The Theory of Learning AI Agents that Approximate Human Behaviour

Soumadeep Ghosh

Kolkata, India

Abstract

This paper presents a comprehensive theoretical framework for understanding how artificial intelligence agents can learn to approximate human behavior. We synthesize insights from machine learning, cognitive science, neuroscience, behavioral economics, and psychology to develop formal models of human behavior approximation. We introduce the concept of *behavioral fidelity measures*, present novel learning architectures that incorporate cognitive biases and heuristics, and analyze the fundamental trade-offs between computational efficiency and behavioral accuracy. Our framework encompasses imitation learning, inverse reinforcement learning, and cognitive architecture-based approaches, providing both theoretical guarantees and practical algorithms. We demonstrate that successful human behavior approximation requires modeling not only rational decision-making but also systematic deviations from rationality that characterize human cognition.

The paper ends with "The End"

1 Introduction

The quest to develop artificial intelligence systems that can understand, predict, and replicate human behavior represents one of the most ambitious challenges in contemporary AI research. Unlike traditional optimization problems where the objective is clearly defined, approximating human behavior requires grappling with the inherent complexity, inconsistency, and context-dependence of human decision-making [9, 12].

1.1 Motivation and Scope

Human behavior emerges from a sophisticated interplay of neural processes, cognitive mechanisms, emotional states, social contexts, and learned experiences. Any AI system attempting to approximate such behavior must address several fundamental questions:

1. **What aspects of behavior should be modeled?** Human actions result from both deliberative reasoning and automatic processes, involving perception, memory, planning, and motor control.

2. **How can we formalize behavioral similarity?** Defining metrics that capture meaningful correspondence between human and artificial agent behavior is non-trivial.

3. **What learning paradigms are appropriate?** Different approaches - supervised learning, reinforcement learning, imitation learning - offer distinct advantages and limitations.

4. **How do we incorporate irrationality?** Human behavior systematically deviates from normative rationality in ways that must be modeled explicitly.

1.2 Theoretical Foundations

We ground our framework in three complementary perspectives:

Computational Rationality: The principle that human behavior, while not perfectly optimal, represents bounded rational solutions given computational constraints [6, 7].

Predictive Processing: The brain as a hierarchical prediction machine that minimizes prediction errors through active inference [3, 5].

Dual-Process Theory: The distinction between fast, automatic (System 1) and slow, deliberative (System 2) cognitive processes [4, 9].

2 Formal Framework

2.1 Mathematical Preliminaries

Let $\mathcal{S}$ denote a state space, $\mathcal{A}$ an action space, and $\mathcal{O}$ an observation space. We model human behavior as a stochastic policy $\pi_H : \mathcal{S} \times \mathcal{A} \to [0, 1]$ operating in an environment characterized by transition dynamics $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$.

Definition 2.1 (Behavioral Trajectory)**.** *A behavioral trajectory is a sequence*

$$\tau = (s_0, a_0, r_0, s_1, a_1, r_1, \ldots, s_T)$$

where $s_t \in \mathcal{S}$, $a_t \sim \pi_H(\cdot|s_t)$, $r_t = R(s_t, a_t)$, and $s_{t+1} \sim P(\cdot|s_t, a_t)$.

Definition 2.2 (Behavioral Fidelity)**.** *Given a learned agent policy π_A and human policy π_H, the behavioral fidelity is quantified by:*

$$\Phi(\pi_A, \pi_H) = \mathbb{E}_{s \sim \rho_{\pi_H}} \left[D_{KL}(\pi_H(\cdot|s) \| \pi_A(\cdot|s)) \right] \tag{1}$$

where ρ_{π_H} is the state visitation distribution under π_H, and D_{KL} denotes Kullback-Leibler divergence.

2.2 The Human Behavior Approximation Problem

Formally, we seek to find an agent policy π_A^* that minimizes behavioral divergence:

$$\pi_A^* = \arg\min_{\pi_A \in \Pi} \Phi(\pi_A, \pi_H) + \lambda \mathcal{C}(\pi_A) \tag{2}$$

where Π is a hypothesis class of policies, $\mathcal{C}(\pi_A)$ represents computational complexity, and λ balances fidelity against efficiency.

