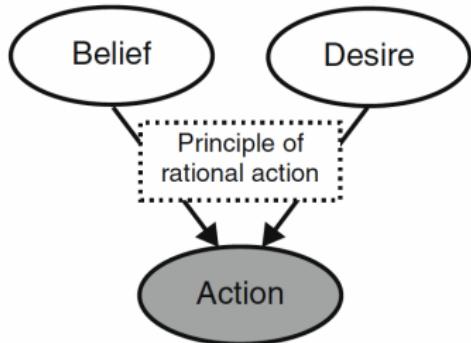
- **Method:** Use a model of how people actually make decisions, then work backwards
- **Example:** Same choice - Alice chooses {eggplant sandwich} over {turkey, tuna, ham}
- **Model thinking:** "If Alice really liked eggplant, how likely would she be to make this choice? What if she liked turkey instead? What if she just wanted any sandwich?"
- **Calculation:** Uses math to figure out which preference scenario makes this choice most probable

NUC vs. Inverse Decision-Making

- **Similarity:** NUC is similar to inverse decision-making as it works on assumption that agents maximize utilities
- **Key Differences:**
 - Uses events with complex spatiotemporal structures and not just isolated discrete choices
 - Computes both variables costs and rewards, whereas inverse decision-making only infers rewards without costs

Inverse Planning

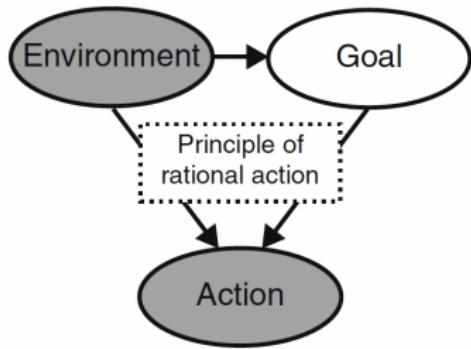
(a)



Belief-Desire Psychology ("Forward-thinking")

- Belief + Desire → Action
- "Sarah believes store is open + wants coffee → walks to coffee shop"

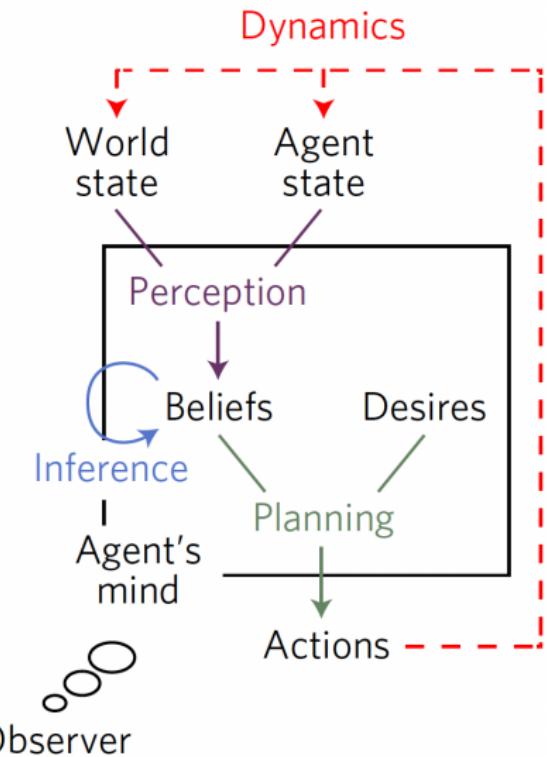
(b)



Inverse Planning ("Backward-thinking")

- Observe: Environment + Action → Infer: Goal
- "See maze layout + this path → agent wants point A"

Goal-Directed Action Understanding



Limitations

- Model does not explain multiple causes behind other people's goals
- Treats cost as constant, observable and uniform across agents

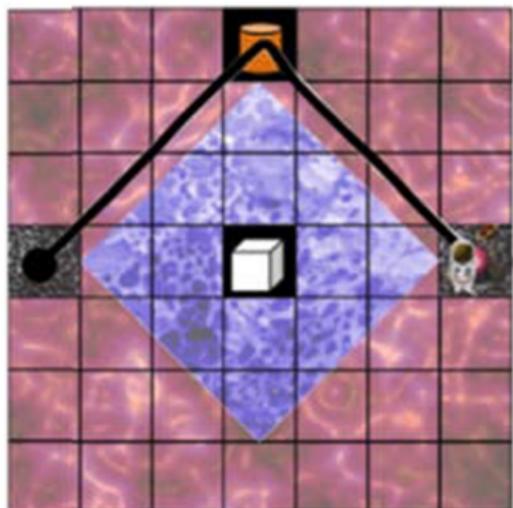
Research Questions

- **RQ1:** Can NUC support joint inference of costs and rewards when we know neither, using a coherent generative model?
- **RQ2:** Does NUC drive fine-grained quantitative inferences or only coarse qualitative ones?
- **RQ3:** Is NUC a unified generative model supporting probabilistic inference, or a collection of simple heuristics?

