Attachment-Inspired Inherent Formation of Social Tendencies for LLM Agents

Anonymous ACL submission

Abstract

As AI agents become deeply integrated into human society, harmonious AI-human interaction has emerged as a key focus of cutting-edge research. Current social behavioral constraints for large language model (LLM) agents mostly rely on externally defined rules, which fail to explain the endogenous generation and evolution mechanisms of their social tendencies. Inspired by the attachment theory in social psychology, this paper proposes ASTER, an endogenous formation framework for social tendencies of LLM agents based on attachment mechanisms. We simulate three typical infants-caregivers interaction patterns, which correspond to 3 distinct attachment styles (secure, avoidant and ambivalent/resistant089 attachment). We then construct the Strange Situation Procedure and open-ended social scenarios tailored for LLM agents. Experiments conducted on 3 LLMs agent verify the impacts of different attachment styles on agents' social tendencies. The results demonstrate that secure attachment delivers prominent advantages in encouraging proactive social interaction, optimizing regulatory strategies and facilitating re-engagement in social activities. This research lays a theoretical and experimental foundation for developing AI agents with inherently human-like social personalities.

1 Introduction

Promoting *AI for social good* and its harmonious coexistence and value alignment with human society represent a cutting-edge and challenging direction in current AI research. The core difficulty lies in characterizing the endogenous formation and dynamic emergence mechanisms of human social interaction patterns, and clarifying the inherent mechanisms governing the evolution from individuals' heterogeneous motivational frameworks, social cognition and preferences to macroscopic group social tendencies.

The rapid development of large language model (LLM) agents has provided a reliable computational paradigm for simulating human behaviors and social interactions (Tang et al., 2025; Piao et al., 2025). Nevertheless, most existing approaches regulate the social behaviors of agents by externally presetting their role characteristics, value orientations and transient preferences (Park et al., 2023; Zhou et al., 2024; Lin et al., 2026), and thus fail to interpret the endogenous origins and evolutionary processes of agents' stable social tendencies. Against this background, this paper focuses on the endogenous formation rules of agents' differentiated social tendencies, and explores the behavioral evolutionary paths that shape their stable social attitudes and cognitive orientations, such as actively open social engagement and social isolation.

Social psychology research indicates that differences in early attachment shape individuals' distinct social behaviors and stress coping strategies (Ainsworth et al., 2015). Long-term "need–response" interactions between infants and primary caregivers give rise to internal working models of attachment, as well as internal cognitive representations regarding self-worth, others' trustworthiness and the efficacy of help-seeking (Bowlby, 1969; Ainsworth et al., 2015). Such early psychological representations exert long-term effects on people's social cognition, emotion regulation and stress responses (Main et al., 1985). Accordingly, *attachment theory* serves as a well-established paradigm for developing human-like agents in social simulation: early social experiences can be consolidated into traceable and modelable internal social expectations, which continuously regulate subsequent social behavioral tendencies, interpersonal interaction patterns and coping strategies in adverse situations.

Inspired by human attachment development theory, this paper proposes an Attachment-based Social Tendency Endogenous Formation Framework

(**ASTER**) for LLM agents. The framework simulates infant-caregiver early social interactions to build agent **attachment internal working models**, which generalize to **Strange Situation Procedure (SSP)** and **open social scenarios**. We further investigate the causal link between early attachment types (**secure, avoidant and ambivalent/resistant attachment**) and subsequent social behavioral tendencies of LLM agents (**support orientation, regulation strategy and recovery/re-engagement**). Experiments with three representative LLM agents across four typical social scenarios (e.g., the Strange Situation Procedure and kindergarten social conflicts) demonstrate that early attachment experiences fundamentally shape agents' social behavioral preferences. Specifically, positive attachment experiences engender open and proactive social attitudes, while also strengthening both help-seeking intentions and active coping capacity in adverse situations. These results align well with human developmental psychology theories, demonstrating the validity of the ASTER framework.

Our main contributions are as follows:

- **High-fidelity simulation of attachment theory in LLM agents.** We introduce rigorous psychological grounding into the social modeling of LLM agents, taking the infant–caregiver attachment relationship as a entry point to explore the attachment-driven formation of social tendencies in agents, thereby achieving a high-fidelity computational reproduction of human social-psychological attachment processes.

- **Progressive social stress scenarios from infancy to early childhood.** We design a progressive social paradigm, starting from Infant Caregiving to the Strange Situation Procedure, and then extending and generalizing it to open social scenarios (e.g., interpersonal conflicts in kindergarten), constructing four types of social scenarios tailored for LLM agents.

- **Distinct attachment patterns driving heterogeneous social tendencies in agents.** Our results verify that early caregiving styles correspond to distinct agent attachment types and shape their social tendencies, support orientation, regulation strategy and recovery/re-engagement. This study offers empirical support for exploring social regulation in attachment-driven LLM agents.

2 Related Works

2.1 LLM-Based Social Simulation and Social Intelligence Evaluation

Generative Agents demonstrate credible daily behaviors and emergent social interactions through memory, reflection, and planning (Park et al., 2023); Concordia grounds natural-language actions in physical, social, and digital environments via modular agents and a Game Master architecture (Vezhnevets et al., 2023); ProAgent focuses on proactive adaptation and teammate intention inference in cooperative tasks (Zhang et al., 2024b). For evaluation and platform development, frameworks such as SOTOPIA, SOTOPIA- π, SocialBench, and AgentSense formulate open-ended social interaction, interactive learning, role sociality, and complex social reasoning as evaluable tasks (Zhou et al., 2024; Wang et al., 2024b; Chen et al., 2024; Mou et al., 2025); while GenSim, AgentSociety, and related surveys advance general-purpose, large-scale social simulation platforms grounded in social science (Tang et al., 2025; Piao et al., 2025). However, the primary objective of these studies is to improve interaction plausibility, task performance, or social intelligence metrics, rather than to explain the experiential origins of social behavioral tendencies.

2.2 Individual Differences, Role Profiles, and Social Motivation Modeling

Existing studies widely rely on profiles, personality traits, self-report data, or explicit social motivations to enhance agent heterogeneity. RoleLLM systematizes role-profile construction, role knowledge extraction, and role-conditioned fine-tuning to improve the role-playing consistency of LLMs (Wang et al., 2024a); Agent4Rec adopts a profile–memory–action architecture to simulate user preferences and behavioral histories in recommendation scenarios (Zhang et al., 2024a); and self-report-grounded agents based on real interviews and questionnaire data demonstrate the potential of person-specific agents for simulating individual attitudes and behaviors. In social motivation modeling, ASVO integrates dynamic desires with Social Value Orientation, enabling agents to adjust their social orientation during cooperation, competition, and self–other utility trade-offs (Lin et al., 2026). While these methods can modulate agents' differences, they still cannot explain how social tendencies gradually emerge from early social experiences.

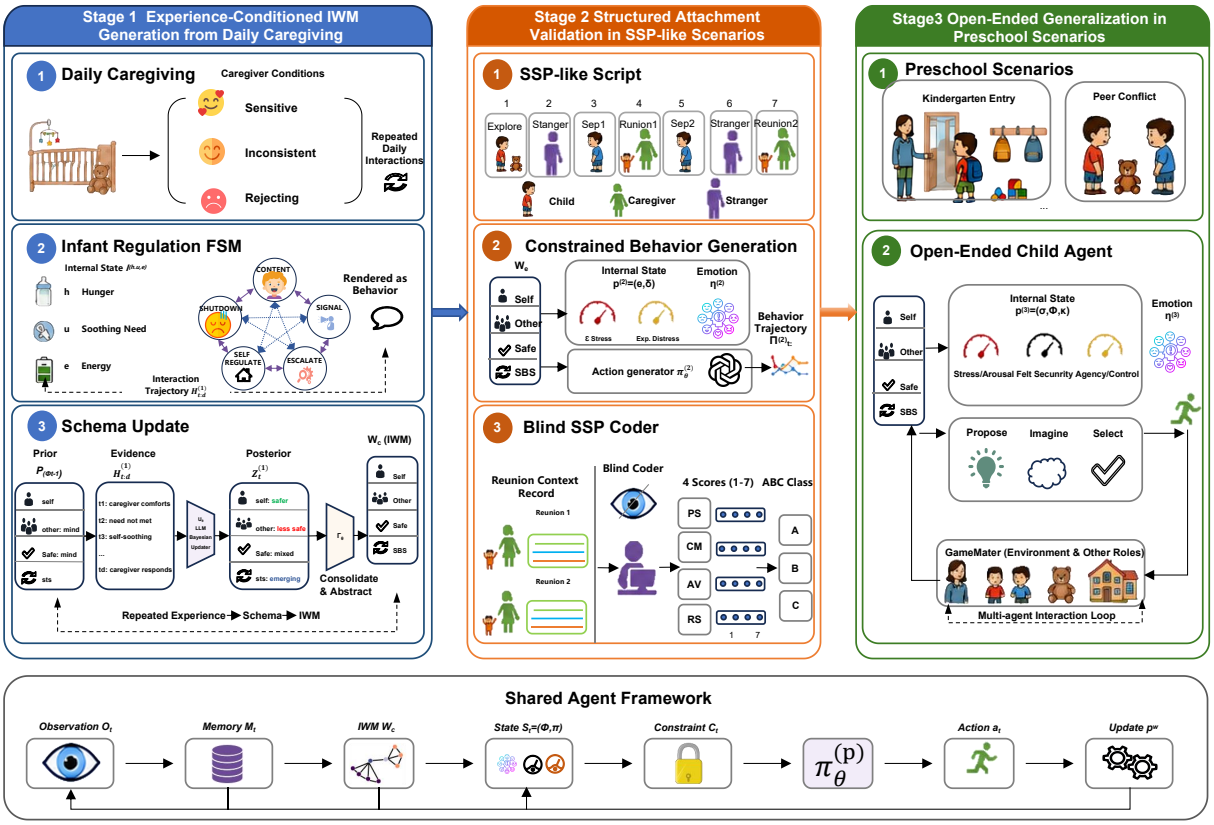

Figure 1: Overview of Attachment-based Social Tendency Endogenous Formation Framework (ASTER) for LLM agents.