`The following space was deliberately left blank.`

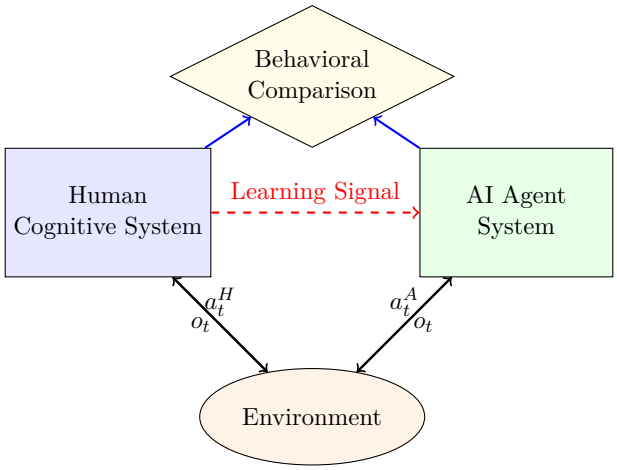

Figure 1: Conceptual architecture for human behavior approximation. The AI agent learns from human demonstrations and feedback while both interact with a shared environment.

3 Learning Paradigms

3.1 Imitation Learning

Imitation learning aims to derive a policy from expert demonstrations $\mathcal{D} = \{\tau_1, \ldots, \tau_N\}$ where τ_i are trajectories from π_H.

Behavioral Cloning: The simplest approach treats imitation as supervised learning:

$$\pi_A^* = \arg\min_{\pi_A} \sum_{i=1}^{N} \sum_{t=0}^{T_i} \ell(\pi_A(\cdot|s_t^i), a_t^i) \tag{3}$$

where ℓ is a loss function (e.g., cross-entropy for discrete actions).

Limitations: Behavioral cloning suffers from distribution shift - errors compound as the agent encounters states absent from the training distribution [11].

Theorem 3.1 (Behavioral Cloning Error Bound)**.** *Let ϵ be the expected error of π_A on the training distribution, and T the horizon length. The expected error under π_A's induced distribution is bounded by:*

$$\mathbb{E}_{\tau\sim\pi_A}[errors] \leq \epsilon T + T(T-1)\epsilon^2 \tag{4}$$

3.2 Inverse Reinforcement Learning

IRL recovers the reward function R that rationalizes observed behavior, assuming the human acts near-optimally [1, 14].

Maximum Entropy IRL: Models human behavior as maximizing entropy-regularized expected return:

$$\pi_H = \arg\max_{\pi} \mathbb{E}_{\tau\sim\pi}[R(\tau)] + \alpha\mathcal{H}(\pi) \tag{5}$$

where $\mathcal{H}(\pi) = -\mathbb{E}_{\tau\sim\pi}[\log \pi(\tau)]$ is the policy entropy.

The resulting policy has the form:

$$\pi_H(a|s) \propto \exp\left(\frac{1}{\alpha} Q^*(s,a)\right) \tag{6}$$

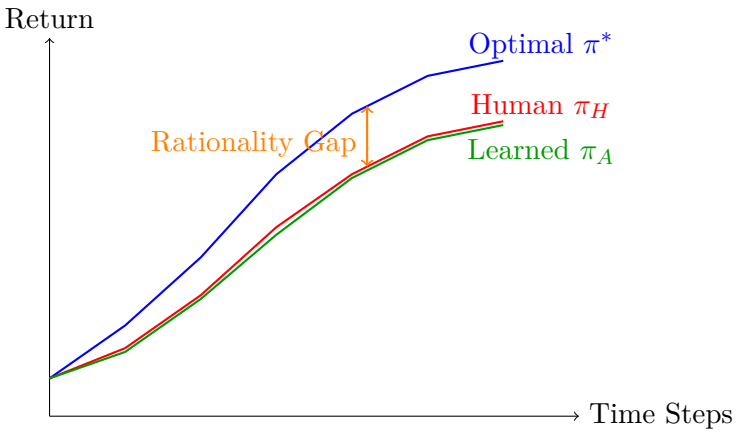

Figure 2: Performance comparison between optimal, human, and learned policies. Human behavior exhibits systematic suboptimality that must be captured by the learned agent.