Table of Contents

- 1 Introduction
- 2 Related Work
- 3 Experiments 1 and 2
- 4 Naive Utility Calculus
- 5 Experiment 5
- 6 Results
- 7 Discussion

Interactive Experiment



You will watch astronauts land on an alien planet with novel terrains and travel towards their space station. The astronauts can always travel across all types of terrain and can collect up to one care package on their way home.

Your Task: Determine the astronauts'

- **Ability** to travel each kind of terrain (costs)
- **Desire** to collect each care package (rewards)

Quick Check

- ① Do astronauts have identical or different abilities?
- ② Do astronauts have identical or different care package preferences?
- ③ Can astronauts always cross all terrains, even if exhausting?
- ④ How many care packages must/can astronauts collect?

Quick Check

- ① Do astronauts have identical or different abilities?

Answer: Different astronauts have different abilities

- ② Do astronauts have identical or different care package preferences?

Answer: Different astronauts have different preferences

- ③ Can astronauts always cross all terrains, even if exhausting?

Answer: Yes, all astronauts can travel across all terrains

- ④ How many care packages must/can astronauts collect?

Answer: Up to one care package (optional)

Questions to Answer

For Terrain

"How easy is it to cross this terrain?"

Scale: 0 to 10

- **0** = Extremely easy
- **5** = Average
- **10** = Extremely exhausting

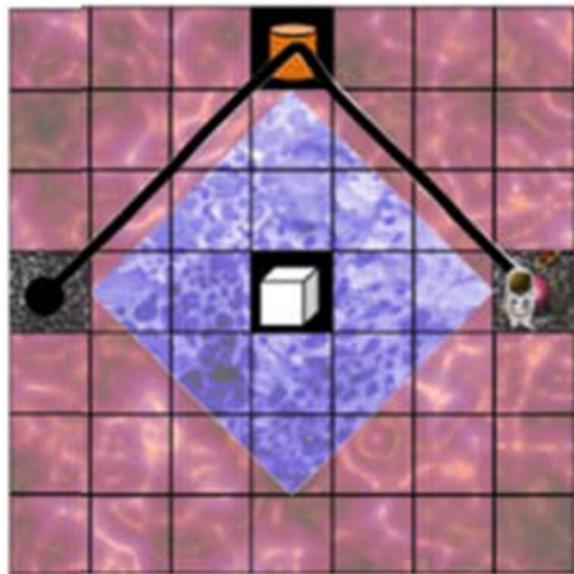
Care Package

"How much does the astronaut like this container?"

Scale: 0 to 10

- **0** = Not at all
- **5** = Average
- **10** = A lot

Experiment 1a-1



Based on this astronaut's path, rate on a scale of 0-10:

1. Red terrain



How easy is it to cross?

2. Blue terrain



How easy is it to cross?

3. White care package



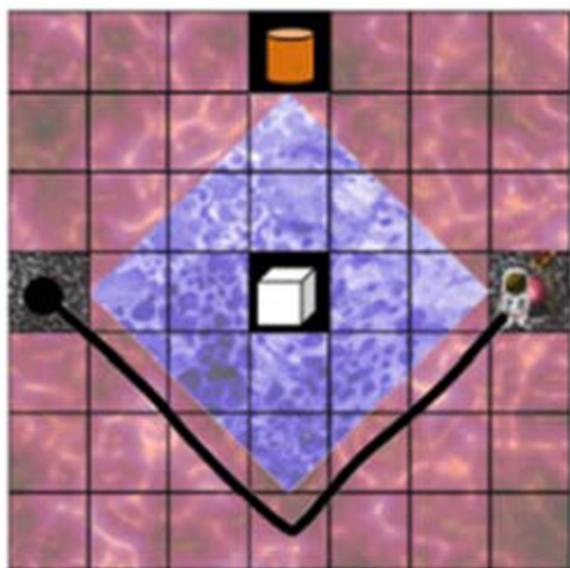
How much does the astronaut like it?

4. Orange care package



How much does the astronaut like it?

Experiment 1a-2



Based on this astronaut's path, rate on a scale of 0-10:

1. Red terrain



How easy is it to cross?

2. Blue terrain



How easy is it to cross?

3. White care package



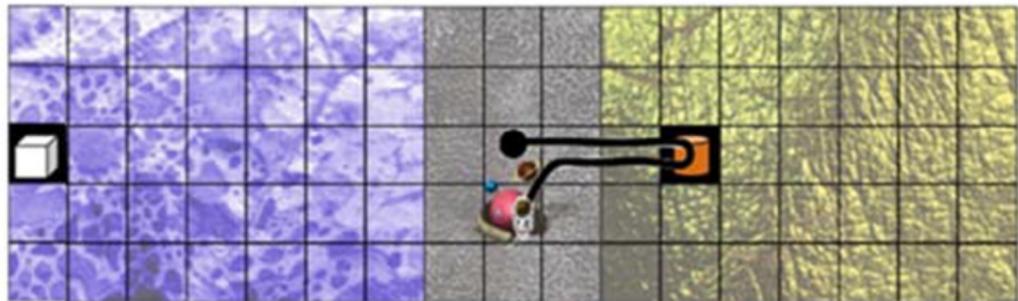
How much does the astronaut like it?

4. Orange care package



How much does the astronaut like it?