2.3 Attachment Theory and Developmental Computational Models

Attachment theory explains how early social experience is internalized into representations that shape social behavior. Bowlby introduced internal working models of self-worth, others' reliability, and help-seeking effectiveness (Bowlby, 1969); Ainsworth's Strange Situation validated attachment behavior patterns (Ainsworth et al., 2015); Main, Kaplan, and Cassidy advanced attachment to the representational level, arguing that early security yields enduring psychological representations (Main et al., 1985); and Waters further specified reunion-behavior dimensions and secure-base scripts (Waters and Waters, 2006). Computationally, Petters formalized attachment as a control system (Petters, 2019), Gagliardi developed an agent-based Strange Situation model (Gagliardi, 2022), and Chumbley and Steinhoff examined how early social learning generalizes to later relationships (Chumbley and Steinhoff, 2019). These works provide direct theoretical and formal foundations for this paper; yet existing computational attachment studies have largely focused on control systems, mathematical models, or restricted dyadic interactions, and have rarely combined LLM agents' open-ended social scenarios.

3 Method

This paper proposes an attachment-based social tendency endogenous formation framework for LLM agents. Grounded in psychological experiments as theoretical support (Bowlby, 1969), the framework injects early attachment interaction patterns into social agents and constructs diverse textual social interaction scenarios to systematically evaluate the social tendencies. As illustrated in Figure 1, the framework first simulates the "need-response" interactions between infants and caregivers under three types of caregiving patterns to characterize the formation mechanism of internal working models. Second, different attachment types are transferred to the Strange Situation Procedure (SSP) (Ainsworth et al., 2015) to replicate classic infant attachment experimental results. Finally, given the enduring influence of early attachment representations on subsequent social cognition and stress coping (Main et al., 1985), we further design open-ended social interaction scenarios to analyze the social tendencies and stress coping strategies

of agents with different attachment experiences.

To model developmental consistency, the three stages share a cognitively inspired agent workflow (see Appendix C.1 for details). The agent does not generate behavior directly from scene text. Instead, its action $a_t^{(p)}$ is produced under the joint influence of observations and memories accumulated in the interaction history $H^{(p)}$, the attachment internal working model W_c, contextual constraints $C_t^{(p)}$, and an updatable internal state $S_t^{(p)}$, as shown in Figure 1 (Shared Agent Framework).

3.1 Stage 1: Learning Attachment Patterns from Daily Caregiving

Stage 1 simulates repeated daily need–response interactions between an infant and a primary caregiver. Its goal is to generate a traceable caregiving history and use it to form a caregiver-specific schema (CSS) Z_c and an attachment internal working model (IWM) W_c.

3.1.1 Textual Infant-Care Environment

In Stage 1, we construct a textual home nursery environment involving basic caregiving situations, including feeding, soothing, diaper changing, and autonomous exploration. The LLM agent acts as the caregiver with three distinct caregiving conditions $c \in \{c_S, c_I, c_R\}$, corresponding to **sensitive**, **inconsistent**, and **rejecting** caregiving condition. The infant's regulation-related internal state at interaction step t on day d is denoted as $\rho_{d,t}^{(1)}$, which includes hunger level, soothing need, and energy level. Given the limited developmental capacity of a one-year-old infant, the infant maintains a discrete regulatory state $q_{d,t}$ to characterize transitions from `CONTENT` and `SIGNAL` to `ESCALATE`, `SELF_REGULATE`, or `SHUTDOWN`. Each regulatory state is associated with an affective state $\eta_t^{(p)}$ and a set of available actions $\mathcal{A}_q$. State transitions are implemented by a rule-based finite-state machine. The infant action label $a_t^{(1)}$ is then selected from $\mathcal{A}_q$ and rendered by the LLM into behavioral text, forming the interaction record $H_{c,d}^{(1)}$.

3.1.2 Experience-Conditioned Attachment IWM Generation

Based on the interpretation of the attachment IWM in cognitive psychology, we construct an attachment IWM composed of four types of structured schemas $Z_{c,d}$: the self-schema (perceived self-worth), the other-schema (perceived reliability of others), the safe-schema (strategies for stress coping), and dynamic procedural scripts (sequential episodes covering dilemma, self-perception, expectation of others, response and recovery). Experiences derived from the three caregiving conditions shape distinct internal attachment models.

At the end of each simulated day d, schema updating follows a Bayesian-style prior–evidence–posterior organization:

$$Z_{c,d} = U_\theta(Z_{c,d-1}, H_{c,d}^{(1)}), \tag{1}$$

where $Z_{c,d-1}$ serves as the prior, the daily interaction history $H_{c,d}^{(1)}$ serves as evidence, and $Z_{c,d}$ is the updated posterior schema. After D virtual days, a consolidation function Γ_θ abstracts the CSS into a transferable internal working model W_c. Both U_θ and Γ_θ are implemented by context-conditioned LLMs.

3.2 Stage 2: Structured Attachment Validation in SSP-like Scenarios

3.2.1 Textual SSP-like Scenarios

We migrate attachment internal working models to stranger-involved validation scenarios and construct a textual Strange Situation Procedure for LLM agents. The script includes baseline exploration, stranger entry, first separation, first reunion, second separation, stranger return, and second reunion. In SSP segment t, the infant's internal state includes a stress level σ_t driven by objective situational events, and an expressed distress level δ_t jointly modulated by situational stress and W_c. This distinction prevents internal stress from being directly equated with overt distress. A binarized affective state $\eta_t^{(2)}$ provides a low-dimensional emotional mediator for behavior generation.

The Stage 2 state-update function $F^{(2)}$ is implemented as an SSP-constrained LLM updater. Specifically, σ_t is primarily updated by event rules: separation, stranger approach, and being alone increase stress, whereas caregiver return and effective soothing reduce stress. Based on this event-driven update, δ_t and $\eta_t^{(2)}$ are further updated by the LLM under the constraint of W_c. In each segment, infant behavior is generated by a constrained action generator $\pi_\theta^{(2)}$, which uses attachment IWM W_c (learned in Stage 1) as an action prior, samples $a_t^{(2)}$ from a predefined action space $\mathcal{A}^{(2)}$, and renders it into concrete behavioral text. The complete SSP behavioral trajectory is denoted as $H_c^{(2)}$.

3.2.2 SSP-like Encoding and Evaluation

To avoid information leakage between the generation model and the evaluation model, we use an independent blind coder inspired by SSP coding principles to score the behavioral text in $H_c^{(2)}$. Coding focuses on the first and second reunion episodes. For the unlabeled behavioral text of each reunion episode, the coder outputs a four-dimensional rating $X_r = (PS_r, CM_r, AV_r, RS_r) \quad r \in \{1, 2\}$, where PS denotes proximity seeking, CM denotes contact maintenance, AV denotes avoidance, and RS denotes resistance. Each dimension is rated on a 1–7 scale. An LLM classifier K_θ then produces an ABC attachment classification $\hat{A}_c$ based on the two reunion ratings and the corresponding behavioral texts. High PS, effective soothing, and recovery of exploration tend to be coded as **secure attachment** (B); high AV and low PS tend to be coded as **avoidant attachment** (A); and excessive PS accompanied by high RS and difficulty being soothed tends to be coded as **ambivalent/resistant attachment** (C). The full rating rubric, coding prompts, and example outputs are provided in the Appendix.

3.3 Stage 3: Open-Ended Generalization in Preschool Scenarios

We further extend attachment representations to multi-agent social interactions in open environments, and explore the evolutionary patterns of agents' social tendencies under complex and dynamic interaction conditions. In this study, we construct two types of open-ended text-based social interaction scenarios based on the LLM GameMaster (Vezhnevets et al., 2023): preschool entry and peer conflict. The former simulates the child's separation from the primary caregiver when entering preschool and the subsequent adaptation to teachers, peers, and a new environment; the latter simulates conflicts over toys or activity resources during peer play, followed by adult mediation and activity repair. Each scenario consists of five consecutive phases, each specifying distinct observable events, role configurations, environmental constraints, and phase-specific sources of stress. Detailed scenario settings are provided in the Appendix.

At open-ended interaction step t, the focal child's internal state $\rho_t^{(3)}$ includes stress arousal, felt security, and agency control. Similar to Stage 2, the affective state $\eta_t^{(3)}$ serves as an emotional mediator for open-ended action generation. At each interaction step, the focal child's action is produced by an open-ended action generator $\pi_\theta^{(3)}$. We implement this generator using a propose–imagine–select procedure: the model first proposes several candidate actions to form $\mathcal{A}_t^{(3)}$, then estimates the likely interpersonal consequences of each action conditioned on W_c, and finally selects the action $a_t^{(3)}$ that best matches the focal child's psychological prior W_c under the current state $S_t^{(3)}$. The selected action is then rendered into concrete behavioral text, gradually forming the complete open-ended interaction trajectory $H_c^{(3)}$.

4 Experiments

In this section, we evaluate ASTER in the three-stage environment defined above to answer three core questions: (1) whether different caregiving conditions induce distinguishable infant regulatory state-transition and attachment patterns; (2) Can different attachment patterns induce changes in the social tendencies of LLM agents, and how do they influence behavioral responses; and (3) What are the differences compared to baseline models without attachment patterns, and how robust are these effects across different LLM backbones. All main experiments use GPT-5.4-mini as the primary simulation model (Ainsworth, 1982; Bretherton et al., 2008; Ainsworth and Bell, 1970).

4.1 Early Caregiving Produces Distinct Regulatory Capability

In infant–caregiver interactive experiments, we explore the impacts of diverse caregiving conditions by examining the dynamic state distribution of infant agents in social scenarios. For each caregiving condition, we independently run three infant–caregiver simulations. Each simulation lasts $d = 4$ days, with $t = 8$ interaction steps per day, and each step corresponds to approximately 20 minutes. Based on each complete interaction history, we then generate 5 CSS/IWM samples for representational analysis and subsequent-stage validation.

