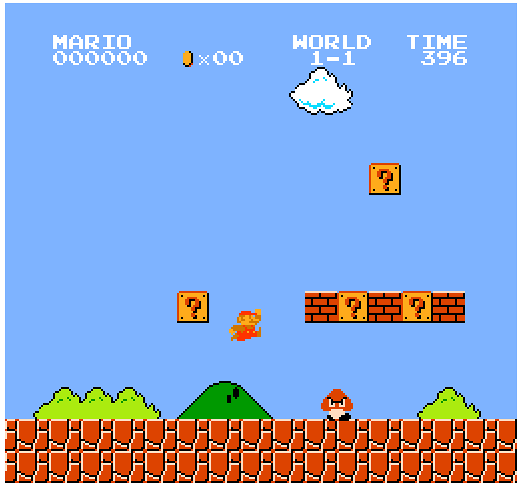
Cautious Curiosity: A Novel Approach to a Human-Like Gameplay Agent

Zhou, C., Machado, T., Harteveld, C. (2023)

Proceedings of the *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*

Abstract

- Authors compared agents using “Super Mario Bros”.
- Authors asked participants to watch videos that agent or human is playing 1-1, and to evaluate how human-like they are.
- As a result, authors succeed in presenting a new reward function that leads to **human-like** behaving



Motivation

- Game AI agents are usually trained to win as fast (correct) as possible.
- This leads to **unhuman-like** behavior (skipping content, ingoring enemies, etc.).
- Developers need **human-like** testers to evaluate player experience and game balance.
- Goal: Creating an agent that balances human-likeness and competitiveness.

Related Works

- **Human-like Agents:** mimic average players, not mastes or novices.
- **Curiosity-based RL:** agents motivated by information gain.
- **Information Gap Theory:** curiosity arises when there’s a gap between what we know and what we want to know.
- **Inverted-U Relationship:** too much or too little gap reduces curiosity.

Methods

- Built on ICM + new reward function.
- Reward increases with prediction error until a threshold ψ
- Beyond ψ, reward decreases (This is “Cautious” behavior)
- Encourages exploring familiar but slightly uncertain areas, not random ones.

Reward function

$$\min_{\theta_I,\theta_P,\theta_F} [-\lambda E_{\pi(S_t;\theta_p)}[\textstyle\sum_t r_t] + (1-\beta)L_I + \beta L_F] \quad (12)$$

Normal ICM

$$r_t^i = \frac{\eta}{2}\|\hat{\phi}(s_{t+1}) - \phi(s_{t+1})\|_2^2$$

Curious curiosity

When the prediction error is smaller than the threshold ψ_t, the reward will be:

$$r_t^i = \eta * ||\widehat{\phi}(S_{t+1}) - \phi(S_{t+1})||_2^2 \quad (10)$$

where η is a scaling factor larger than 0. When the prediction is larger than the threshold, we have:

$$r_t^i = -\eta * ||\widehat{\phi}(S_{t+1}) - \phi(S_{t+1})||_2^2 + 2 * \psi_t \quad (11)$$