Experiment 1b-1



Based on this astronaut's path, rate on a scale of 0-10:

1. Blue terrain



How easy is it to cross?

2. Yellow terrain



How easy is it to cross?

3. White care package



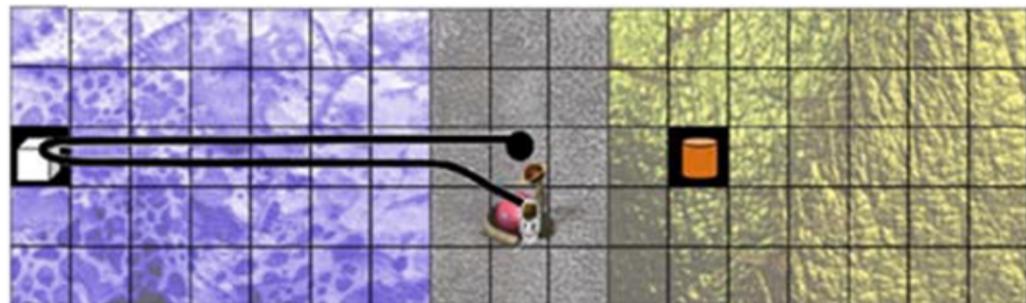
How much does the astronaut like it?

4. Orange care package



How much does the astronaut like it?

Experiment 1b-2



Based on this astronaut's path, rate on a scale of 0-10:

1. Blue terrain



How easy is it to cross?

2. Yellow terrain



How easy is it to cross?

3. White care package



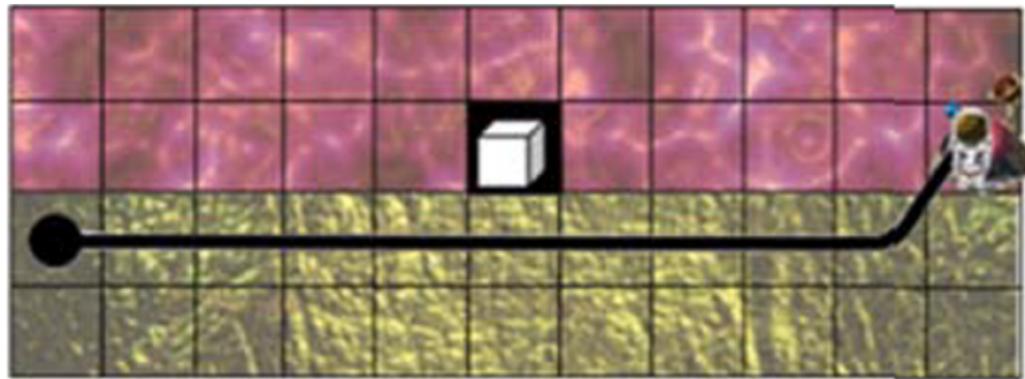
How much does the astronaut like it?

4. Orange care package



How much does the astronaut like it?

Experiment 1c-1



Based on this astronaut's path, rate on a scale of 0-10:

1. Red terrain



How easy is it to cross?

2. Yellow terrain



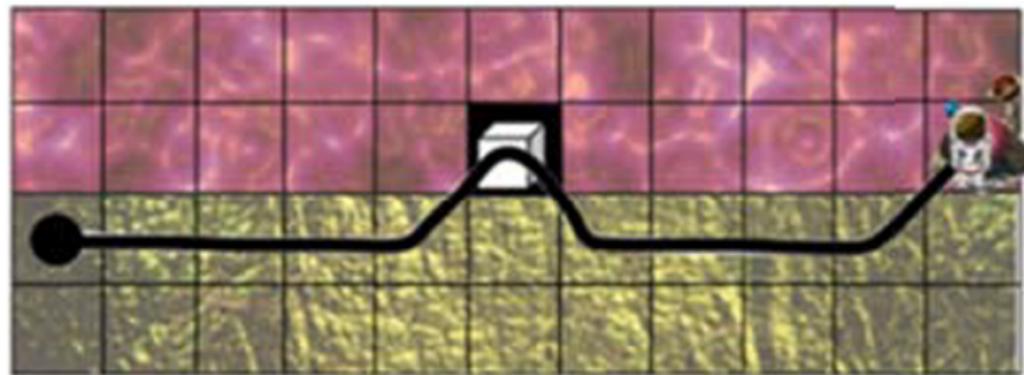
How easy is it to cross?

3. White care package



How much does the astronaut like it?

Experiment 1c-2



Based on this astronaut's path, rate on a scale of 0-10:

1. Red terrain



How easy is it to cross?

2. Yellow terrain



How easy is it to cross?

3. White care package



How much does the astronaut like it?