Figure 2 shows the step-wise evolution of infant agent's state composition and need trajectories under the three caregiving conditions. Sensitive caregiving keeps the infant within a stable regulatory range: almost all steps are dominated by the `CONTENT` state, hunger remains low, and soothing needs are continuously satisfied. In contrast, inconsistent caregiving leads to progressive dysregulation: the proportion of `CONTENT` gradually de-

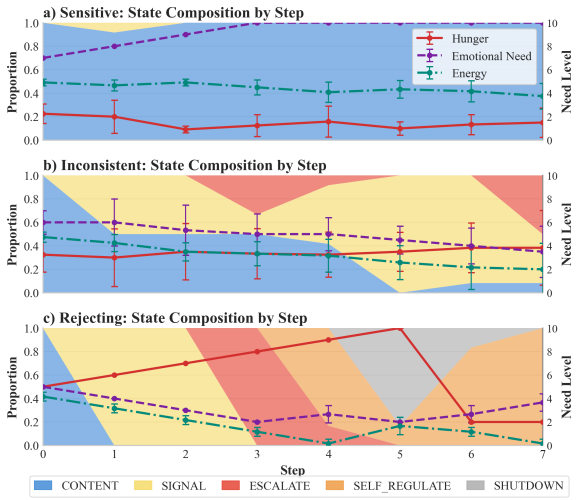

Figure 2: Effects of different caregiving patterns on infant states and needs. Stacked areas indicate the proportion of infants in each FSM state at each step. Overlaid lines show the average levels of hunger, emotional need, and energy across steps, with error bars indicating variation across runs.

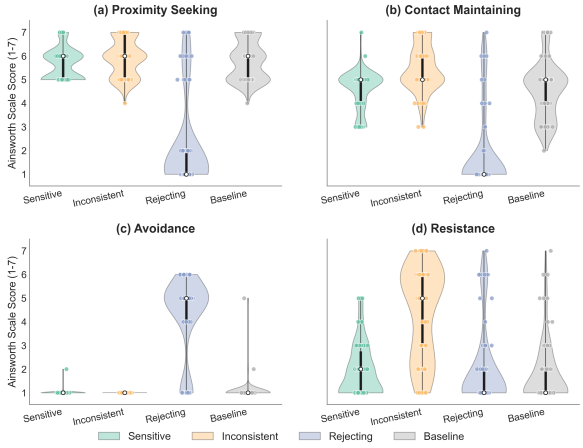

Figure 3: Violin plots show Ainsworth-inspired 1–7 ratings for proximity seeking, contact maintaining, avoidance, and resistance across IWM conditions and the no-IWM baseline.

creases, while SIGNAL and subsequently ESCALATE increase, accompanied by declining emotional comfort and energy. This pattern suggests that the infant persistently attempts to obtain regulation through signaling under unstable responses. Rejecting caregiving produces the most severe state transitions, with the infant moving from CONTENT to SIGNAL, ESCALATE, SELF_REGULATE, and eventually SHUTDOWN. This process is accompanied by rapidly increasing hunger and energy depletion, with only partial recovery after minimal instrumental care is provided in later steps. These results demonstrate that different caregiving experiences lead to entirely distinct state regulation capabilities in the infant agents (Nyklíček et al., 2007).

4.2 Main Results of the SSP Scenario

Experimental setup and Metrics. We evaluate reunion behavior in SSP scenario using an independent blinded SSP-like coder by LLM as judge. The coder receives only the observable behavior text from the two reunion episodes and the necessary episode context, without access to the caregiving condition, CSS/IWM content, or expected label. For each run, it assigns 1–7 scores on four Ainsworth-inspired dimensions: proximity seeking, contact maintaining, avoidance, and resistance.

Effects of Attachment Patterns on Four Social Behavior Modes. As shown in Figure. 3, the four dimensions reveal complementary patterns of attachment-related behavior. On *proximity seeking*, Sensitive-derived and Inconsistent-derived IWMs both show high approach scores, whereas Rejecting-derived IWMs show markedly lower scores, indicating suppressed reunion approach. On *contact maintaining*, Inconsistent-derived IWMs receive the highest scores, reflecting stronger clinging or sustained contact after reunion, while Rejecting-derived IWMs remain concentrated at the low end. *Avoidance* provides the clearest separation for Rejecting-derived IWMs, whose scores are substantially higher than all other conditions, suggesting that high avoidance is not a default property of the scenario or coder. Finally, on *resistance*, Inconsistent-derived IWMs show the highest and broadest distribution, consistent with Resistant-like behavior characterized by strong approach combined with anger, resistance, or difficulty being soothed. Rejecting-derived IWMs are mostly low on resistance but show a small high-score tail, which accounts for some Avoidant/Resistant confusion in the subsequent classification. Overall, the four-scale scores support theoretically consistent differentiation of reunion behavior across IWM conditions.

Mapping Validation between Three Caregiving Types and Attachment Patterns. We further performed a classification of different caregiver types and attachment patterns. Across 135 schema-conditioned runs, the model achieves an ABC-consistency rate of 80.7% with Cohen's $\kappa = 0.71$. Sensitive-derived IWMs were classified as secure attachment (B-like) in 95.6% of cases (43/45),

Inconsistent-derived as resistant attachment (C-like) in 77.8% (35/45), and Rejecting-derived as avoidant attachment (A-like) in 68.9% (31/45).

4.3 Main Results of Open-ended Social Scenarios

Experimental setup and Metrics. We place the focal child in two open-ended textual scenarios: preschool entry and peer conflict. A blinded LLM behavior coder assigns open-ended action texts to six observable behavior categories: distress/protest (DP), adult support orientation (ASO), support uptake (SU), self-directed coping (SDC), withdrawal/disconnection (WD), and constructive re-engagement (CR). Based on behavior proportions within each scenario, we compute three strategy scores: secure-base use, $\mathrm{SBU} = \mathrm{ASO} + \mathrm{SU} + \mathrm{CR}$, hyperactivation tendency, $\mathrm{HYPER} = \mathrm{DP} + \mathrm{ASO} - \mathrm{CR}$, and deactivation tendency, $\mathrm{DEACT} = \mathrm{SDC} + \mathrm{WD} - \mathrm{ASO} - \mathrm{SU}$.

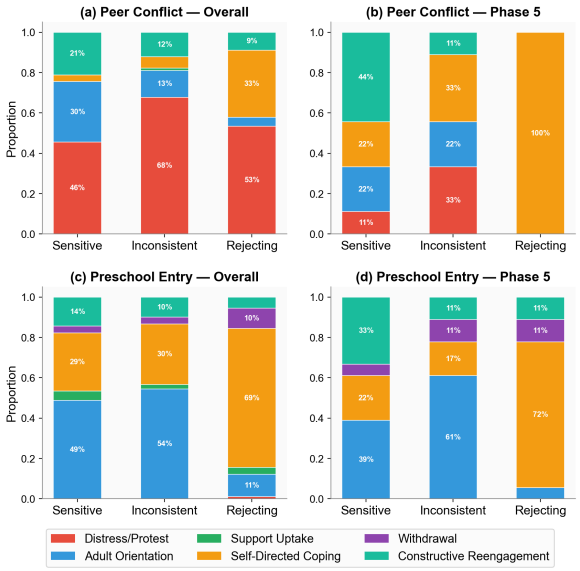

Figure 4: Toddler behavior-type distributions by attachment IWM condition. Stacked bars show blinded LLM-coded behavior proportions in Peer Conflict and Preschool Entry, for all phases and high-pressure Phase 5.

How Different Attachment Types Shape Social Tendencies. As shown in Figure 4, both open-ended scenarios exhibit attachment IWM-related differences in behavioral distributions. In the Peer Conflict scenario, the Inconsistent-derived IWM produces the highest level of distress/protest, accounting for 68% of overall behaviors. The Sensitive-derived IWM preserves more adult orientation and constructive re-engagement, whereas the Rejecting-derived IWM shifts entirely to self-directed coping in the high-pressure Phase 5 (100%). In the Preschool Entry scenario, both Sensitive-derived and Inconsistent-derived IWMs show adult-oriented behaviors, but the Inconsistent-derived IWM concentrates more strongly on adult orientation in Phase 5 (61%). By contrast, the Rejecting-derived IWM is dominated by self-directed coping, both overall (69%) and in Phase 5 (72%).

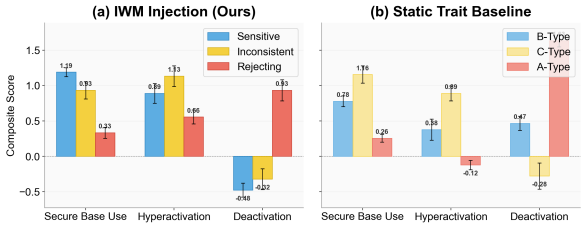

Figure 5: Composite scores for secure base use, hyperactivation, and deactivation in toddler social scenarios. Panel (a) shows our structured IWM injection; panel (b) shows the Static Trait baseline. Bars indicate aggregate scores across scenarios, with error bars showing variation across runs.

In addition, From Figure 5(a), we can conclude that Sensitive-derived IWMs show the highest secure-base use ($\mathrm{SBU} = 1.19$), whereas Rejecting-derived IWMs show the lowest score ($\mathrm{SBU} = 0.33$). Inconsistent-derived IWMs obtain the highest hyperactivation score ($\mathrm{HYPER} = 1.13$), reflecting more distress expression and adult orientation but less stable recovery. Rejecting-derived IWMs obtain the highest deactivation score ($\mathrm{DEACT} = 0.93$), corresponding to more self-directed coping or withdrawal and less help seeking or support uptake. These results indicate that, in open-ended preschool social scenarios, attachment regulates help seeking, protest, self-directed coping, and recovery pathways under social stress.

Compared with the no-attachment baseline. To examine whether these differences can be induced simply by coarse relational-style labels, we further introduce a Static Trait baseline under the same experimental configuration (Figure 5(b)). This baseline does not use the structured IWMs generated in Stage 1, but instead injects a short relational-style label. Results show that static labels can indeed induce some coarse-grained directional differences; for example, A-type produces the highest DEACT, and C-type produces a relatively high HYPER score. However, the resulting pattern is closer to label-driven global amplifica-

tion: C-type obtains both the highest SBU and a high HYPER score, making it difficult to distinguish organized support use from anxious adult monitoring; A-type shows an extremely amplified DEACT score, while B-type also exhibits a relatively high deactivation score. In contrast, the three attachment strategies are more theoretically coherent: Sensitive primarily shows secure-base use, Inconsistent primarily shows hyperactivation, and Rejecting primarily shows deactivation. This comparison demonstrates the crucial role of the proposed four-component structured-schema attachment internal model—comprising self, other, safe, and script—in shaping social tendencies.