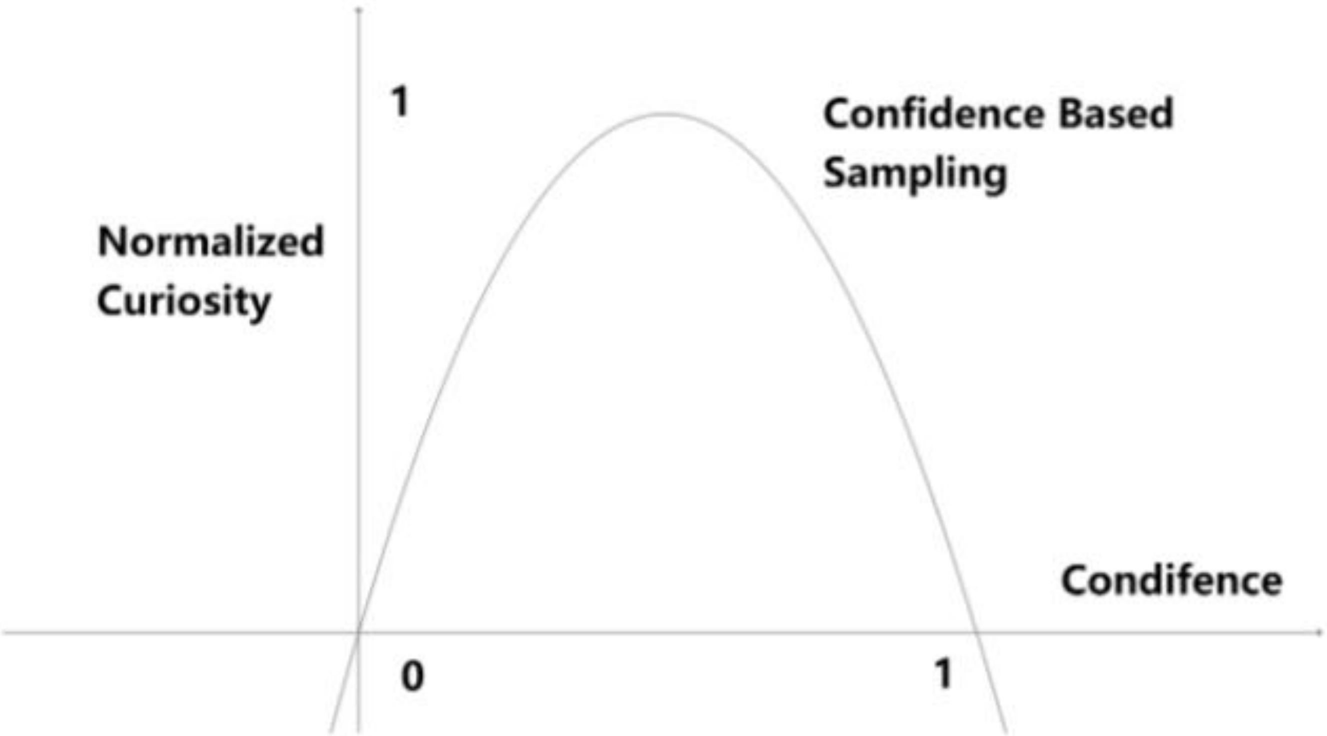
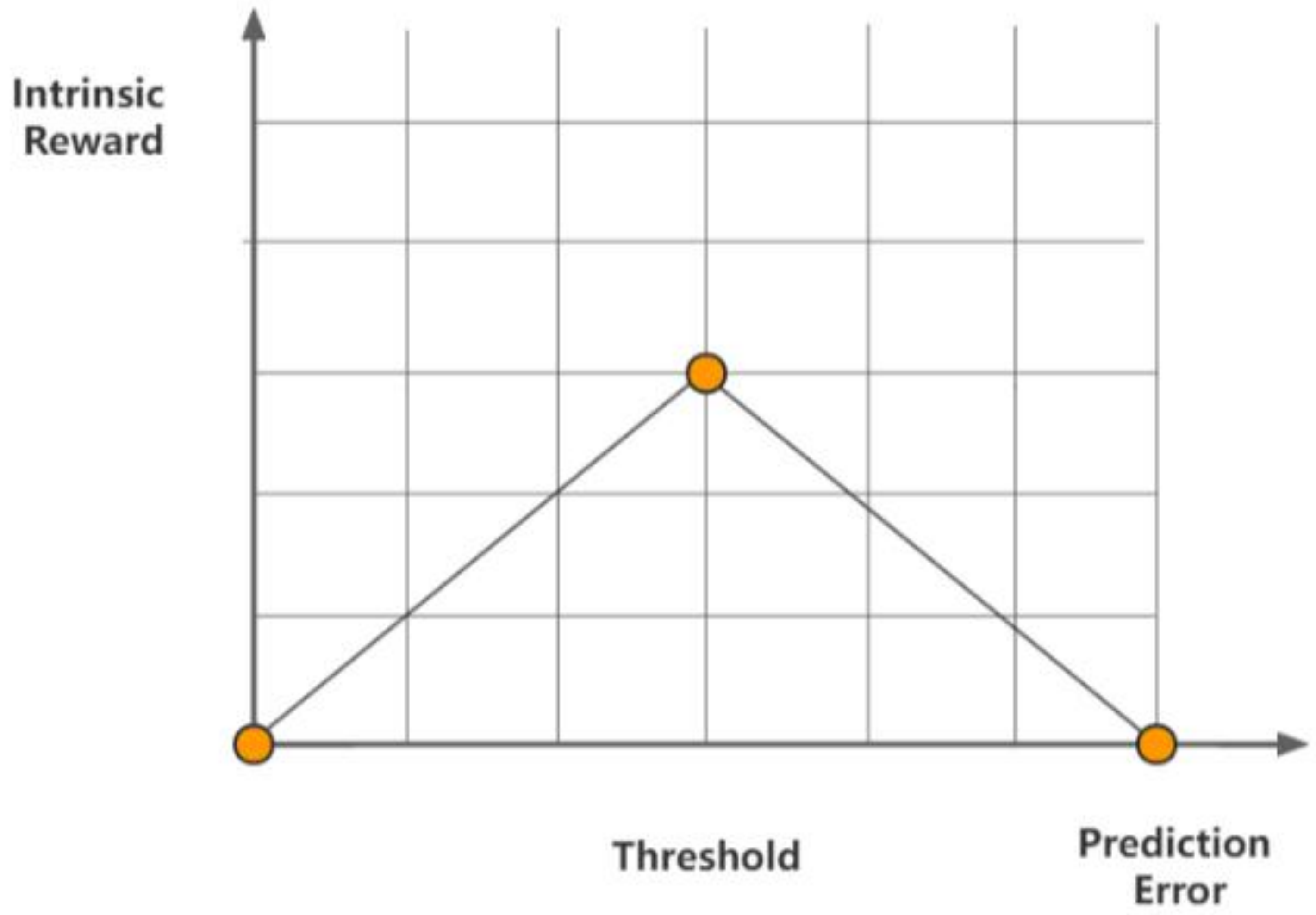