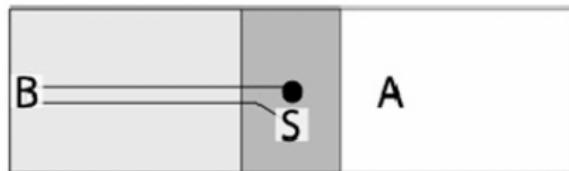
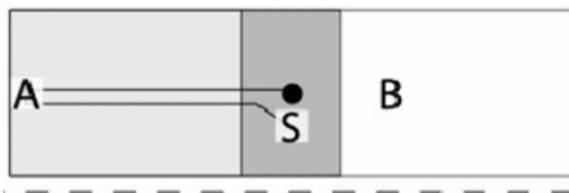
Experiment Types

Participants watched agents navigate worlds with different terrains and care packages to infer cost and reward functions. Three different geometries were used:

- **Experiment 1a:** Detours could be explained by either terrain costs or reward differences
- **Experiment 1b:** Forced terrain crossings vs. optional detours to collect care packages
- **Experiment 1c:** Minimal physical detours that still reveal agent preferences

Responses evaluated against Naïve Utility Calculus model and two alternative models.

Experiment 2



Based on this astronaut's path across both maps, rate on a scale of 0-10:

1. White terrain

How easy is it to cross?

2. Gray terrain

How easy is it to cross?

3. Reward A

How much does the astronaut like it?

4. Reward B

How much does the astronaut like it?

Why Experiment 2?

Unlike Experiment 1's single events, real-world observations involve multiple actions across different contexts that let us revise beliefs about agents' costs and rewards.

Experiment 2 tests inference from multiple action events with two purposes:

- Test if inferences over repeated events follow the full Naïve Utility Calculus model
- Test if a simpler single-event model (using first or most recent observation) explains inferences equally well

Processing Human Data to "Ground Truth"

Step 1: Raw Data Collection

- Participants rated on 0-10 scales: terrain difficulty (0=easy, 10=exhausting) and care package preference (0=not at all, 10=a lot)

Step 2: Z-Score Within Each Participant

- Remove individual scale usage differences by standardizing each participant's responses: $z = (\text{raw_response} - \mu_{\text{participant}}) / \sigma_{\text{participant}}$

Step 3: Separate by Question Type

- Split z-scored responses into cost judgments and reward judgments

Step 4: Average Across Participants

- For each trial, average z-scores across all participants to create the "ground truth" score

Discussion Questions

- ① What limitations are there to using humans as the "ground truth" here?
- ② Is it fair to isolate actions that result in costs and rewards? e.g. What changes if terrains have rewards or care packages have costs?

Table of Contents

- 1 Introduction
- 2 Related Work
- 3 Experiments 1 and 2
- 4 Naive Utility Calculus
- 5 Experiment 5
- 6 Results
- 7 Discussion

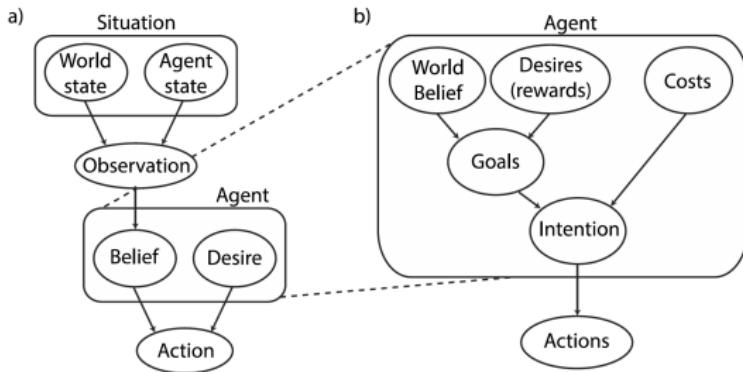
Key Question

How do we design an algorithm that can give us the cost and reward function for each terrain and care package?

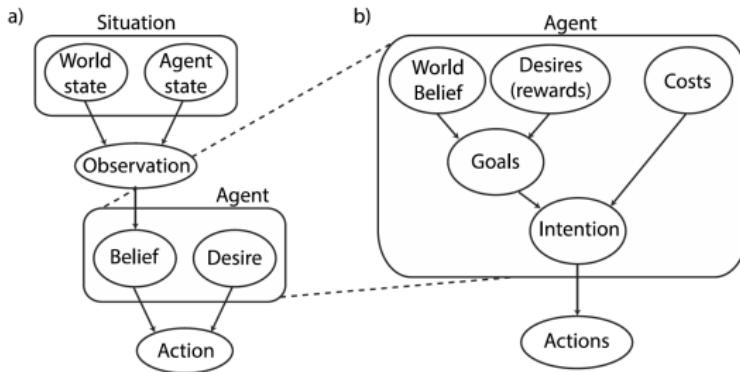
The intuitive solution involves:

- ① **Generation (Forward):** Given cost and reward → predict actions
- ② **Inference (Backward):** Given the actions → Infer the costs / rewards