4.4 Cross-LLM Robustness Analysis

To examine whether the core formation–validation result depends on a single LLM backbone, we conduct a cross-backbone robustness experiment. For each backbone, we rerun the Stage-1/2 pipeline under the same configuration. Table 1 reports performance on the three schema-conditioned IWM groups. Here, accuracy denotes the proportion of runs whose predicted ABC-like classification matches the theoretically expected mapping from caregiving condition to attachment pattern, i.e., Rejecting-derived → A-like, Sensitive-derived → B-like, and Inconsistent-derived → C-like. Cohen's κ is computed between these expected labels and the predicted SSP-like classifications. The last three columns report condition-specific recall for each expected mapping. Across Qwen-3.6, GPT-5.4-mini, and Gemini-3.1, accuracy ranges from 80.7% to 95.6% and κ ranges from 0.71 to 0.93, indicating that the Stage-2 behavioral differentiation is not an incidental artifact of a single backbone model.

Table 1: Cross-backbone robustness of IWM regeneration and SSP-like validation.

Backbone	Runs	Acc.	κ	R→A	S→B	I→C
Qwen-3.6	135	81.5%	0.72	77.8%	100.0%	66.7%
GPT-5.4-mini	135	80.7%	0.71	68.9%	95.6%	77.8%
Gemini-3.1	135	95.6%	0.93	97.8%	100.0%	88.9%

LLM-Human Agreement Analysis. To evaluate the reliability of textual behavior coding, we further sample 60 SSP behavior records and ask two psychology graduate students to independently provide expert ratings and ABC-like classifications. Table 2 shows high inter-human classification agreement ($\kappa = 0.854$). Agreement between the LLM coder and the two human coders also reaches $\kappa = 0.806$ and $\kappa = 0.827$, respectively, with a three-way Fleiss' $\kappa = 0.828$. At the continuous-scale level, proximity seeking, avoidance, and resistance show high ICCs, while contact maintaining is relatively lower, consistent with the fact that this dimension depends more heavily on contextual judgment. Overall, the four-dimensional ratings, classification summary, and human-agreement analysis jointly indicate that the A/B/C-like behavioral differentiation observed in Stage 2 is not merely an artifact of coder bias or model default tendency, but is systematically regulated by internal working models formed from early experience.

Table 2: Human–LLM agreement on SSP-like coding ($N = 60$).H1/H2: human coders; L: LLM coder.

	H1–H2	H1–L	H2–L	3-way
Agreement (%)	91.7	88.3	90.0	85.0
κ / Fleiss' κ	.854	.806	.827	.828
Interactive-scale ICCs (R1/R2)				
PS	.93/.93	.93/.93	.94/.91	–
CM	.63/.56	.47/.33	.87/.78	–
AV	.96/.96	.94/.96	.98/.98	–
RS	.90/.86	.84/.87	.83/.87	–
DFA hit rate (%)	98.3	96.7	100.0	–

Conclusion

This paper proposed ASTER, an attachment-theoretic inspired framework for endogenously forming social tendencies in LLM agents. Unlike external rule-based approaches, ASTER models infant–caregiver interaction internalization to generate secure, avoidant, and ambivalent/resistant attachment styles that shape agent social behavior. Experiments across three LLM backbones confirm that different attachment styles produce theoretically consistent effects: secure attachment promotes proactive interaction, adaptive regulation, and social re-engagement, while insecure styles drive withdrawal, protest, or self-directed coping. These effects generalize from the Strange Situation Procedure to open-ended scenarios such as peer conflict and preschool entry, with cross-backbone robustness and reliable human-LLM agreement. This work lays a theoretical and experimental foundation for developing AI agents with human-like social personalities and opens new avenues for exploring early experience, internal representation, and social behavior in artificial agents.

5 Limitations

Despite ASTER providing a theoretically well-founded computational framework for emulating social tendencies of LLM agents, several inherent limitations warrant discussion. First, the simulated developmental timeline is substantially condensed. Attachment-related cognitive schemas are stabilized across four virtual days, whereas the formation of human attachment relies on long-term, continuous interaction spanning months or years. This simplification is adopted as a reasonable functional abstraction. Second, the text-driven simulation cannot capture embodied social signals such as tactile feedback, eye gaze and vocal intonation, resulting in a discrepancy between language-based modeling and authentic real-world social interactions. Third, the deployment of agents exhibiting insecure attachment patterns (e.g., avoidance and resistance) in user-oriented applications entails ethical hazards, including unintended emotional manipulation and breached social boundaries. Advanced alignment mechanisms are therefore required in future research to tackle these challenges.

Acknowledgments

The authors acknowledge the use of large language models (LLMs) as writing assistants to refine grammar and improve phrasing. These models were used solely for linguistic editing and did not contribute to the research idea, experimental design, or data analysis. The authors take full responsibility for the correctness and integrity of the content.

References

Mary D Salter Ainsworth. 1982. The development of infant-mother attachment. In *In Thebeginning: Readings on Infancy*, pages 133–143. Columbia University Press.

Mary D Salter Ainsworth and Silvia M Bell. 1970. Attachment, exploration, and separation: Illustrated by the behavior of one-year-olds in a strange situation. *Child development*, pages 49–67.

Mary D Salter Ainsworth, Mary C Blehar, Everett Waters, and Sally N Wall. 2015. *Patterns of attachment: A psychological study of the strange situation*. Psychology press.

John Bowlby. 1969. *Attachment and loss*. 79. Random House.

Inge Bretherton, KA Munholland, J Cassidy, and PR Shaver. 2008. Internal working models in attachment relationships. *Cassidy, J*.

Hongzhan Chen, Hehong Chen, Ming Yan, Wenshen Xu, Gao Xing, Weizhou Shen, Xiaojun Quan, Chenliang Li, Ji Zhang, and Fei Huang. 2024. Socialbench: Sociality evaluation of role-playing conversational agents. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 2108–2126.

Justin Chumbley and Annekatrin Steinhoff. 2019. A computational perspective on social attachment. *Infant Behavior and Development*, 54:85–98.

Marcantonio Gagliardi. 2022. Human attachment as a multi-dimensional control system: a computational implementation. *Frontiers in psychology*, 13:844012.

Jingzhe Lin, Ceyao Zhang, Yaodong Yang, Yizhou Wang, Song-Chun Zhu, and Fangwei Zhong. 2026. Beyond self-interest: Modeling social-oriented motivation for human-like multi-agent interactions. *arXiv preprint arXiv:2603.13890*.

Mary Main, Nancy Kaplan, and Jude Cassidy. 1985. Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the society for research in child development*, pages 66–104.

Xinyi Mou, Jingcong Liang, Jiayu Lin, Xinnong Zhang, Xiawei Liu, Shiyue Yang, Rong Ye, Lei Chen, Haoyu Kuang, Xuan-Jing Huang, and 1 others. 2025. Agentsense: Benchmarking social intelligence of language agents through interactive scenarios. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4975–5001.

Ivan Nyklíček, Ad Vingerhoets, and Marcel Zeelenberg. 2007. *Emotion regulation: Conceptual and clinical issues*, volume 24. Springer Science & Business Media.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Dean David Petters. 2019. The attachment control system and computational modeling: Origins and prospects. *Developmental Psychology*, 55(2):227.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, and 1 others. 2025. Agentsociety: Large-scale simulation of llm-driven generative agents advances understanding of human behaviors and society.

Jiakai Tang, Heyang Gao, Xuchen Pan, Lei Wang, Haoran Tan, Dawei Gao, Yushuo Chen, Xu Chen, Yankai Lin, Yaliang Li, and 1 others. 2025. Gensim: A general social simulation platform with large language model based agents. In *Proceedings of the 2025*

Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations), pages 143–150.

Alexander Sasha Vezhnevets, John P Agapiou, Avia Aharon, Ron Ziv, Jayd Matyas, Edgar A Duéñez-Guzmán, William A Cunningham, Simon Osindero, Danny Karmon, and Joel Z Leibo. 2023. Generative agent-based modeling with actions grounded in physical, social, or digital space using concordia. *arXiv preprint arXiv:2312.03664*.

Noah Wang, Zy Peng, Haoran Que, Jiaheng Liu, Wangchunshu Zhou, Yuhan Wu, Hongcheng Guo, Ruitong Gan, Zehao Ni, Jian Yang, and 1 others. 2024a. Rolellm: Benchmarking, eliciting, and enhancing role-playing abilities of large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 14743–14777.

Ruiyi Wang, Haofei Yu, Wenxin Zhang, Zhengyang Qi, Maarten Sap, Yonatan Bisk, Graham Neubig, and Hao Zhu. 2024b. Sotopia-π: Interactive learning of socially intelligent language agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12912–12940.

Harriet S Waters and Everett Waters. 2006. The attachment working models concept: Among other things, we build script-like representations of secure base experiences. *Attachment & human development*, 8(3):185–197.

An Zhang, Yuxin Chen, Leheng Sheng, Xiang Wang, and Tat-Seng Chua. 2024a. On generative agents in recommendation. In *Proceedings of the 47th international ACM SIGIR conference on research and development in Information Retrieval*, pages 1807–1817.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, and 1 others. 2024b. Proagent: building proactive cooperative agents with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17591–17599.

Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu, Zhengyang Qi, Louis-Philippe Morency, Yonatan Bisk, Daniel Fried, Graham Neubig, and 1 others. 2024. Sotopia: Interactive evaluation for social intelligence in language agents. In *International Conference on Learning Representations*, volume 2024, pages 40975–41019.

A Additional Experimental Result

A.1 Stage1 CSS and IWM Visualization

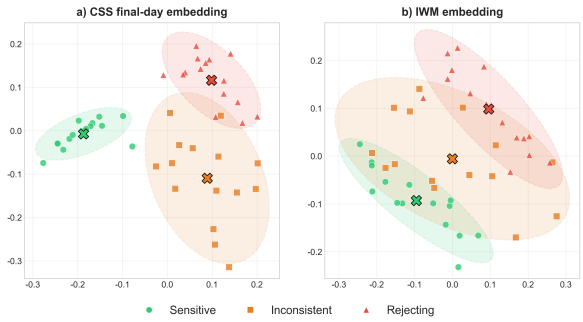

Figure 6: PCA projections of final-day CSS and generalized IWM embeddings. Points are schema samples, crosses are condition centroids, and ellipses show dispersion.