Figure 1: Relationship between confidence (i.e., information gap) and curiosity based on Dubey et al. (Dubey and Griffiths 2017).


Intrinsic Reward
Threshold
Prediction Error

Agents compared

- CC (Cautious Curious)
- ICM(Intrinsic Curiosity Module)
- A3C(Asynchronous Advantages Actor Critic algorithm)
- Human players (average of 3 players who is not beginner)
- Random agent

How to train

- All agents were trained 50 episodes in 1-1 (each episode representing a complete game trajectory).
- Visual input: RGB images → grayscale 84 x 84

Benchmarking

Name	Pass the game	Time used, in M(SD)	Mistakes, in M (SD)
CC	Yes	65 (2.82)	4 (0)
ICM	Yes	58.3 (0.94)	1 (0.81)
A3C	Yes	60 (0)	1.33 (0.47)
Human	Yes	96 (0.94)	3.67 (0.47)
Random	No	400 (= time limit)	

Table 1: Average time used and average mistakes made by each player.

	A3C	Human	ICM	Random	CC
A3C	1	0.025	0.18	1.9e-5	0.16
Human	0.025	1	0.002	2.9e-10	0.16
ICM	0.18	0.002	1	7.1e-4	0.03
Random	1.9e-5	2.9e-10	7.1e-4	1	1.05e-7
CC	0.16	0.16	0.03	1.05e-7	1

Table 3: Details of the post hoc dunn test

How to evaluate

After watching 15 videos, (5 models x 3 videos)

Q1. Please watch this video then rank the human-likeness of the player. (From 0 to 10, 0 means the player is not a human at all and 10 means the player is a real human.)

Q2. State briefly (in one or two sentences) why you think the player above is a human or an AI player.