Generative model

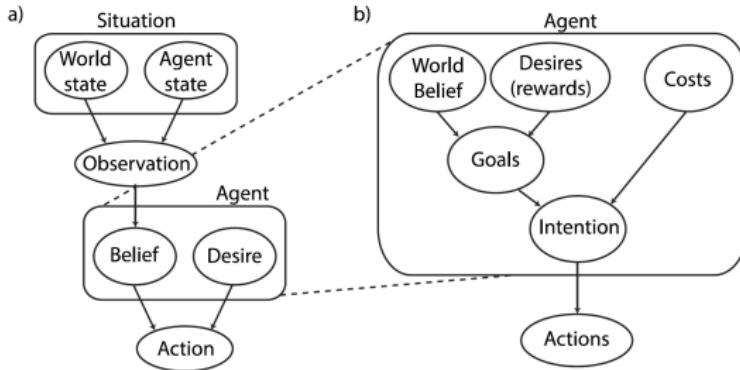


Generative model



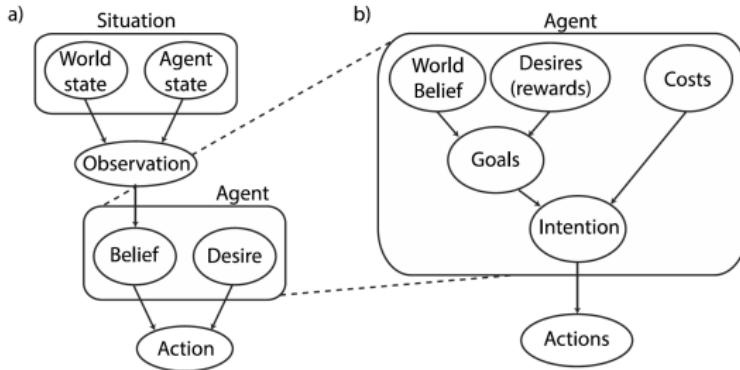
- **Desires = Reward Function**

Generative model



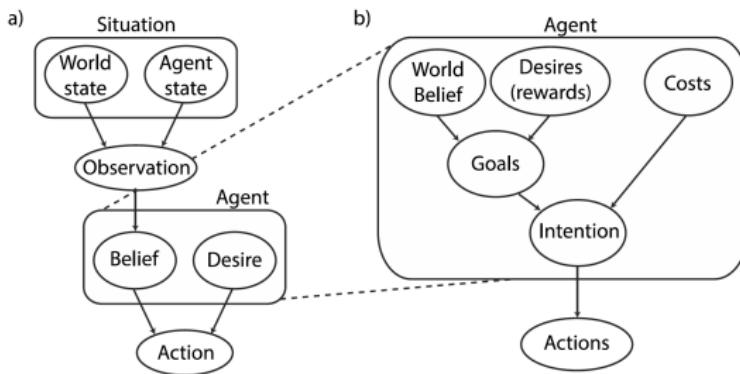
- **Desires** = Reward Function
- **Goals** = States of the world that the agent believes yield rewards

Generative model



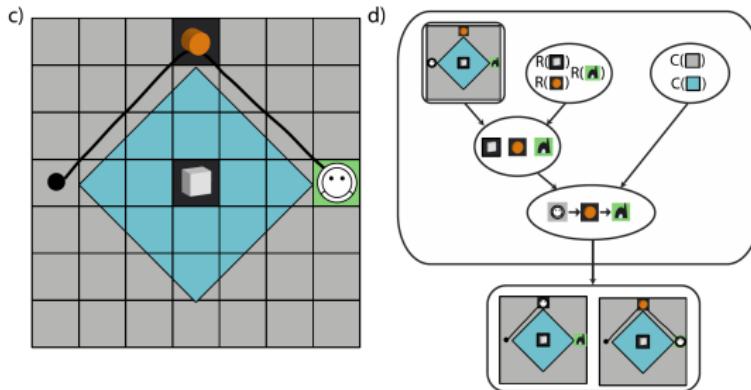
- **Desires** = Reward Function
- **Goals** = States of the world that the agent believes yield rewards
- **Intentions** = Ordered sequences of goals intended to maximize utilities

Generative model

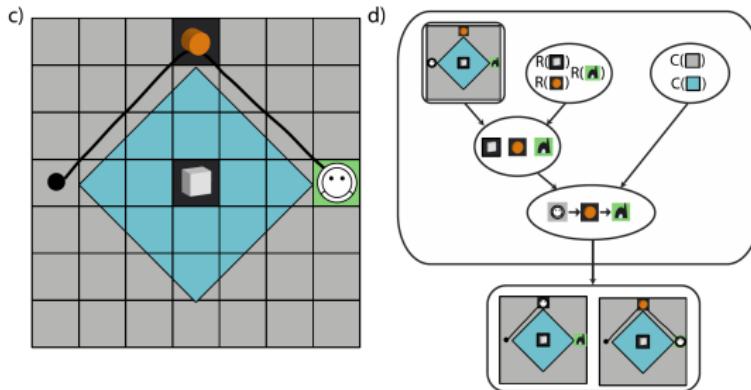


- **Desires** = Reward Function
- **Goals** = States of the world that the agent believes yield rewards
- **Intentions** = Ordered sequences of goals intended to maximize utilities
- **Actions** = mappings from states to action which when executed sequentially fulfill each goal in the intention