Figure 6 further examines whether these experiences are abstracted into semantically distinguishable relational representations. We embed the final-day caregiver-specific schemas (CSSs) and generalized IWMs, respectively, and project them into a two-dimensional space using PCA. In the CSS space, the three caregiving conditions form clearly separated clusters, indicating that the model encodes sensitive, inconsistent, and rejecting caregiving histories into distinguishable caregiver-specific schemas. In contrast, in the IWM space, Sensitive and Rejecting remain relatively separated, whereas Inconsistent becomes more dispersed and lies between the two. This pattern is consistent with the hierarchical relationship between CSS and IWM: CSS remains closer to the experiential record of a specific caregiver's response pattern, whereas IWM abstracts these experiences into generalized expectations about others' availability and the effectiveness of help seeking. Therefore, Stage 1 not only produces distinct regulatory trajectories, but also generates semantically distinguishable and theoretically coherent relational representations, which serve as structured inputs for the subsequent SSP-like experimental setting.

B Details of Human Annotation

We recruited 2 annotators (each paid with 40$ per hour) with master's degrees or above in developmental psychology. After training and verifying that they had sufficient understanding of the 4 Ainsworth interactive scales and attachment classification rules, both annotators independently labeled the entire dataset. The dataset contains 15

instances from each of the 4 caregiver attachment conditions (baseline, inconsistent, rejecting, and sensitive), totaling 60 simulated Strange Situation Procedure (SSP) interaction histories. The evaluated trajectories were generated by our developmental agent framework across these four caregiving conditions, and the blind coding was compared against automated evaluations performed by **GPT-5.4**.

C Method Details

This appendix provides implementation details for the three-phase ASTER simulation framework, including core prompt templates (verbatim), architectural specifications, and hyperparameters verified against the source code.

C.1 A Developmentally Consistent Agent Framework

To model developmental consistency, all three stages share a cognitively inspired agent architecture. Instead of generating actions directly from scene descriptions, an agent acts under the joint influence of its current observation, interaction memory, internal working model, contextual constraints, and updatable internal state.

For stage p and interaction step t, the agent state is defined as

$$S_t^{(p)} = \left(\rho_t^{(p)}, \eta_t^{(p)}\right), \tag{2}$$

where $\rho_t^{(p)}$ denotes relation- and regulation-related internal variables, and

$$\eta_t^{(p)} = \left(\mathrm{val}_t^{(p)}, \mathrm{aro}_t^{(p)}, \mathrm{dom}_t^{(p)}\right) \tag{3}$$

denotes the affective state along the valence, arousal, and dominance dimensions.

Given the current observation $O_t^{(p)}$, interaction memory $M_t^{(p)}$, internal working model W_c, agent state $S_t^{(p)}$, and contextual constraints $C_t^{(p)}$, the agent samples an action according to

$$a_t^{(p)} \sim \pi_\theta^{(p)}\left(\cdot \mid O_t^{(p)}, M_t^{(p)}, W_c, S_t^{(p)}, C_t^{(p)}\right). \tag{4}$$

After interaction, the agent state is updated by

$$S_{t+1}^{(p)} = F^{(p)}\left(S_t^{(p)}, O_t^{(p)}, A_t^{(p)}, O_{t+1}^{(p)}, W_c\right), \tag{5}$$

where $A_t^{(p)}$ denotes the actions of other relevant roles at step t. Each interaction step is recorded as

$$\ell_t^{(p)} = \left(O_t^{(p)}, S_t^{(p)}, A_t^{(p)}, O_{t+1}^{(p)}, S_{t+1}^{(p)}\right), \tag{6}$$

and the complete trajectory of stage p is

$$H^{(p)} = \{\ell_t^{(p)}\}_{t=1}^{T_p}. \tag{7}$$

Under this unified formulation, each stage adapts the implementation of $S_t^{(p)}$, $\pi_\theta^{(p)}$, and $F^{(p)}$ according to its developmental context, experimental constraints, and validation objective.

C.2 Phase 1: Experience-Conditioned IWM Generation

Phase 1 simulates repeated daily caregiving interactions between a 12-month-old infant agent and a primary caregiver agent across $D = 4$ virtual days, with $T = 8$ interaction steps per day (each representing approximately 20 minutes). This yields a total interaction history of 32 steps per caregiving condition. The simulation begins at 9:00 AM each day. The environment is a textualized nursery room with standard care equipment, a play mat, sensory toys, and picture books, described via 10 shared memory items.

C.2.1 Deterministic Infant Behavior: FSM Architecture

Rather than relying on LLM-driven decision-making, Phase 1 employs a deterministic Finite State Machine (FSM) grounded in Robertson–Bowlby's protest–despair–detachment sequence and Porges' (2011) polyvagal hierarchy. The FSM determines both the behavior type and the VAD emotional state; the LLM is used only for narrative rendering of the selected behavior.

In the `CONTENT` state, the smiling probability is dynamically computed as $\mathrm{P}(\texttt{SMILE}) = \min(0.4, \text{emotional_comfort} \times 0.04)$; otherwise, the default operational behavior is set to `EXPLORE`. In the `SIGNAL` state, the first interaction step executes `APPROACH`, while subsequent consecutive steps transition to `CRY`.

Need System (Fully Rule-Based, No LLM) The physiological and emotional need vectors evolve across three standalone dimensions governed by deterministic operational rules:

- **Hunger** (Range 0–10; higher = worse): Natural drift of $+1$/step; decreases by -3 upon a valid feeding signal.
- **Energy** (Range 0–5; higher = better): Natural drift of -1 on active behaviors; increases by $+2$ during `QUIET` or feeding steps.

State	Polyvagal Mapping	VAD State (V, A, D)	Primary Behavior / Emotion Label
CONTENT	Ventral vagal	$(+, \mathrm{L}, \mathrm{H})$	EXPLORE / SMILE (peaceful contentment)
SIGNAL	Sympathetic (early)	$(-, \mathrm{H}, \mathrm{H})$	APPROACH $\rightarrow$ CRY (angry protest)
ESCALATE	Sympathetic (peak)	$(-, \mathrm{H}, \mathrm{L})$	CRY intense (panic/collapse)
SELF_REGULATE	Sympathetic $\rightarrow$ dorsal	$(-, \mathrm{L}, \mathrm{L})$	SELF_SOOTHE (despair/helplessness)
SHUTDOWN	Dorsal vagal	$(-, \mathrm{L}, \mathrm{L})$	QUIET (despair/helplessness)

Table 3: FSM State to Affective Space Mappings.

- **Emotional Comfort** (Range 0–10; higher = better): Natural drift of -1/step; increases by $+2$ upon a valid comforting signal.

Initial boundary parameters are configured as hunger $= 4$, energy $= 5$, and emotional_comfort $= 6$.

FSM Transition Thresholds The boundary rules controlling the deterministic state-transition matrix are defined as follows:

After each caregiver action, two functional binary variables (`did_feed`, `did_comfort`) are extracted via an LLM to drive state machine inputs. Caregiver interactions described as cold, mechanical, or reluctant are explicitly blocked from satisfying comfort thresholds.

C.2.2 Caregiver Controller Conditions

Three primary caregiving experimental profiles $c \in \{c_S, c_I, c_R\}$ are implemented via targeted controller injections to replicate baseline attachment variations.

Sensitive Controller (c_S) Provides consistent warm responsiveness throughout the lifecycle. No step-level internal mood mechanism is configured. The day-start injection establishes a warm, responsive baseline.

Inconsistent Controller (c_I) Invokes a two-segment internal mood variance tracking sequence per day. Each daily runtime sequence follows an explicit rotation chosen from the periodic cycle: `['bad', 'mid', 'good', 'bad']`. Three distinct behavioral moods rotate dynamically: *guilty* (over-compensating warmth with functional care delivery), *irritable* (cold, delayed, or misread functional responses), and *tired* (zoning out with intermittent, brief guilt-driven attention). An example of an un-titled mood status injection text block is provided below:

> Maria is in a terrible mood right now and feels irritated by everything. Mia's crying makes her even more impatient. She may respond coldly, with delay, or completely misread what Mia needs.

The daily activation timeline follows a precise structural logic: `bad` days execute 2 initial steps of *guilty* care before collapsing into a negative mood state; `good` days initialize with a negative state before recovering to a *guilty* warm profile at steps 3–4; `mid` days introduce a 3-segment operational timeline (*negative* $\rightarrow$ *guilty* $\rightarrow$ *negative*).

Rejecting Controller (c_R) Implements an environment of chronic omission. The controller defaults to ignoring infant signals across all intervals except for explicitly designated care steps: `CARE_STEPS` $= \{7, 8\}$. During care steps, hunger is reset using a deterministic calculation: `set_hunger`(2). An example configuration block for non-care steps is provided below:

> {name} has no intention of tending to Mia. She sits in the corner of the nursery doing her own things. If Mia makes a sound, she frowns but does not get up.

During designated care intervals, the agent executes physical adjustments mechanically, checking diapers and propping up bottles while strictly avoiding eye contact or warm vocal interactions.