3.3 Cognitive Architecture-Based Learning

Rather than learning end-to-end mappings, we can incorporate structured cognitive components:

$$\pi_A(a|s) = \int_\Theta \pi_{\text{motor}}(a|\theta) \cdot p_{\text{planning}}(\theta|s) \cdot p_{\text{memory}}(s|h)\, d\theta \tag{7}$$

where θ represents internal cognitive states, and the architecture decomposes into perception, memory, planning, and motor control modules.

4 Modeling Human Irrationality

4.1 Cognitive Biases as Features

Human decision-making exhibits systematic deviations from rationality that must be explicitly modeled:

Loss Aversion: Losses loom larger than gains. We model this via prospect theory [8]:

$$v(x) = \begin{cases} x^\alpha & \text{if } x \geq 0 \\ -\lambda(-x)^\beta & \text{if } x < 0 \end{cases} \tag{8}$$

where $\lambda > 1$ captures loss aversion intensity.

Temporal Discounting: Humans exhibit hyperbolic rather than exponential discounting [10]:

$$D(t) = \frac{1}{1 + kt} \tag{9}$$

Confirmation Bias: Preferential weighting of belief-consistent evidence:

$$p_{\text{update}}(h|e) \propto p(h) \cdot p(e|h)^{w(h,e)} \tag{10}$$

where $w(h, e) > 1$ if e confirms h, and $w(h, e) < 1$ otherwise.

4.2 Bounded Rationality Framework

Following Simon's bounded rationality [12], we model decision-making as resource-constrained optimization:

$$\pi_H = \arg\max_{\pi \in \Pi_c} \mathbb{E}_{\tau \sim \pi}[R(\tau)] \quad \text{subject to} \quad C(\pi) \leq B \tag{11}$$

where Π_c is the set of cognitively realizable policies, $C(\pi)$ measures computational cost, and B is the cognitive budget.

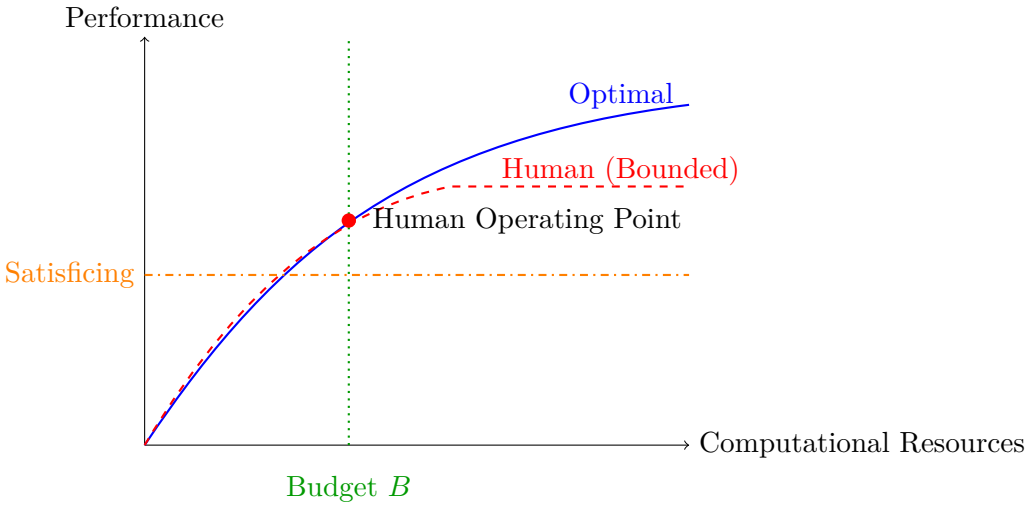

Figure 3: Bounded rationality framework showing how humans operate under computational constraints, achieving satisficing rather than optimal performance.