Generative Model

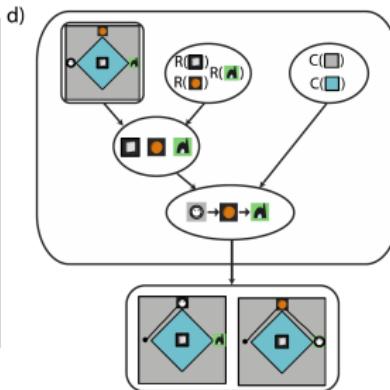
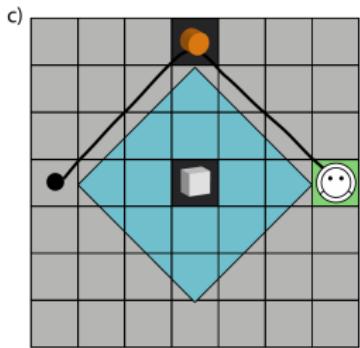


Generative Model



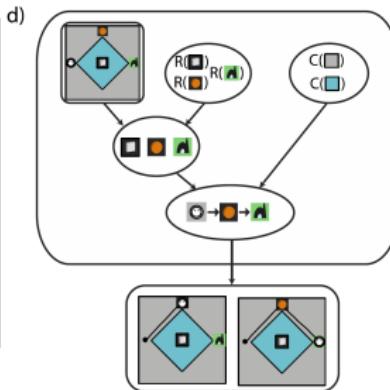
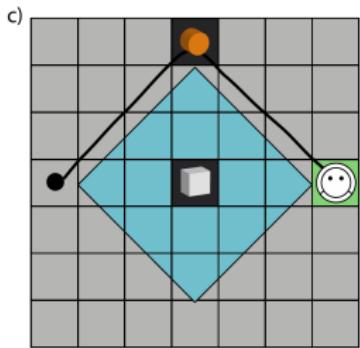
- **Beliefs + Rewards → Candidate goals**

Generative Model



- **Beliefs + Rewards → Candidate goals**
- **Goals + Action costs → Specific intentions to act**

Generative Model



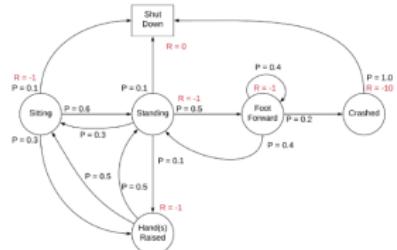
- **Beliefs + Rewards → Candidate goals**
- **Goals + Action costs → Specific intentions to act**
 - Constraint: space station must be final goal
 - Results in 5 possible intentions: direct path, single package routes, dual package routes

Discussion Question

Q1. The NUC model assumes that rational agents undergo a hierarchical planning process: First select an intention (a sequence/plan of goals), then take actions to achieve a goal. Would it be possible to model agents using a "flat" planning process, where agents directly optimize for subjective utility (rewards minus costs), without going through the process of selecting goals? If so, how would that generative model differ in its predictions?

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)



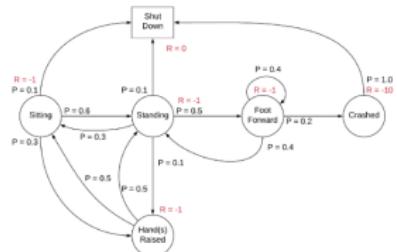
In NUC context:

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)

- **States (S):** Agent positions



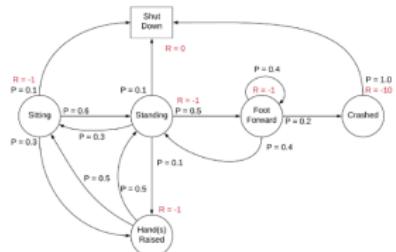
In NUC context:

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)

- **States (S):** Agent positions
- **Actions (A):** 8 movement directions



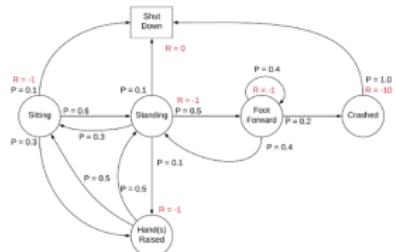
In NUC context:

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)

- **States (S):** Agent positions
- **Actions (A):** 8 movement directions
- **Transition probabilities (P_a):** Not important here (deterministic)



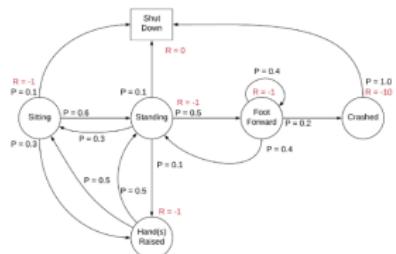
In NUC context:

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)

- **States (S):** Agent positions
- **Actions (A):** 8 movement directions
- **Transition probabilities (P_a):** Not important here (deterministic)
- **Rewards (R_a):** Only at goals



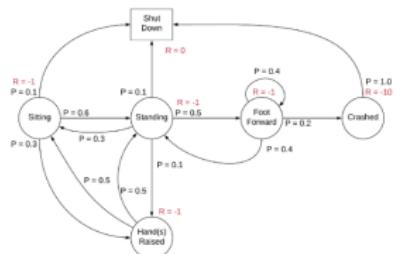
In NUC context:

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Markov Decision Processes (MDPs)

Formal definition: 4-tuple (S, A, P_a, R_a)

- **States (S):** Agent positions
- **Actions (A):** 8 movement directions
- **Transition probabilities (P_a):** Not important here (deterministic)
- **Rewards (R_a):** Only at goals



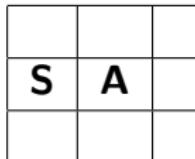
In NUC context:

- Actions are deterministic
- Costs vary by terrain type
- Each goal has its own MDP

Purpose: Compute optimal policies to reach goals efficiently, which observers use to infer agent preferences from observed behavior.