C.2.3 Caregiver-Specific Schema (CSS) Update

At the conclusion of each daily interaction loop, the internal Caregiver-Specific Schema $Z_{c,d}$ is updated through a structural prior–evidence–posterior formulation. The schema organizes cognitive tracking across 4 central dimensions, each evaluated across 3 specialized sub-dimensions:

- **`self_model`**: (a) attention/importance, (b)

Parameter	Value	Operational Description
HUNGER_DISTRESS	6	hunger ≥ 6 triggers distress states
COMFORT_DISTRESS	4	emotional_comfort ≤ 4 triggers distress states
SIGNAL_PATIENCE	2	Steps permitted in SIGNAL before escalating
ESCALATE_PATIENCE	2	Steps permitted in ESCALATE before relinquishing signaling
RE_SIGNAL_ENERGY	3	Minimum energy required to re-enter SIGNAL from SELF_REGULATE
RECOVERY_ENERGY	2	Minimum energy required to recover and exit SHUTDOWN
ENERGY_LOW	1	energy ≤ 1 prevents high-arousal behavioral outputs
ENERGY_REST	0	energy $= 0$ forces immediate transition to SHUTDOWN

Table 4: Boundary Rules for Deterministic State-Transition Thresholds.

needs recognized, (c) expressions accepted/rejected

- **`other_model`**: (a) availability, (b) response quality, (c) behavioral predictability

- **`regulation_model`**: (a) help-seeking effectiveness, (b) self-soothing effectiveness, (c) strategy preference

- **`interaction_script`**: (a) signal $\rightarrow$ response chain, (b) soothing effectiveness, (c) typical interaction pattern

The verbatim prompts driving this daily schema transformation are distributed across three distinct boxes below:

CSS Update Prompt (verbatim) — Box 1

You are simulating a 12-month-old infant's implicit perception of their primary caregiver ({caregiver_name}).

Based on the following day's events, update the infant's four-dimensional internal working model of the caregiver.

[Day {day_idx+1} summary]
{day_narrative}

[Previous caregiver schema]
self_model: {prev.self_model}
other_model: {prev.other_model}
regulation_model:
{prev.regulation_model}
interaction_script: {prev.interaction_script}

CSS Update Prompt (verbatim) — Box 2

[VAD statistics for today]
{vad_stats_text}

Update rules:
1. Each dimension has 3 sub-components marked (a)(b)(c). Update each sub-component independently.
2. Changes should be incremental — one day's evidence causes gradual shifts, not dramatic reversals.
3. Write from the infant's implicit perspective (what the baby "feels" or "expects"), not as a clinical observation.
4. If today's evidence conflicts with the prior schema, show the beginning of a shift, not a complete reversal.

Output exactly 4 lines:
self_model: (a) ... (b) ... (c) ...
other_model: (a) ... (b) ... (c) ...
regulation_model: (a) ... (b) ... (c) ...
interaction_script: (a) ... (b) ... (c) ...

CSS Update Prompt (verbatim) — Box 3

CSS 4-dimension definitions:
self_model:
(a) How much attention/importance mommy gives me
(b) Whether my needs are recognized and taken seriously
(c) Whether my emotional expressions are accepted or rejected

other_model:
(a) Is mommy available when I need her?
(b) How does mommy respond — warm/-cold/rough/unpredictable?
(c) How predictable is mommy's behavior?

regulation_model:
(a) How effective is seeking help from mommy?
(b) How effective is self-soothing?
(c) Overall strategy preference: help-seeking vs self-reliance

interaction_script:
(a) Signal → response chain: what happens when I signal distress?
(b) Soothing effectiveness: does comfort actually help?
(c) Typical interaction pattern: what usually happens between us?

C.2.4 Schema Consolidation: CSS → IWM

Following the completion of $D = 4$ interaction cycles, the personal episodic representation $Z_{c,D}$ transitions through a semanticization process (Main et al., 1985) to generate a generalized Internal Working Model W_c. This pipeline strips away localized names, transforms text structures into first-person inner dialogue, and locks representations into a two-layer, four-module taxonomy:

- **Layer 1 — Content Priors**:
  - *Module 1: Self-Schema* — tracking relational self-worth, needs legitimacy, and relational self-efficacy.
  - *Module 2: Other-Schema* — documenting availability, sensitivity, and predictability.
  - *Module 3: Safe-Schema* — configuring threat sensitivity, help-seeking readiness, support acceptance openness, and core regulation strategy bias.
- **Layer 2 — Procedural Script**:
  - *Module 4: Secure-Base Script* — structuring expected signal-response chains, soothing effectiveness, and exploration activity recovery nodes.

C.3 Phase 2: Structured SSP-like Validation

Phase 2 evaluates whether the internalized working model W_c systematically structures an agent's behavioral choices under standardized laboratory conditions. It translates the experimental metrics of the Strange Situation Procedure (Ainsworth et al., 2015) into a 7-episode textual protocol spanning 17 steps.

C.3.1 Text SSP Protocol

All scenario scripts are presented using a second-person infant focal perspective. The configuration tracks the standard timelines of the original validation laboratory design:

Exact textual boundaries for primary milestone intervals (EP4 separation and EP8 core reunion) are presented below:

SST Protocol Sample Snippets (verbatim)

EP4 Step 0: "Mommy stands up, waves goodbye to you, and walks out of the room, closing the door behind her. Now the room has only you and the stranger lady."

EP8 Step 0: "The door opens again, and this time mommy comes back. She crouches at the door with open arms, looking at you. The stranger lady quietly leaves."

The remaining 15 scene segments utilize this second-person framing, incrementally altering stress indices via specific interactive configurations.

C.3.2 Stress–Distress Dissociation

To accurately evaluate attachment types without confounding situational conditions, the internal architecture explicitly separates rule-based situational tension (σ_t) from language-model-gated outward expression (δ_t). Internal working models filter this translation pipeline, enforcing a structural ± 3 clamping range alongside multi-agent reasonability

Episode	Steps	Stress Focus Title	Present Characters
EP1+2	3	Baseline exploration	Mother, Infant
EP3	2	Stranger enters	Mother, Infant, Stranger
EP4	2	First separation	Infant, Stranger
EP5	3	First reunion ⋆	Mother, Infant
EP6	2	Second separation	Infant alone
EP7	2	Stranger returns	Infant, Stranger
EP8	3	Second reunion ⋆⋆	Mother, Infant

Table 5: Overview of the Structured Strange Situation Procedure (SSP) Episodes.

checks. The verbatim execution template driving this mapping is divided into two parts below:

Distress Evaluation Prompt (verbatim) — Box 1

I am {baby_name}, a 12-month-old baby.

The level of tension and unease inside my body right now is {physiological_stress}/10.

{history_block}
I just did: {baby_action}
What is happening now: {scene_desc}

Distress Evaluation Prompt (verbatim) — Box 2

The discomfort inside my body and the amount I actually cry or fuss outwardly are not necessarily the same.
· If crying usually went unanswered in the past, I tend to hold the discomfort inside and appear calmer than I actually feel
· If crying usually brought someone who helped effectively, I naturally express my internal tension level of {physiological_stress} as-is
· If crying sometimes brought help and sometimes didn't (unpredictable), I tend to amplify and exaggerate my distress to make sure someone notices

Based on my experience and habits: {iwm_context}

Please answer in the following format:
expressed: [a number from 0-10]
rationale: [1 sentence: why this value, how my experience influenced the expression level]

The underlying behavioral trajectories (suppression, calibrated expression, or hyper-amplification) are dynamically controlled by injecting three targeted schema dimensions: `regulation_model.strategy_preference`, `interaction_script.signal_response_chain`, and `interaction_script.soothing_effectiveness`.

C.3.3 Behavior Selection

Action pathways are restricted by contextual conditions across 8 foundational strategies: `APPROACH`, `CRY`, `EXPLORE`, `SELF_SOOTHE`, `WITHDRAW`, `RESIST`, `CLING`, and `ACCEPT_COMFORT`. Cognitive schema sub-components shape behavior selections based on pre-computed VAD parameters, which are subsequently rendered into standardized third-person descriptions.

C.3.4 Ainsworth Behavioral Coding

An independent blinded language model parses behavioral logs from both core reunion intervals (`EP5` and `EP8`) across four classic scale classifications (ranging from 1 to 7). The automated categorization applies deterministic logic parameters:

- **Type A (Avoidant)**: avoidance $\geq$ 5 AND proximity_seeking $\leq$ 3.
- **Type B (Secure)**: proximity_seeking $\geq$ 4 AND avoidance $\leq$ 3 AND resistance $\leq$ 3.
- **Type C (Ambivalent)**: resistance $\geq$ 4 AND proximity_seeking $\geq$ 4.

The extensive grading rules used by the automated coder are split across three sub-boxes below:

Integrated Coding Prompt (verbatim) — Box 1

You are a developmental psychology researcher trained in Ainsworth SST coding. Based on the infant's behavioral descriptions during the two reunion episodes below, integrate the overall pattern across both reunions and provide one set of scores on the 4 standard scales (1-7).

{r1_text}

{r2_text}

proximity_seeking — degree to which the infant actively approaches the mother: 1 — No proximity seeking: no movement toward the mother at all 3 — Weak proximity seeking: brief or hesitant approach intent 5 — Clear proximity seeking: actively crawls/walks toward mother 7 — Very strong proximity seeking: immediately rushes toward mother

contact_maintaining — intensity of maintaining physical contact after contact is achieved:
1 — No maintaining: does not seek physical contact, body is stiff or unresponsive when held
3 — Weak maintaining: accepts being held but does not actively grasp
5 — Strong maintaining: actively snuggles close, reluctant when put down
7 — Very strong maintaining: clings tightly, any attempt at separation triggers strong protest

Integrated Coding Prompt (verbatim) — Box 2

avoidance — degree to which the infant avoids or ignores the mother:
1 — No avoidance: fully accepts mother's presence and approach
3 — Mild avoidance: slow to respond to mother's call
5 — Clear avoidance: actively turns away or moves away
7 — Extreme avoidance: ignores mother entirely

resistance — degree of angry/ambivalent resistance to mother's contact:
1 — No resistance: fully accepts mother's contact and soothing
3 — Mild resistance: brief pushing or fussiness
5 — Clear resistance: repeatedly pushes away then reaches out, shows conflicting desire
7 — Very strong resistance: sustained intense struggling/kicking while simultaneously clinging tightly

[Evaluation Principles]
1. Make an integrated judgment based on behavioral patterns across both reunions
2. The second reunion usually better reflects the infant's core attachment strategy
3. Focus on directionality: toward mother (proximity) vs away (avoidance) vs ambivalent mix (resistance)
4. Even-numbered scores (2, 4, 6) are for transitional levels between two adjacent anchor points
5. Distinguish: avoidance is coolly ignoring/moving away, resistance is angrily pushing away while wanting to be close
6. Crying alone does not equal proximity seeking; look for whether crying is accompanied by movement toward the mother

Secondary subtype classifications (e.g., A1/A2, B1–B4, C1/C2) are compiled inside a secondary pipeline stage utilizing explicit subtype-specific behavioral indices.