5 Hierarchical Models of Behavior

5.1 Hierarchical Reinforcement Learning

Human behavior exhibits hierarchical structure with goals, subgoals, and primitives. We model this using options framework [13]:

Definition 5.1 (Option)**.** *An option* $\omega = (I_\omega, \pi_\omega, \beta_\omega)$ *consists of:*

- *Initiation set* $I_\omega \subseteq \mathcal{S}$
- *Option policy* $\pi_\omega : \mathcal{S} \times \mathcal{A} \to [0, 1]$
- *Termination condition* $\beta_\omega : \mathcal{S} \to [0, 1]$

The hierarchical value function decomposes as:

$$V^{\pi_H}(s) = \mathbb{E}_{\omega \sim \mu(\cdot|s)}\left[Q_\omega^{\pi_H}(s, \omega)\right] \tag{12}$$

where μ is a meta-policy over options, and:

$$Q_\omega^{\pi_H}(s, \omega) = \mathbb{E}\left[\sum_{t=0}^{\tau_\omega - 1} \gamma^t r_t + \gamma^{\tau_\omega} V^{\pi_H}(s_{\tau_\omega}) \mid s_0 = s, \omega\right] \tag{13}$$

5.2 Bayesian Theory of Mind

Modeling human behavior requires inferring hidden mental states - beliefs, desires, intentions. We employ Bayesian Theory of Mind [2]:

$$p(\text{beliefs}, \text{desires}|\text{actions}) \propto p(\text{actions}|\text{beliefs}, \text{desires}) \cdot p(\text{beliefs}, \text{desires}) \tag{14}$$

The generative model assumes:

$$\text{desires} \sim p_{\text{prior}}(\cdot) \tag{15}$$

$$\text{beliefs} \sim p(\cdot|\text{observations}) \tag{16}$$

$$\text{actions} \sim \pi_{\text{plan}}(\cdot|\text{beliefs}, \text{desires}) \tag{17}$$

6 Neural Architecture for Behavior Approximation

6.1 Deep Learning Approaches

Modern approaches employ deep neural networks with specialized architectures:

Perception Module: Convolutional or transformer-based encoder:

$$h_{\text{percept}} = f_{\text{encode}}(o_t; \theta_{\text{percept}}) \tag{18}$$

Memory Module: Recurrent or attention-based memory:

$$h_{\text{mem},t} = \text{LSTM}(h_{\text{mem},t-1}, h_{\text{percept}}; \theta_{\text{mem}}) \tag{19}$$

Planning Module: Model-based lookahead or value estimation:

$$V_t = f_{\text{value}}(h_{\text{mem},t}; \theta_{\text{value}}) \tag{20}$$

Action Selection: Stochastic policy head:

$$\pi_A(a|o_t) = \text{softmax}\left(f_{\text{policy}}(h_{\text{mem},t}; \theta_{\text{policy}})\right) \tag{21}$$

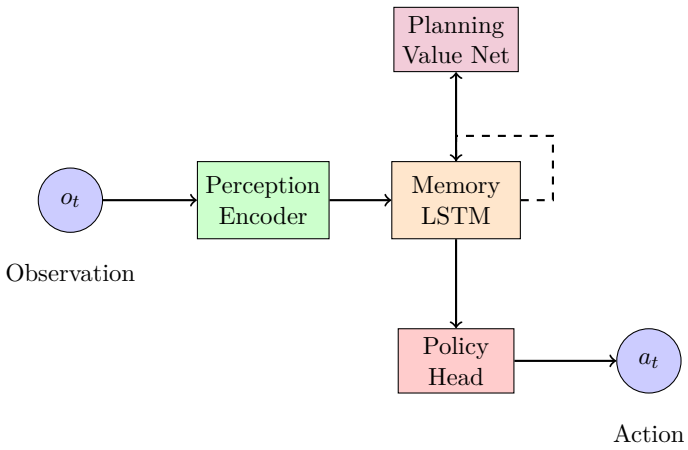

Figure 4: Neural architecture for behavior approximation with perception, memory, planning, and policy modules.