Solving an MDP

3x3 Grid Example:



A = Agent (center)

S = Space Station (goal)

Optimal Solution: Agent moves left

Policy $\pi(s)$:

$$\pi^*(s) = \arg \max_a Q^*(s, a)$$

Action-Value $Q(s, a)$:

$$Q^*(s, a) = R + \gamma \sum_{s'} PV^*(s')$$

Value Function $V(s)$:

$$V^*(s) = \max_a Q^*(s, a)$$

Bellman Equation:

$$V^*(s) = \max_a [R + \gamma \sum_{s'} PV^*(s')]$$

Optimal Policy π^* : Maps each state to best action ("What should I do here?") • **Value Function** V^* : Expected reward from state with optimal policy ("How good is this state?") • **Action-Value** Q^* : Expected reward for specific action then optimal policy ("How good is this action?")

Intentions to Action Policies

How do we "fulfill" an intention?

Answer: Execute goal-specific MDPs for each goal until complete

Example Intention: Orange care package → Space station

- **Step 1:** Solve MDP to reach orange care package
- **Step 2:** Solve MDP from orange package to space station
- Execute these policies sequentially

Softmax Policy (Realistic Behavior):

- People don't always act perfectly - they make mistakes
- **Softmax:** Agents are more likely to choose good actions, less likely to choose bad ones

-

$$p(a|s) \propto \exp(\alpha \cdot \text{Value of action})$$

- $\alpha = 0.1$ (low rationality parameter → nearly optimal behavior)

Inference over the Generative Model

The Problem: We observe actions, but need to infer hidden costs and rewards

The Solution: Bayesian inference - "work backwards" from actions

$$p(\text{Costs, Rewards} | \text{Actions}) \propto p(\text{Actions} | \text{Costs, Rewards}) \times p(\text{Costs, Rewards})$$

How it works:

- **Prior:** What costs/rewards are plausible? (uniform for novel terrains)
- **Likelihood:** How probable are these actions given specific costs/rewards?
- **Posterior:** Which costs/rewards best explain the observed actions?

Two levels of rationality:

- **Rational choice:** Pick intentions that maximize utility
- **Rational action:** Execute actions that fulfill intentions efficiently

Implementation: Monte Carlo sampling - test many cost/reward combinations and see which ones make the observed actions most likely

Table of Contents

1 Introduction

2 Related Work

3 Experiments 1 and 2

4 Naive Utility Calculus

5 Experiment 5

6 Results

7 Discussion

Recap: Research Questions

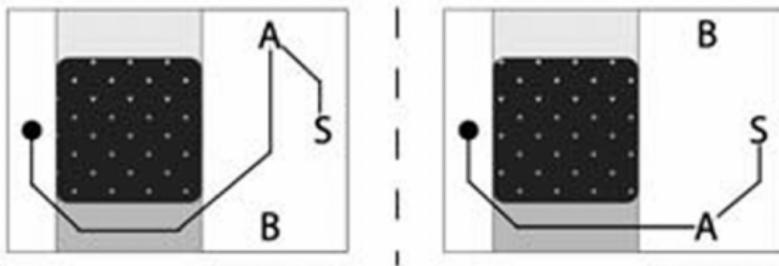
- **RQ1:** Can NUC support joint inference of costs and rewards when we know neither, using a coherent generative model?
- **RQ2:** Does NUC drive fine-grained quantitative inferences or only coarse qualitative ones?
- **RQ3:** Is NUC a unified generative model supporting probabilistic inference, or a collection of simple heuristics?

Experiment 1: Compare model predictions with human joint inferences of agent's costs and rewards

Experiment 2: Same comparison when multiple events are combined to make inferences

Experiment 5: Test how participants infer whether an agent is knowledgeable or ignorant (answers RQ3)

Experiment 5: Knowledge vs. Ignorance



Three Questions Asked:

- ① **Attention check:** "Did the position of the care packages switch on the second day?" (Must answer correctly to proceed)
- ② **Knowledge of rewards:** "Did the astronaut already know the rewards on day 1?"
- ③ **Knowledge of costs:** "Did the astronaut already know the costs on day 1?"