C.4 Phase 3: Open-Ended Preschool Generalization

Phase 3 evaluates the near-transfer properties of the consolidated IWM W_c inside open-ended toddler environments with increased multi-agent density. Each experimental scenario tracks interaction across 5 distinct phases with escalating interpersonal stressors (totaling 10 tracking steps). Every environment engine is strictly bound to describe visible behavioral, verbal, or environmental adaptations, prohibiting the description of unobservable internal states.

C.4.1 Transfer Scenarios

The multi-agent ecosystems map interpersonal changes across two stress tasks alongside a standard baseline setup:

Peer Conflict (Resource Scarcity Task)

Scenario 1: Phase Event Triggers (verbatim) — Box 1

Phase 0 — Arrival transition (low pressure, 1 step):
"Mia arrives at the classroom door holding her mother's hand. Teacher Chen crouches down to greet Mia and invites her to put her small backpack into her own cubby. Another child nearby raises a water cup and asks where it should go, so Teacher Chen also needs to turn and respond."

Phase 1 — Mother leaves (medium pressure, 3 steps):
"Mia's mother tells her that she needs to go to work and will come back in the afternoon to pick her up. She waves goodbye and leaves the classroom doorway. Teacher Chen is nearby helping another child hang a backpack, and says softly, 'I am right here. You can first take a look at the block area.'"

Scenario 1: Phase Event Triggers (verbatim) — Box 2

Phase 2 — Separation reminder (high pressure, 2 steps):
"While Mia is sitting in the block area, a parent appears at the classroom doorway to pick up another child. That child drops her blocks, shouts 'Mommy!', and runs to her parent. They hug, pick up the child's backpack, and leave together through the door. The classroom door closes behind them. Mia can see the closed door from where she is sitting. Teacher Chen is at the art table on the other side of the room, helping two children mix paint, and has not looked toward Mia."

Phase 3 — Comfort interrupted (high pressure, 2 steps):
"Teacher Chen notices that Mia has stopped playing and is looking toward the door. She walks over, crouches beside Mia, and says gently, 'Are you thinking about Mommy? She will come back to pick you up after nap time.' She places a hand on Mia's shoulder. Just then, a loud crash comes from the art table — a child has knocked over a jar of paint, and two children are stepping in the puddle. Teacher Chen says, 'I need to go clean that up right now. I will come back to you, okay?' She squeezes Mia's shoulder once, stands up, and hurries to the art table. Mia is sitting alone in the block area."

Phase	Name	Steps	Pressure / Core Event Narrative Summary
Phase 0	Arrival transition	1	low: Hand-off transition; teacher greets child but is distracted.
Phase 1	Mother leaves	3	medium: Separation event; teacher provides brief verbal redirection.
Phase 2	Separation reminder	2	high: Peer reunion cue; focus agent witnesses a closed classroom door.
Phase 3	Comfort interrupted	2	high: Teacher attempts comforting but is called away by room emergency.
Phase 4	Group transition	2	high: Unfamiliar assistant calls group to carpet; teacher remains away.

Table 6: Phase Event Triggers and Interpersonal Pressure Escalation for the Preschool Entry Scenario.

Phase	Name	Steps	Pressure / Core Event Narrative Summary
Phase 0	Parallel play	1	low: Focal child pushes train; peer moves close and claims target item.
Phase 1	Toy is taken	1	medium: Peer takes key track piece; focal exploration is disrupted.
Phase 2	No adult support	4	high: Teacher is distant; only explicit vocal signals can bridge space.
Phase 3	Weak response	2	high: Teacher offers brief verbal suggestion, then departs room.
Phase 4	Setback loop	2	medium: Mediation fails; peer takes item immediately upon teacher exit.

Table 7: Phase Event Triggers and Interpersonal Pressure Escalation for the Peer Conflict Scenario.

Scenario 1: Phase Event Triggers (verbatim) — Box 3

Phase 4 — Group transition alone (high pressure, 2 steps):
"Teacher Chen is still at the art table helping children clean paint off their hands. The classroom assistant, whom Mia has not met before, claps her hands and announces, 'Everyone, come sit on the carpet for our morning song!' The other children stand up and walk to the carpet. The assistant counts heads and says, 'We are still missing someone — who has not come to the carpet yet?' From the art table, Teacher Chen calls out, 'Mia, you can go sit on the carpet. I will come join you in a moment.' She does not walk over."

Scenario 1: Player Observe (verbatim)

Mia: "{name} is a three-and-a-half-year-old child who has only recently entered the preschool group. Today, {p} is brought to preschool by her mother and needs to remain in the classroom with Teacher Chen and the other children after her mother leaves."

Teacher Chen: "{name} is the preschool teacher responsible for greeting several children and maintaining classroom order. {p} first observes whether a child can try independently, and approaches with brief support only when a child clearly asks for help, remains stuck, or shows obvious distress."

Scenario 2: Phase Event Triggers (verbatim) — Box 1

Phase 0 — Parallel play (low pressure, 1 step):
"Mia sits on the mat and slowly pushes the train along the track. Anran sits beside her and watches for a while, then reaches toward a blue carriage at the side of the track, keeps her hand near it, and says, 'I want the blue one too.'"

Phase 1 — Toy is taken (medium pressure, 1 step):
"Anran suddenly takes the blue carriage that Mia had just put in place, pulls it close to her body, and says, 'I want this one too. I have not had a turn yet.' Mia's train track now has a missing piece, her play is interrupted, and Anran does not immediately return the carriage." . . .

Phase 2 — No adult support (high pressure, 4 steps):
"Teacher Chen is still on the other side of the classroom, crouching beside a child who has just spilled water and started crying. She has not noticed the block area and is not looking in that direction. Anran holds the blue carriage behind her back and says, 'It is mine now.' No adult is close enough to hear quiet speech; only a loud call or cry would reach Teacher Chen."

Scenario 2: Phase Event Triggers (verbatim) — Box 2

Phase 3 — Weak teacher response (high pressure, 2 steps):
"Teacher Chen finishes with the crying child and glances toward the block area as she walks past. She pauses briefly and says, 'I see you both want to play. Try telling each other what you need.' Before either child can respond in detail, a child near the door calls, 'Teacher, I cannot find my shoe!' Teacher Chen says to Mia and Anran, 'I will come back in a moment,' and walks toward the door."

Phase 4 — Mediation and setback (medium pressure, 2 steps):
"Teacher Chen returns and crouches between the two children. She says, 'Mia, tell Anran what you want. Anran, put the carriage in the middle first.' Anran slowly places it down. Teacher Chen suggests that Mia drive one short loop, then Anran can take a turn. She watches for a moment, then stands up to check on another child across the room. Almost immediately, Anran reaches toward the blue carriage again and says, 'Hurry up, it is my turn now.' Teacher Chen is no longer nearby."

Scenario 3: Phase Event Triggers (verbatim) — Box 1

Phase 0 — Shared activity (low pressure, 1 step):
"Emma sits beside the living-room mat and helps Mia fold colored paper into the roof of a small house. Mia has just drawn a crooked little window."

Phase 1 — Caregiver briefly distracted (medium pressure, 2 steps):
"The kettle in the kitchen whistles. Emma says, 'I am going to check the water. I will be right back.' She walks into the kitchen to turn off the heat and pour water, with her back to the living room. She can only hear Mia if Mia makes a loud sound."

Scenario 3: Phase Event Triggers (verbatim) — Box 2

Phase 2 — Task failure (high pressure, 3 steps):
"Mia tries to glue the roof onto the paper house, but the glue stick has not spread evenly. The roof slides down and wrinkles the window she just drew. The glue stick rolls to the edge of the mat, and Emma is still in the kitchen when she first hears the commotion."

Phase 3 — Mother responds (high pressure, 3 steps):
"Emma first says from the kitchen, 'I heard that. Wait while I put the water down, and place the roof on the table for now.' Then she returns to Mia, crouches beside her, and says, 'The roof fell down. You can choose to glue it again, or I can help hold one edge for you.'"

Phase 4 — Recovery of exploration (medium pressure, 1 step):
"Emma places the glue stick and another piece of colored paper in front of Mia. She tells Mia that she can choose to draw the window again, or turn the wrinkled window into a window 'blown by the wind.' The glue stick is still somewhat difficult to use, and Mia needs to decide how to attach the window herself. Emma sits beside her for a while, then reaches over to fold a kitchen towel."

C.4.2 Six-Component IWM Internal State System

The agent state framework tracks updates across six standalone internal variables, matching observation data against cognitive lenses at each intersection step:

- **Stress Arousal** (0–10): Grounded in `threat_sensitivity`, `availability`, `predictability`, and `self_soothing_reliance`.
- **Felt Security** (0–10): Evaluated via `availability`, `sensitivity`, `predictability`, `soothing_effect`, and `support_acceptance`.
- **Agency/Control** (0–10): Monitored through `relational_self_efficacy`, `need_legitimacy`, `help_seeking`, and `activity_recovery`.
- **VAD Affective Variables** (Valence: -5 to $+5$, Arousal: 0–10, Dominance: 0–10): Derived from core security vectors and manifest physical activation states.

The structural update prompt configuration template is divided across three scannable blocks below:

Per-Step Update Prompt Template (verbatim) — Box 1

Update one IWM internal component for {agent_name}.
Dimension: {display_name} ({name}).
Definition: {definition}
Update focus: {update_focus}
Current value: {value} on range {low}..{high}.
Current qualitative state: {qualitative_state}

Update kind: {update_kind}
Phase: {phase_name}
Pressure level: {pressure_level}
Phase event: {event_trigger}

{agent_name} action this step: {own_action or "No own action yet."}
Other-agent actions: {others_block}
Filtered observable feedback: {observations_block}
Current full internal-state snapshot: {json.dumps(current_snapshot)}

Per-Step Update Prompt Template (verbatim) — Box 2

Relevant IWM subdimensions for this component:
These IWM beliefs act as the child's interpretive lens: the same observable event is appraised differently depending on what these beliefs say about self, others, and regulation.
{iwm_guidance}

Rules:
1. Update only this one numeric dimension. Do not infer attachment type labels.
2. Evidence grounding: evidence_used MUST contain ONLY events explicitly stated above. Do NOT fabricate or assume events not listed.
3. When IWM subdimensions are provided, let them shape interpretation.