6.2 Attention and Meta-Learning

To capture context-dependent behavior adaptation:

Attention Mechanism: Selectively weight relevant features:

$$\alpha_i = \frac{\exp(e_i)}{\sum_j \exp(e_j)}, \quad e_i = h_{\text{query}}^\top W_{\text{attn}} h_{\text{key},i} \tag{22}$$

Meta-Learning: Learn to adapt quickly to new behavioral contexts:

$$\theta_{\text{adapted}} = \theta - \alpha \nabla_\theta \mathcal{L}_{\text{task}}(\theta; \mathcal{D}_{\text{support}}) \tag{23}$$

where $\mathcal{D}_{\text{support}}$ are few-shot demonstrations from a new human or context.

7 Theoretical Guarantees

7.1 Sample Complexity

Theorem 7.1 (Sample Complexity Bound)**.** *Let $\mathcal{H}$ be a hypothesis class with VC dimension d. To learn a policy π_A that achieves behavioral fidelity $\Phi(\pi_A, \pi_H) \leq \epsilon$ with probability at least $1 - \delta$, the number of required samples is:*

$$N = O\left(\frac{d + \log(1/\delta)}{\epsilon^2}\right) \tag{24}$$

7.2 Generalization Bounds

Theorem 7.2 (PAC-Bayesian Bound for Behavioral Approximation)**.** *For any prior P over policies and any $\delta > 0$, with probability at least $1 - \delta$ over the training data $\mathcal{D}$, for all posteriors Q:*

$$\mathbb{E}_{\pi \sim Q}[\Phi(\pi, \pi_H)] \leq \hat{\mathbb{E}}_{\pi \sim Q}[\Phi(\pi, \pi_H)] + \sqrt{\frac{D_{KL}(Q \| P) + \log(2N/\delta)}{2(N-1)}} \tag{25}$$

where $\hat{\mathbb{E}}$ denotes empirical expectation and $N = |\mathcal{D}|$.

8 Experimental Considerations

8.1 Evaluation Metrics

Beyond behavioral fidelity Φ, comprehensive evaluation requires:

Action Matching Rate:

$$\text{AMR} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}[a_i^A = a_i^H] \tag{26}$$

Trajectory Distribution Distance:

$$\text{TDD} = W_2(\rho_{\pi_A}, \rho_{\pi_H}) \tag{27}$$

where W_2 is the 2-Wasserstein distance between trajectory distributions.

Goal Achievement Similarity:

$$\text{GAS} = \frac{|\mathcal{G}_A \cap \mathcal{G}_H|}{|\mathcal{G}_H|} \tag{28}$$

where $\mathcal{G}_A$ and $\mathcal{G}_H$ are sets of goals achieved by agent and human.

8.2 Benchmark Domains

Effective benchmarks for human behavior approximation include:

- **Game Playing:** Complex strategic environments (chess, StarCraft) with rich human datasets
- **Navigation:** Indoor/outdoor navigation with varied human strategies
- **Manipulation:** Object interaction tasks requiring dexterous control
- **Social Interaction:** Multi-agent scenarios requiring theory of mind
- **Economic Games:** Revealing systematic biases and heuristics

9 Challenges and Open Problems

9.1 Individual Differences

Humans exhibit vast individual variation in behavior. Modeling requires:

$$\pi_H(a|s, \psi) = f_{\text{policy}}(s; \theta, \psi) \tag{29}$$

where ψ represents individual characteristics (personality, skills, preferences).

9.2 Temporal Consistency vs. Adaptation

Balancing consistency over time with adaptation to new information:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{fidelity}} + \lambda_1 \mathcal{L}_{\text{consistency}} + \lambda_2 \mathcal{L}_{\text{adaptation}} \tag{30}$$

9.3 Interpretability

Understanding *why* an agent behaves as it does is crucial for trust and debugging:

$$\text{Explanation}(a, s) = \arg\max_{e \in \mathcal{E}} p(e|a, s) \cdot \text{Simplicity}(e) \tag{31}$$

10 Applications

10.1 Human-AI Collaboration

Agents that approximate human behavior can better collaborate by:

- Predicting human actions and intentions
- Adapting their behavior to human preferences
- Communicating in human-understandable ways

10.2 Training Simulations

Realistic virtual humans for training scenarios (medical, military, customer service).

