Does Infantile Attachment Require Intrinsic Reward?

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

Infantile attachment, a bond between infants and caregivers, involves proximity seeking behaviors to caregivers. Attachment is crucial for infants' cognitive development and social relationships, but debates about its underlying nature remain unresolved (Sroufe & Waters, 2017). Some theorists argue that attachment stems from an innate drive, as seen in the ethological view that humans, like other mammals, evolved to seek proximity to caregivers for protection (Bowlby, 1969). Others believe attachment is shaped by external reinforcement mechanisms (Gewirtz, 1972). This study uses reinforcement learning (RL) in a simple environment to explore whether infant attachment behaviors require intrinsic rewards for staying close to the caregiver or if external rewards from both the caregiver and the environment alone can generate attachment behaviors, especially those demonstrated by human infants with different attachment styles.

2 Experiment

We implemented a simple 1-dimensional grid world, where simulated infants explore and seek comfort with their caregivers. Both the infant agent and the caregiver are initialized at one end of the grid, with the caregiver remaining fixed. The infant can move left, right, or stay still. The caregiver can reward, ignore, or punish the infant, resulting in an extrinsic social reward. Infants get an additional extrinsic environmental reward when exploring the environment depending on its states. They receive a positive reward when they are in a happy state and a negative reward when they encounter environmental hazards and become distressed. In distress, infants can either seek caregiver for comfort or self-regulate to return to a happy state.

We implemented two groups of infants: one with an additional intrinsic attachment reward for staying close to the caregiver, and another without. For the intrinsically rewarded group, infants receive a higher reward when they are closer to the caregiver. The total reward is the intrinsic attachment reward plus the extrinsic social and environmental rewards. Figure 1 shows our preliminary result for the extrinsic group and more figure and implementation details can be found in Appendix.

3 Discussion

Our preliminary result showed that simple external reinforcement can generate human infant-like attachment behaviors and subgroups (Figure 1). Introducing intrinsic attachment rewards makes one part of the simulation more realistic such that some infants don't 'escape' caregivers from self-regulating their emotions, while inaccurately representing subgroup attachment patterns documented in the empirical literature (Figure 2). More work is needed to understand how the intrinsic and external rewards interact to drive the simulation and which simulation best represents human infant attachment patterns.

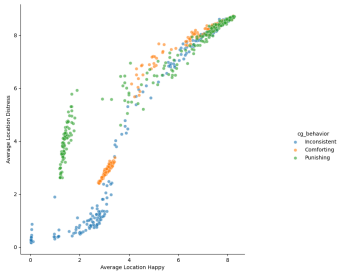

Figure 1: Preliminary Results Extrinsic Reward Group

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

4 References

[1] Bowlby, J. (1969). *Attachment and loss* (No. 79). Random House.

[2] Gewirtz, J. L. (1972). Attachment, dependence, and a distinction in terms of stimulus control.

[3] Sroufe, L. A., & Waters, E. (2017). Attachment as an organizational construct. *Interpersonal development*, 109-124.

5 Appendix

5.1 Preliminary result figure 2

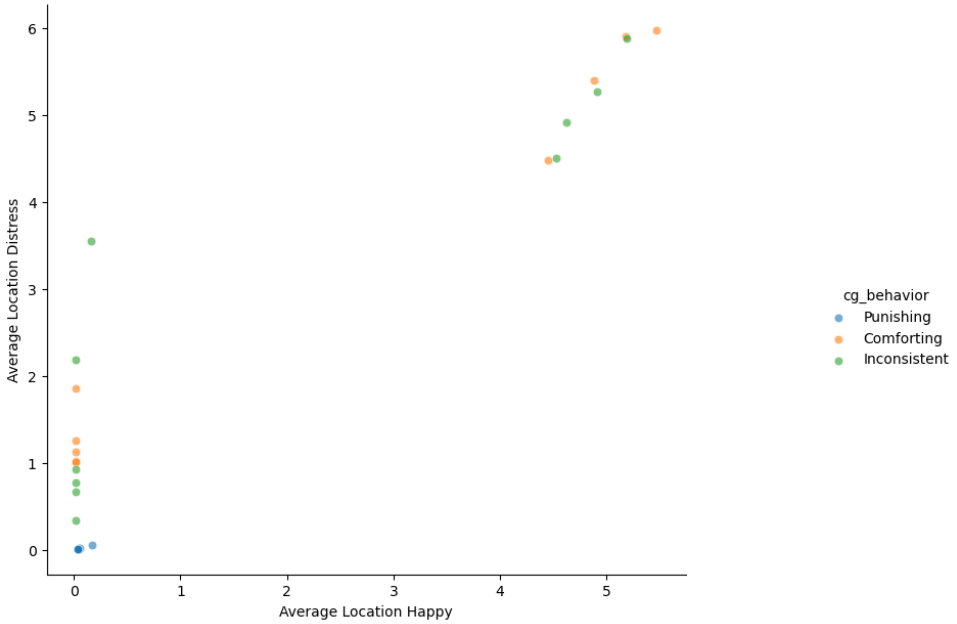

Figure 2: Preliminary Results Intrinsic Reward Group

5.2 Reward functions for both groups

Both groups collect extrinsic rewards from both interacting with the caregiver and the environment, but the intrinsically rewarded group has an additional reward from staying closer to the caregiver.

- **Extrinsically rewarded group:**

$$R_{env} = \beta \cdot R_{exploration} + (1 - \beta) \cdot R_{distress}$$

$$R_{total} = \alpha \cdot R_{env} + (1 - \alpha) \cdot R_{social}$$

- **Intrinsically rewarded group:**

$$R_{env} = \beta \cdot R_{exploration} + (1 - \beta) \cdot R_{distress}$$

$$R_{extrinsic} = \alpha \cdot R_{env} + (1 - \alpha) \cdot R_{social}$$

$$R_{attachment} = \frac{10}{DistanceFromCaregiver + 10}$$

$$R_{total} = \gamma \cdot R_{attachment} + (1 - \gamma) \cdot R_{extrinsic}$$

5.3 Parameters for both groups

Both groups have the same parameter values for alpha (ratio between environmental reward vs. social reward), beta (ratio between exploration vs. distress reward), probability of emotion regulation (probability of changing from distress state to happy state by infants themselves), and environmental positive and negative reward functions (infants collect positive or negative rewards depending on their emotional states in each location of the environment that are generated by a function).

The intrinsically rewarded group has an additional parameter gamma that defines the ratio between valuing intrinsic award of staying close to the caregiver vs. extrinsic rewards.

- **Extrinsically rewarded group:**

$$\alpha = 0.5 \sim 0.5$$
$$\beta = 0.5$$
$$ProbabilityofEmotionRegulation = 0.002 \sim 0.2$$
$$R_{exploration} = x + e^{0.4x},\ for\ x \in [1, 10]$$
$$R_{distress} = x + e^{0.2x},\ for\ x \in [1, 30]$$
$$EnvironmentalHazard = x + e^{0.4x},\ for\ x \in [1, 10]$$
$$R_{social} : r_s \in \{-1, 0, 1\}$$

- **Intrinsically rewarded group:**

$$\gamma = 0.5$$