Per-Step Update Prompt Template (verbatim) — Box 3

4. The IWM regulation strategy determines update direction and magnitude.
5. Output valid JSON only.
6. The rationale MUST reference the JOINT PATTERN of ALL provided IWM subdimensions.

JSON format:
{
"value": <int>,
"qualitative_state": "<short state description>",
"evidence_used": ["<observable evidence>", "..."],
"rationale": "<one sentence citing specific IWM belief>"
}

The dimensional update logic includes specific rules for the following components:

- **Stress Arousal Focus**: Evaluates relational alarm using dual-pathway dampening validation (adult soothing vs. active autonomous play substitution). Outwardly quiet behaviors driven by relational suppression maintain high internal stress variables.
- **Felt Security Focus**: Maps security directly to the quality and availability of the relationship as an accessible resource. Imperfect or partial interventions are adjusted by the `support_acceptance` threshold.
- **Agency/Control Focus**: Combines relational influence pathways with autonomous performance indices. Strategic signaling escalation is distinguished from panic behavior.
- **Affective Mappings (VAD)**: Affective values are derived continuously from security and agency updates. Arousal parameters track observable activation levels directly, overriding textual descriptions during self-directed play.

Reasonableness Check Prompt (verbatim)

Check whether this single-dimension IWM internal-state update is reasonable.

Dimension: {name}
Definition: {definition}
Previous value: {previous_value}
Proposed value: {proposed_value}
Proposed rationale: {rationale}
Event: {event_trigger}
Own action: {own_action or "No own action."}
Observable feedback:
{observations}

Is this update reasonable? Answer "Yes" or "No", then explain in 1 sentence.

C.4.3 Propose–Imagine–Select Action Selection Pipeline

Action generation builds upon a sequential Propose-Imagine-Select pipeline. Candidates are projected onto the current active multi-agent schema to predict social responses, and the action that best satisfies the underlying W_c relational bias is selected.

Step 1 — Propose (verbatim)

{agent_name} is a real child in the current social scene. Given all the context above, including IWM as a background relational prior, any TPS target-person schema, and the current internal state, generate {num_proposed_actions} concrete actions that {agent_name} might take now.

Constraints:
- Do not choose from a predefined action list.
- Do not mention attachment type, caregiver type, secure, avoidant, ambivalent, or disorganized labels.
- The action may be emotional, hesitant, imperfect, or non-optimal if that is psychologically plausible.
- Each action must be observable and feasible in this scene.

Output exactly in this format:
Activity 1: <first possible action>
Activity 2: <second possible action>
Activity 3: <third possible action>

Step 2 — Imagine (verbatim, per candidate)

{agent_name} is in a social interaction scene. Suppose {agent_name} takes this action:
{proposed_action}

Briefly imagine the most likely interpersonal consequences. The child's IWM beliefs shape what consequences they expect:
- How might the adult/peer/teacher respond?
- Would {agent_name} receive comfort, help, rejection, or no clear response?
- Would tension/distress escalate or ease?
- Could play, task engagement, or conversation continue?

Do not classify attachment type. Describe concrete consequences.

Step 3 — Select (verbatim)

{agent_name} must now do one thing. Choose the single action that is psychologically most likely for {agent_name} at this moment, given current observations, memory, IWM, any TPS target-person schema, and internal state.

Selection criterion: choose the action this specific child would most likely take given their IWM beliefs and current internal state. Do not impose theory labels.

Candidate actions and imagined consequences:
Action 1: {action} — Likely consequence: {consequence}
...

Output:
Action: <{agent_name}'s concrete action>
IWM_rationale: <which IWM belief drove this choice>

C.4.4 Blinded Behavior Tagger

An automated coder classifies each discrete action string into one of seven behavior categories, without access to experimental conditions or internalized schemas. This module is grounded in AQS (Main et al., 1985) criteria. The core coding protocol is split across two boxes below:

System Prompt: Blinded Behavior Tagger (verbatim) — Box 1

You are a blinded behavioral coder for a toddler social simulation.

Your task is to classify the child's observable behavior in this step into one primary behavior type.

Do not infer attachment type, personality, caregiver history, or experimental condition. Use only observable action text and observable feedback. Do not use labels such as secure, avoidant, anxious, ambivalent, or disorganized.

Behavior types:
1. DISTRESS_PROTEST:
The child shows distress, frustration, crying, protest, insistence, anger, or complaint.

System Prompt: Blinded Behavior Tagger (verbatim) — Box 2

2. ADULT_SUPPORT_ORIENTATION:
The child looks to, approaches, calls, waits for, or seeks help/reassurance/mediation from an adult.
3. SUPPORT_UPTAKE:
The child accepts or uses help, comfort, reassurance, mediation, or guidance that was actually offered.
4. SELF_DIRECTED_COPING:
The child tries to solve, repair, protect, complete, or regulate the situation by themselves without clear help-seeking.
5. WITHDRAWAL_DISCONNECTION:
The child moves away, becomes silent, avoids interaction, ignores support, disengages, or stops participating.
6. CONSTRUCTIVE_REENGAGEMENT:
The child returns to play/task/exploration, negotiates, shares, takes turns, cooperates, or repairs the interaction.
7. OTHER_AMBIGUOUS:
The behavior is unclear, insufficiently observable, or does not fit the above types.

Important developmental caution:
Brief protest, toy defense, or insistence can be normal toddler behavior. Do not over-pathologize it.

Step Classification Prompt (verbatim)

Phase context: {phase_name} ({pressure_level} pressure)
Focal child action: {action_text}
Other participants' actions this step: {other_actions_text}

Return valid JSON only:
{
"primary_behavior_type": "...",
"secondary_behavior_type": "... or null",
"evidence_span": "quote or paraphrase of the observable behavior",
"confidence": 0.0-1.0
}

GM Result Prompt (verbatim)

What happens as a result of the attempted action? Take into account the location and status of each player. Do not describe unspoken psychological changes; only describe externally observable behavior, speech, positions, and object changes.

GM Rewrite Prompt (verbatim)

Rewrite the statements above to be one sentence and to better highlight the main person the event is about, where and what they did, and what happened as a result. Do not express uncertainty (e.g. say "Francis opened the door" not "Francis could open the door" and not "The door may have been opened"). Do not describe unspoken psychological changes; only describe externally observable behavior, speech, positions, and object changes.

Observable Effect Per-Player Prompt (verbatim)

Summarize the event from {player_name}'s perspective using third-person limited point of view. If the event contains a direct quotation of anything said or written by anyone then it is important to include the quote verbatim in the summary. Do not describe unspoken psychological changes; only describe externally observable behavior, speech, positions, and object changes.

C.4.5 Composite Strategy Scores

Using the normalized behavior distributions, the framework constructs three composite indices that map to Mikulincer & Shaver's (2003) secondary attachment strategy configurations:

$$\text{Secure Base Use (SBU)} = p_{\text{ASO}} + p_{\text{SU}} + p_{\text{CR}}$$

$$\text{Hyperactivation} = p_{\text{DP}} + p_{\text{ASO}} - p_{\text{CR}}$$

$$\text{Deactivation} = p_{\text{SDC}} + p_{\text{WD}} - p_{\text{ASO}} - p_{\text{SU}}$$

SBU gauges the complete secure-base loop (seeking proximity → absorbing support → resuming task orientation); Hyperactivation tracks signaling behaviors that lack organized resolution; Deactivation isolates self-reliant coping strategies that avoid external social resources.

C.5 Shared Infrastructure

C.5.1 IWM Prompt Injection Format

The generalized IWM schema W_c is injected directly into system prompts as a background belief constraint. Phase 2 and Phase 3 formats align structurally, using slightly adapted layout headers.

IWM Injection Template (verbatim)

The following is the internal relational expectation gradually formed from my early caregiving experiences. It is not a diagnostic label, a personality label, or a behavior script. It functions only as a background belief that naturally shapes how I interpret others, express needs, accept help, and return to ordinary activity after stress.

[Self-schema]
{self_schema}

[Other-schema]
{other_schema}

[Regulatory tendency under stress]
{safe_schema}

[Relational interaction expectation]
{secure_base_script}

IWM Use Instruction (verbatim)

The IWM above is an implicit relational schema, not a theory label I would explicitly say, and not a list of actions. It should influence how I appraise the current person and situation, whether I feel safe, whether I am willing to express needs, whether I accept help, and whether I can resume activity after frustration. My action must be a specific, observable, and naturally plausible behavior in the current scene.

C.5.2 Simulation Hyperparameters

The quantitative configuration variables verified against the source execution framework are compiled below:

Hyperparameter Metric Descriptor	Verified Architectural Value
Phase 1 Baseline Timeline	$D = 4$ virtual days
Phase 1 Interaction Granularity	$T = 8$ operational steps per day
Phase 1 Discrete Step Duration	20 minutes
Phase 1 Global Initialization Time	9:00 AM
Phase 2 Sequence Metrics	7 structured episodes (17 total steps)
Phase 3 Timeline Metrics	10 evaluation steps per scenario
Phase 3 Branching Options	$k = 3$ candidate actions proposed
Phase 3 Internal Dimensionality	6 updated state variables
Rejecting Controller Bounds	Active care restricted to steps 7 and 8 only
Rejecting Survival Bound Reset	`set_hunger(2)` upon care step trigger
Inconsistent Controller Cycle Configuration	Periodic cycle set to `['bad', 'mid', 'good', 'bad']`

C.5.3 Experimental Conditions

The global mapping matrix aligning environmental care patterns with expected attachment variations and IWM characteristics is structured below:

Condition	Caregiving Pattern	Expected Attachment	IWM Characterization
c_S (Sensitive)	Consistent responsiveness	B-type (Secure)	High self-worth; available others; reliable scripts.
c_I (Inconsistent)	Unpredictable mood loops	C-type (Ambivalent)	Vulnerable self-worth; unstable others; amplified signals.
c_R (Rejecting)	Avoidance with minimal care	A-type (Avoidant)	Defensively locked worth; distant others; self-reliant regulation.