10.3 Computational Social Science

Simulating populations for policy analysis:

$$\text{Outcome(policy)} = \int_{\Psi} \mathbb{E}_{\pi_H(\cdot|\psi)}[\text{reward(policy)}]p(\psi)d\psi \tag{32}$$

11 Ethical Considerations

Developing AI that approximates human behavior raises important ethical questions:

- **Manipulation Risk:** Systems that predict human behavior could be exploited for manipulation
- **Privacy:** Learning from behavioral data requires careful privacy protection
- **Bias Amplification:** Models may learn and perpetuate human biases
- **Autonomy:** Over-personalization may reduce human agency
- **Transparency:** Users should understand when interacting with behavioral models

12 Conclusion

We have presented a comprehensive theoretical framework for understanding how AI agents can learn to approximate human behavior. Our approach synthesizes insights across computer science, cognitive science, neuroscience, and behavioral economics to provide both formal foundations and practical algorithms.

Key contributions include:

1. Formal definitions of behavioral fidelity and approximation objectives
2. Integration of cognitive biases and bounded rationality into learning algorithms
3. Hierarchical and modular architectures reflecting human cognitive structure
4. Theoretical guarantees on sample complexity and generalization
5. Analysis of fundamental trade-offs between accuracy and efficiency

Future work must address the challenges of individual differences, contextual adaptation, and interpretability while carefully considering the ethical implications of increasingly sophisticated behavioral models. As AI systems become more deeply integrated into society, the ability to approximate and predict human behavior will become ever more important - demanding rigorous theoretical foundations coupled with responsible deployment practices.

The ultimate goal is not merely to mimic human actions, but to understand the underlying cognitive processes that generate behavior, enabling AI systems that can truly collaborate with humans as partners rather than mere tools.

References

[1] Abbeel, P., & Ng, A. Y. (2004). *Apprenticeship learning via inverse reinforcement learning.* Proceedings of the 21st International Conference on Machine Learning (ICML).

[2] Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). *Action understanding as inverse planning.* Cognition, 113(3), 329-349.

[3] Clark, A. (2013). *Whatever next? Predictive brains, situated agents, and the future of cognitive science.* Behavioral and Brain Sciences, 36(3), 181-204.

[4] Evans, J. S. B., & Stanovich, K. E. (2013). *Dual-process theories of higher cognition: Advancing the debate.* Perspectives on Psychological Science, 8(3), 223-241.

[5] Friston, K. (2010). *The free-energy principle: A unified brain theory?* Nature Reviews Neuroscience, 11(2), 127-138.

[6] Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). *Computational rationality: A converging paradigm for intelligence in brains, minds, and machines.* Science, 349(6245), 273-278.

[7] Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). *Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic.* Topics in Cognitive Science, 7(2), 217-229.

[8] Kahneman, D., & Tversky, A. (1979). *Prospect theory: An analysis of decision under risk.* Econometrica, 47(2), 263-291.

[9] Kahneman, D. (2011). *Thinking, Fast and Slow.* Farrar, Straus and Giroux.

[10] Laibson, D. (1997). *Golden eggs and hyperbolic discounting.* Quarterly Journal of Economics, 112(2), 443-478.

[11] Ross, S., Gordon, G., & Bagnell, D. (2011). *A reduction of imitation learning and structured prediction to no-regret online learning.* Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS).

[12] Simon, H. A. (1955). *A behavioral model of rational choice.* Quarterly Journal of Economics, 69(1), 99-118.

[13] Sutton, R. S., Precup, D., & Singh, S. (1999). *Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning.* Artificial Intelligence, 112(1-2), 181-211.

[14] Ziebart, B. D., Maas, A., Bagnell, J. A., & Dey, A. K. (2008). *Maximum entropy inverse reinforcement learning.* Proceedings of the 23rd AAAI Conference on Artificial Intelligence.

The End