Response scale: 0-10 ("Definitely did not" → "Maybe" → "Definitely did")

Table of Contents

1 Introduction

2 Related Work

3 Experiments 1 and 2

4 Naive Utility Calculus

5 Experiment 5

6 Results

7 Discussion

Alternative Models

1. NUC Variations:

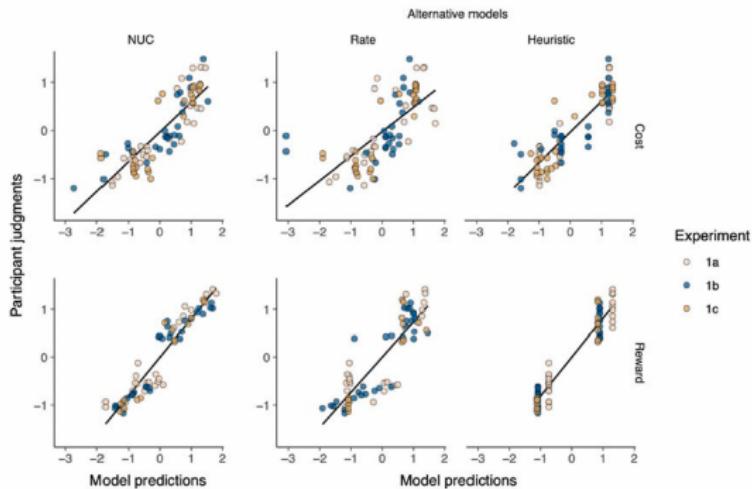
- **Rate-based model:** $U = R/C$ (maximize reward rate) instead of $U = R - C$
- **Single-event model:** Use only first or most recent observation instead of integrating all events
- **Basic goal attribution:** Simple inverse planning without cost-benefit analysis

2. Domain-Specific Heuristics:

- **Time-based:** Infer costs from time spent in each terrain
- **Collection-based:** Infer rewards from which objects are collected
- **Change-based:** Associate behavior changes with ignorance
- **Action-based:** Judge traits solely from specific actions (helping = nice)

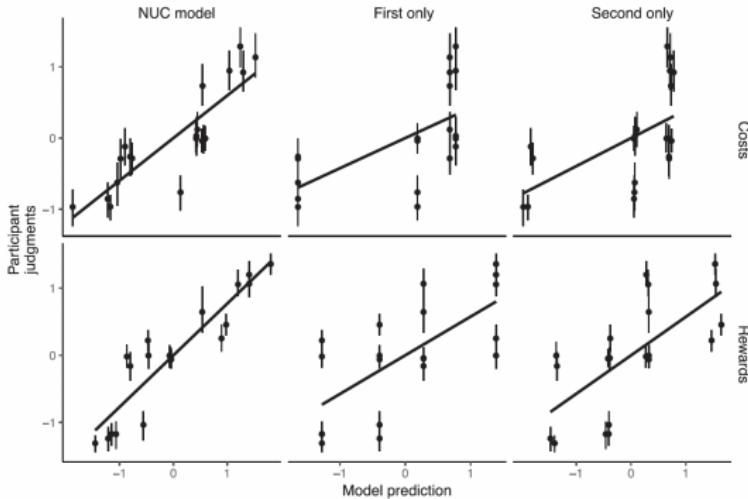
Goal: Test if NUC captures fine-grained inferences that simpler alternatives miss

Experiment 1 Results



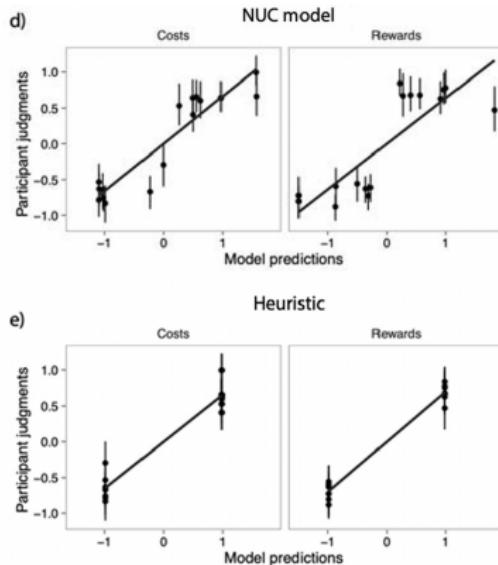
- **NUC model:** High correlations (0.85 costs, 0.95 rewards)
- **Rate model:** Lower correlations (0.70 costs, 0.87 rewards)
- **Heuristic:** High overall correlation but missed fine-grained patterns
- **Conclusion:** People make graded inferences consistent with utility maximization, not just simple heuristics

Experiment 2 Results



- **Full NUC model:** High correlations (0.86 costs, 0.92 rewards)
- **First-only model:** Lower correlations (0.61 costs, 0.69 rewards)
- **Second-only model:** Lower correlations (0.57 costs, 0.68 rewards)
- **Conclusion:** People integrate information from multiple events for more precise inferences

Experiment 5 Results



- **NUC model:** High correlations (0.93 costs, 0.83 rewards)
- **Simple heuristic:** High overall correlation (0.97 costs, 0.99 rewards)
but predicts only bimodal judgments
- **Graded sensitivity:** Participants showed nuanced inferences within heuristic clusters

Table of Contents

1 Introduction

2 Related Work

3 Experiments 1 and 2

4 Naive Utility Calculus

5 Experiment 5

6 Results

7 Discussion

Discussion

- Experiment 5 tests whether the NUC model can infer whether agents are initially ignorant or knowledgeable about their costs or rewards (i.e. their desires). However, the paper itself doesn't describe how knowledge/ignorance about one's own desires is accounted for in the generative model. How do you think the generative model presented in Section 2 could be modified to account for knowledge/ignorance of one's own rewards or costs?