Quantitative: Human-Likeness Evaluation

	Video 1	Video 2	Video 3
CC	7 (5)	7 (6)	7 (5)
ICM	5 (6)	6 (6)	4.5 (6)
A3C	5 (6.25)	6 (6)	7 (5)
Human	7 (5.5)	7 (5)	7 (5)
Random	3 (8)	4 (8)	3 (8)

Table 2: Median of ratings (0 to 10), in *Mdn* (*IQR*)

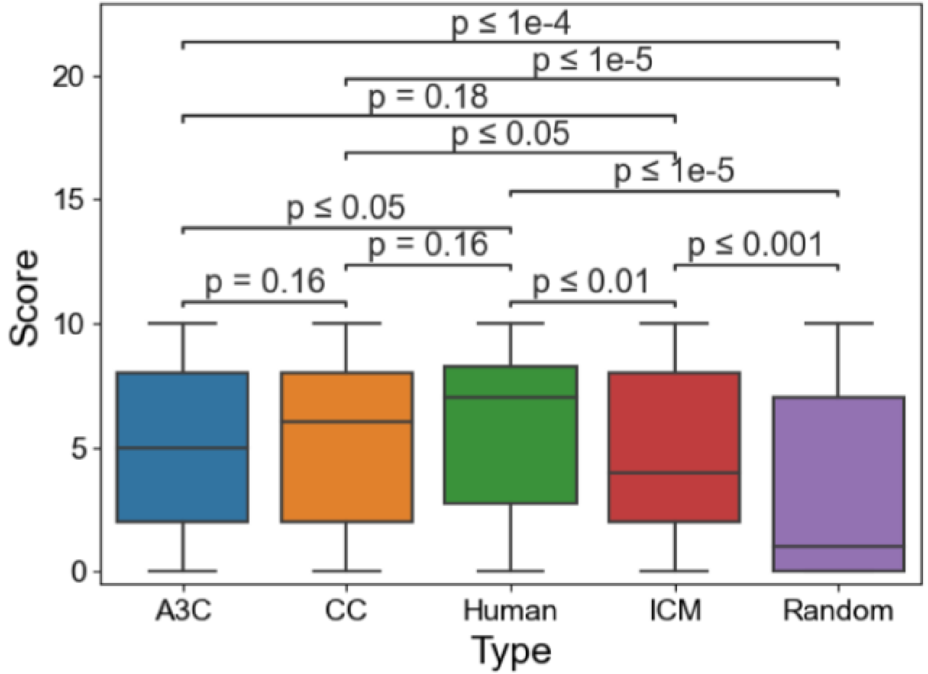


Figure 3: Boxplots of the user study data.

Keyword	Number of occurrence (133 in total)	Distribution $(CC \vert ICM \vert Human \vert A3C \vert Random)$
Mistakes	59 (44%)	19|11|16|13|0
Normal speed and reaction	41 (31%)	9|11|15|5|1
Poor skill	14 (11%)	1|5|2|1|5
Hesitation	8 (6%)	4|3|0|1|0
Others	11 (8%)	4|4|3|0|0

*Others involve overaggressive, unpredictable, and different game patterns.

Table 4: The frequency of top qualitative keywords that identify that the player is a Human.

Keyword	Number of occurrence (341 in total)	Distribution ($CC\|ICM\|$ $Human\|A3C\|Random$)
Fast speed and reaction	127 (37%)	26|40|18|42|1
Useless Movement	105 (31%)	10|9|1|2|83
Perfect skill	50 (15%)	5|14|13|18|0
Ignore game content	39 (11%)	9|8|9|13|0
Jitterness	12 (4%)	0|0|0|0|12
Others	8 (2%)	1|2|3|2|0

*Others involve constant speed, purpose and inconsistent skill.

Table 5: The frequency of top qualitative keywords that identify that the player is an AI.

Conclusion

- Main contribution: A new curiosity-based reward function
- Result: CC agent is the most human-like agent
- Significance: First reinforcement = learning agent grounded in psychological theory

Discussion

- Ignoring game contents
- Limitation: CC struggles in new levels such as 1-2, 1-3.

Any questions?